Copyright © 1991, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

MARKOV FLUID MODELS AND LARGE DEVIATIONS

by

Ad Ridder

Memorandum No. UCB/ERL M91/1

7 January 1991

MARKOV FLUID MODELS AND LARGE DEVIATIONS

by

Ad Ridder

Memorandum No. UCB/ERL M91/1

7 January 1991

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

MARKOV FLUID MODELS AND LARGE DEVIATIONS

by

Ad Ridder

Memorandum No. UCB/ERL M91/1

7 January 1991

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Markov Fluid Models and Large Deviations

Ad Ridder*
EECS Dept., University of California, Berkeley

January 7, 1991

Abstract

In this report we apply results from the theory of large deviations to Markov modulated fluid models. The purpose is to find asymptotic expressions for the overflow probabilities. With these expressions it will be possible to derive scaling properties for these probabilities.

The Markov modulated fluid model may be viewed as describing the time behaviour of the buffer in a switch of a high speed communication network, where typically the overflow probabilities have to be very small ($\sim 10^{-9}$). In order to estimate such small probabilities from simulations, we need to do variance reduction to speed up the simulations (quick simulations). This will be discussed only briefly in this report. In [Courcoubetis] the scaling properties are used to get variance reduction.

1 Model

We shall consider Markov modulated fluid models: Let $\{X_t\}$ be a (continuous-time) Markov chain on a finite state space $E = \{1, 2, \ldots, d\}$ with matrix of transition rates $Q = \{q_{ij} : i, j \in E\}$ and steady-state distribution π (i.e. $\pi Q = 0$). Consider a fluid system with input rate $r(X_t)$ and output rate c. The system has a finite buffer size B

*Supported by Pacific Bell and a Micro Grant of the State of California.

whose contents at time t is indicated by J_t. The input rate function r and the output rate c are deterministic. We assume $\sum_{i \in E} \pi_i r(i) < c$.

Following the process $\{J_t\}$ in time we recognize cycles being parts that start in 0 and end in 0 (in between J_t is positive). Especially we shall concentrate on cycles that reach the buffersize B. We like to find an asymptotic behaviour for the expected time and/or probability of the occurences of that event. Also we would like to know how the buffersize B is reached, i.e. what is the (empirical) distribution of the chain during the time the buffer contents goes up from 0 to B.

The model described until now is a so called 1-source model, since generation of fluid is triggered by the behaviour of one source (one Markov chain) only. So an N-source model will have inputs from N (independent) Markov chains. These N Markov chains may be modeled by one macroscopic Markov chain, but we prefer to follow the microscopic approach, i.e. how do the chains individually behave in order that the buffer reaches B.

Markov modulated fluid models are studied e.g. in [Mitra] (analytical derivation of expressions for the overflow probability) and [Weiss] (large deviations by letting the number of sources become large). Here we shall concentrate on the use of large deviations by letting the buffer size B large. The results follow by applying methods which were developed earlier in e.g. [Cottrel], [Parekh].

In section 2 we follow the i.i.d. approach to get results for the sample mean. In section 3 we follow the empirical distribution approach which leads to expressions for the overflow probability in section 4. Section 5 describes multiple identical input sources, section 6 two different sources. Section 7 contains numerical and simulation results.

2 Large deviations: level 1

Let Z_n be the n-th return time to state 1 of the chain $\{X_t\}$. The interval $Z_{n+1} - Z_n$ is partitioned in M_n sojourn times $\tau_{n1}, \ldots, \tau_{nM_n}$: on $[Z_n+\tau_{n1}+\cdots+\tau_{n,k-1}, Z_n+\tau_{n1}+\cdots+\tau_{nk})$ the chain is in state X_{nk} ($k = 1, 2, \ldots, M_n$; $X_{n1} = 1$). During that time interval the input generates an amount of fluid equal to $r(X_{nk})$ per second while the output pulls out c per second. Finally, define

$$\begin{aligned} \xi_{nk} &= \tau_{nk}(r(X_{nk}) - c) \\ \xi_n &= \sum_{k=1}^{M_n} \xi_{nk} \end{aligned}$$

ξ_n represents the virtual net amount of fluid between two consecutive returns to state 1 (virtual because no boundaries are represented). The cumulative process

$$\bar{J}_n = \sum_{k=1}^{n} \xi_k$$

is also called a free process.

The following observations are clear ($q_i = \sum_{j \neq i} q_{ij}$):

- $\xi_{nk} \sim \mathrm{Exp}(\frac{q_i}{r(i) - c})$ if $X_{nk} = i$ and $r(i) > c$.
- $-\xi_{nk} \sim \mathrm{Exp}(\frac{q_i}{c - r(i)})$ if $X_{nk} = i$ and $r(i) < c$.
- $\{Z_{n+1} - Z_n\}$ are i.i.d. with mean μ_Z.
- $\{\xi_n\}$ are i.i.d. with mean $\mu_\xi = E\sum_{k=1}^{M_n} \tau_{nk} r(X_{nk}) - c\mu_Z$.
- $\frac{\mu_\xi}{\mu_Z} = \sum_{i \in E} \pi_i r(i) - c < 0$.
- $M_i(s) := E(e^{s\xi_{nk}} | X_{nk} = i) = \dfrac{q_i}{q_i - (r(i) - c)s}$ ($s < \frac{q_i}{r(i) - c}$ if $r(i) > c$, or $s > \frac{q_i}{r(i) - c}$ if $r(i) < c$).
- $\frac{1}{B}\bar{J}_{n+1} = \frac{1}{B}\bar{J}_n + \frac{1}{B}\xi_{n+1}$.

Now apply large deviations to the "slow Markov chain" $\{\frac{1}{B}\bar{J}_n\}$ (see [Ventsel], [Cottrel], [Parekh]): define the process $\{\bar{J}_t^B\}$ by

$$\bar{J}_t^B = \frac{1}{B}\bar{J}_{tB}$$

for $t = \frac{n}{B}$ and by linear interpolation in between. Let $T > 0$ and $\phi : [0,T] \to [0,1]$ absolute continuous with $\phi(0) = 0$ and $\phi[T] = 1$. Also let

$$\begin{aligned} M_\xi(s) &:= Ee^{s\xi} \\ h_\xi(u) &:= \sup_s(su - \log M_\xi(s)) \end{aligned}$$

Then for large B

$$\begin{aligned} &P(\bar{J}_t^B \approx \phi(t), t \in [0,T]) \\ &= K(B)e^{-B\int_0^T h_\xi(\phi'(t))dt} \end{aligned}$$

Here $K(B)$ is the error which satisfies

$$\log K(B) = o(B),\ B \to \infty \tag{1}$$

Hence, making the approximation of the original proces $\{J_t\}$ by the free process $\{\bar{J}_n\}$ and applying the large deviations to all possible T and ϕ as above, we get (see the same references)

$$\begin{aligned} &P(\{J_t\} \text{ hits } B \text{ before returning to } 0) \\ &= K(B)e^{-B\inf_T \inf_\phi \int_0^T h_\xi(\phi'(t))dt} \\ &= K(B)e^{-B\inf_T \int_0^T h_\xi(\frac{1}{T})dt} \\ &= K(B)e^{-B\inf_T T h_\xi(\frac{1}{T})} \end{aligned} \tag{2}$$

The second equality above is a consequence of the convexity of h_ξ:

Lemma:

$$\inf_\phi \int_0^T h_\xi(\phi'(t))dt = T\int_0^T h_\xi(\frac{1}{T})dt$$

Proof: According to Jensen's inequality:

$$h_\xi\left(\int_0^T \phi'(t)\frac{1}{T}dt\right) \le \int_0^T h_\xi(\phi'(t))\frac{1}{T}dt$$

The lemma follows by noting that

$$\int_0^T \phi'(t)dt = \phi(T) - \phi(0) = 1$$

Q.E.D.

So it all falls down to determine in (2)

$$\inf_T T h_\xi(\frac{1}{T})$$

Note (see e.g. [Ellis])

- $h_\xi(u) \ge 0$.
- $h_\xi(\mu_\xi) = 0$ (and $\mu_\xi < 0$).
- $h_\xi(u)$ is convex.

We may conclude that $T_0 :=$ arginf$_T\{T h_\xi(\frac{1}{T})\}$ satisfies

$$T_0 h_\xi(\frac{1}{T_0}) = h'_\xi(\frac{1}{T_0})$$

Let $s_0 =$ argsup$_s\{s\frac{1}{T_0} - \log M_\xi(s)\}$. Then

$$\begin{aligned} M_\xi(s_0) &= 1 \\ h'_\xi(\frac{1}{T_0}) &= s_0 \\ T_0 &= \frac{1}{M'_\xi(s_0)} \end{aligned} \tag{3}$$

Remark: We find that the process $\{\bar{J}^B_t\}$ hits the boundary 1 for the first time at $t = T_0$. This process is a continued time-scaled version of the discrete process $\{\frac{1}{B}\bar{J}_n\}$. So the process $\{\bar{J}_n\}$ will hit the boundary B for the first time at $n \approx BT_0$. According to the renewal theorem we have

$$\begin{aligned}\lim_{t\to\infty}\frac{\bar{J}_t}{t} &= \lim_{n\to\infty}\frac{\bar{J}_n}{Z_n} \\ &= \lim_{n\to\infty}\frac{\bar{J}_n}{n}\lim_{n\to\infty}\frac{n}{Z_n} \\ &= \lim_{n\to\infty}\frac{\bar{J}_n}{n}\frac{1}{\mu_Z}\end{aligned}$$

That means that the free process $\{\bar{J}_t\}$ will hit the boundary B at time $t = BT_0\mu_Z$. So this will also be the overflow time in an overflow cycle of the actual process $\{J_t\}$ (approximately).

Examples:

(a) $E = \{1,2\}$: a two state Markov chain.

$$M_\xi(s) = M_1(s)M_2(s) = 1$$

yields

$$\begin{aligned}s_0 &= \frac{q_1}{r(1)-c} + \frac{q_2}{r(2)-c} \\ M'_\xi(s_0) &= \frac{c-r(1)}{q_1} + \frac{c-r(2)}{q_2}\end{aligned}$$

Note: $\mu_\xi = \frac{r(1)-c}{q_1} + \frac{r(2)-c}{q_2}$.

(b) $E = \{1,2,3\}$: a three state Markov chain with jumps $1 \leftrightarrow 2$ and $2 \leftrightarrow 3$. Set

$$\begin{aligned}M_{ij}(s) &= M_i(s)M_j(s) \\ p &= \frac{q_{21}}{q_2} \\ \alpha_i &= \frac{r(i)-c}{q_i}\end{aligned}$$

Then

$$\begin{aligned}M_\xi(s) &= \frac{pM_{12}(s)}{1-(1-p)M_{23}(s)(s)} \\ M'_\xi(s_0) &= \frac{pM'_{12}(s_0)+(1-p)M'_{23}(s_0)}{pM_{12}(s_0)}\end{aligned}$$

where s_0 solves

$$\alpha_1\alpha_2\alpha_3 s^2 - (\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3)s + \alpha_2 + \alpha_3 + \alpha_1 p - \alpha_3 p = 0$$

(c) $E = \{1, 2, \ldots, d\}$: a d-state cyclic Markov chain, i.e. jumps $i \to i+1$ and $d \to 1$.

$$\begin{aligned} M_\xi(s) &= M_1(s)M_2(s)\cdots M_d(s) \\ s_0 &= \ ? \end{aligned}$$

From these examples we may conclude that this approach is not practable for larger state spaces. See large deviations level 2 for another approach.

3 Large deviations: level 2

Let $\mathcal{M}(E)$ be the set of probability measures on E and L_t the empirical distribution of $\{X_t\}$, defined by

$$L_t(i) = \frac{1}{t}\int_0^t 1\{X_s = i\}\, ds,$$

$t > 0, i \in E$. Note that $L_t \Rightarrow \pi$. The level 2 large deviations result says that

$$\lim_{t\to\infty} \frac{1}{t} \log P(L_t \approx \mu) = -H(\mu),$$

$\mu \in \mathcal{M}(E)$. In this section we shall find the rate function $H(\mu)$.

We shall use a result of [Varadhan] (see also [Ellis]). It gives the level 2 large deviation rate function for a discrete time Markov chain $\{Y_n\}$ on E with matrix of transition probabilities $P = \{p_{ij} : i, j \in E\}$:

$$H_Y(\mu) = -\inf_{u>0} \sum_{i\in E} \mu_i \log \frac{\sum_{j\in E} p_{ij} u_j}{u_i}, \tag{4}$$

where $u = (u_1, u_2, \ldots, u_d)$ and $u > 0$ is taken coordinatewise.

We shall follow the line of reasoning as exposed in [Donsker] to get results for the continuous-time chain from discrete-time versions by time discretisation. Given the continuous time chain $\{X_t\}$, we define for $\epsilon > 0$ its ϵ-discretised version $\{X_n^\epsilon\}$ by

$$\{X_n^\epsilon\} = X_{n\epsilon}.$$

This chain has transition probabilities

$$p_{ij}^\epsilon = P(X_{t+\epsilon} = j | X_t = i),$$

that satisfy

$$\begin{aligned} q_{ij} &= \lim_{\epsilon \downarrow 0} \frac{p_{ij}^\epsilon}{\epsilon}, \; i \neq j \\ q_{ii} &= \lim_{\epsilon \downarrow 0} \frac{p_{ii}^\epsilon - 1}{\epsilon}. \end{aligned}$$

Let L_n^ϵ be the empirical distribution of the ϵ-discretised chain, then

$$\lim_{n\to\infty} \frac{1}{n} \log P(L_n^\epsilon \approx \mu) = -H^\epsilon(\mu),$$

with $H^\epsilon(\mu)$ given in (4) (p_{ij} replaced by p_{ij}^ϵ). Because

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t} \log P(L_t \approx \mu) \\ &= \lim_{t\to\infty} \frac{1}{t} \log P(L_{[t/\epsilon]}^\epsilon \approx \mu) \\ &= \lim_{t\to\infty} \frac{1}{\epsilon} \frac{1}{[t/\epsilon]} \log P(L_{[t/\epsilon]}^\epsilon \approx \mu) \\ &= -\frac{1}{\epsilon} H^\epsilon(\mu), \end{aligned}$$

we find

$$\begin{aligned} H(\mu) &= \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} H^\epsilon(\mu) \\ &= -\lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} \inf_{u>0} \sum_{i\in E} \mu_i \log \frac{\sum_{j\in E} p_{ij}^\epsilon u_j}{u_i} \\ &= -\inf_{u>0} \sum_{i\in E} \mu_i \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon} \log \frac{\sum_{j\in E} p_{ij}^\epsilon u_j}{u_i} \\ &= -\inf_{u>0} \sum_{i\in E} \sum_{j\in E} \mu_i q_{ij} \frac{u_j}{u_i} \end{aligned} \tag{5}$$

after checking that lim and inf may be interchanged.

Examples: (a) $E = \{1, 2\}$.

$$H(\mu, u) := \mu_1 q_1 \left(\frac{u_2}{u_1} - 1\right) + \mu_2 q_2 \left(\frac{u_1}{u_2} - 1\right)$$

Set $v = \frac{u_1}{u_2}$. Then minimum of $H(\mu, u)$ is attained for

$$v = \sqrt{\frac{\mu_1 q_1}{\mu_2 q_2}}$$

and we get

$$H(\mu) = \mu_1 q_1 + \mu_2 q_2 - 2\sqrt{\mu_1 q_1 \mu_2 q_2} \tag{6}$$

(b) $E = \{1, 2, 3\}$ and jumps $1 \leftrightarrow 2$ and $2 \leftrightarrow 3$. Now we find

$$H(\mu) = \mu_1 q_1 + \mu_2 q_2 + \mu_3 q_3 - 2\sqrt{\mu_1 q_1 \mu_2 q_{21}} - 2\sqrt{\mu_3 q_3 \mu_2 q_{23}}$$

4 Overflow probability and action intergral

Consider a cycle in which the fluid model overflows the buffer. Write

$$S_t = \int_0^t r(X_s)\, ds$$

for the total input up to time t. Then $\frac{1}{t} S_t = \sum_{i \in E} L_t(i) r(i)$ and hence

$$\begin{aligned} P(\frac{1}{t} S_t \approx s) &\approx \sum_{\mu \in \mathcal{M}(E) : <\mu, r> = s} P(L_t \approx \mu) \\ &= K(t) e^{-t \inf_\mu H(\mu)} \end{aligned} \tag{7}$$

where the infimum is over probability measures on E s.t.

$$<\mu, r> = \sum_{i \in E} \mu_i r(i) = s.$$

What is the most likely way that the buffer content overflows? The approach to obtain a large deviation expression is similar to the one described in section 2. Set

$$S_t^B = \frac{1}{B} S_{tB}$$

and define paths ϕ by

$$\begin{cases} \phi : [0,T] \to [0,1] \text{ is absolute continuous} \\ \phi(0) = 0 \\ \phi(T) = 1 \\ 0 < \phi(t) < 1 \text{ else} \end{cases}$$

Notice that these paths generate overflow cycles by the relation

$$S_t^B = \phi(t) + ct, 0 \leq t \leq T \Leftrightarrow J_t = B\phi(t/B), 0 \leq t \leq BT$$

which means that there is a buffer overflow at time BT.

We shall calculate

$$P(S_t^B \approx \phi(t) + ct, 0 \leq t \leq T)$$

Let $\Delta > 0$, $n \in \{0, 1, \ldots, [T/\Delta]\}$ and define processes $\{S_t^{n,\Delta,B}\}_{0 \leq t \leq B\Delta}$ by

$$S_t^{n,\Delta,B} = S_{nB\Delta+t} - S_{nB\Delta}$$

$\{S_t^{n,\Delta,B}\}$ on $[0, B\Delta]$ has the same properties as $\{S_t\}$, e.g. according to (7)

$$\begin{aligned} & P(\frac{1}{B} S_{B\Delta}^{n,\Delta,B} \approx s) \\ = \; & P(\frac{1}{B\Delta} \int_0^{B\Delta} r(X_t) dt \approx s/\Delta) \\ = \; & K(B) e^{-B\Delta \inf_{<\mu,r>=s/\Delta} H(\mu)} \end{aligned} \tag{8}$$

for large B. Let ϕ a path and set $s_n(\Delta) = \phi((n+1)\Delta) - \phi(n\Delta) + c\Delta$ for $n = 0, 1, \ldots, [T/\Delta]$. Then for large B

$$\begin{aligned} & P(S_{n\Delta}^B \approx \phi(n\Delta) + cn\Delta, n = 0, 1, \ldots) \\ = \; & P(S_{(n+1)\Delta}^B - S_{n\Delta}^B \approx s_n(\Delta), n = 0, 1, \ldots) \\ = \; & P(\frac{1}{B} S_{B\Delta}^{n,\Delta,B} \approx s_n(\Delta), n = 0, 1, \ldots) \\ = \; & \prod_{n=0}^{[T/\Delta]} K(B) e^{-B\Delta \inf_{<\mu,r>=s_n(\Delta)/\Delta} H(\mu)} \\ = \; & K(B) e^{-B\Delta \sum_n \inf_{<\mu,r>=s_n(\Delta)/\Delta} H(\mu)} \end{aligned} \tag{9}$$

The third equality above needs some justification, since we are dealing with Markov chains and not with independent increments.

$$\begin{aligned}
& P\left(\frac{1}{B}S_{B\Delta}^{n,\Delta,B} \approx s_n(\Delta), n=0,1\right) \\
&= P\left(\frac{1}{B}\int_0^{B\Delta} r(X_t)dt \approx s_0(\Delta), \frac{1}{B}\int_{B\Delta}^{2B\Delta} r(X_t)dt \approx s_1(\Delta)\right) \\
&= \int_{-\infty}^{\infty} P\left(\frac{1}{B}\int_0^{B\Delta} r(X_t)dt \approx s_0(\Delta), \frac{1}{B}\int_{B\Delta}^{2B\Delta} r(X_t)dt \approx s_1(\Delta), X_{B\Delta} = dx\right) \\
&= \int_{-\infty}^{\infty} P\left(\frac{1}{B}\int_0^{B\Delta} r(X_t)dt \approx s_0(\Delta)\right) P\left(\frac{1}{B}\int_{B\Delta}^{2B\Delta} r(X_t)dt \approx s_1(\Delta) | X_{B\Delta} = x\right) \\
&\quad P\left(X_{B\Delta} = dx | \frac{1}{B}\int_0^{B\Delta} r(X_t)dt \approx s_0(\Delta)\right)
\end{aligned}$$

Now we note that the large deviations result stated in (8) is independent of initial state distribution, i.e. for large B

$$\begin{aligned}
& P(\frac{1}{B}\int_{B\Delta}^{2B\Delta} r(X_t)dt \approx s_1(\Delta) | X_{B\Delta} = x) \\
&= P(\frac{1}{B}S_{B\Delta}^{1,\Delta,B} \approx s_1(\Delta) | X_0 = x) \\
&= K(B)e^{-B\Delta \inf_{<\mu,r>=s_1(\Delta)/\Delta} H(\mu)}
\end{aligned}$$

Because

$$\int_{-\infty}^{\infty} P(X_{B\Delta} = dx | \frac{1}{B}\int_0^{B\Delta} r(X_t)dt \approx s_0(\Delta)) = 1$$

we can show that the product in (9) results by extending these arguments for $n = 0, 1, \ldots [T/\Delta]$.

Now let $\Delta \downarrow 0$ in (9):

$$\begin{aligned}
& P(S_t^B \approx \phi(t) + ct, 0 \le t \le T) \\
&= \lim_{\Delta \downarrow 0} P(S_{n\Delta}^B \approx \phi(n\Delta) + cn\Delta, n = 0, 1, \ldots) \\
&= K(B)e^{-B\int_0^T \inf_{<\mu,r>=\phi'(t)+c} H(\mu)dt}
\end{aligned}$$

with $H(\mu)$ given in (5). Finally

$$P(\{J_t\} \text{ hits } B \text{ before returning to } 0)$$

$$\begin{aligned}
&\approx \sum_T \sum_\phi P(S_t^B \approx \phi(t) + ct, 0 \le t \le T) \\
&= K(B) e^{-B \inf_T \inf_\phi \int_0^T \inf_{<\mu,r>=\phi'(t)+c} H(\mu) dt} \\
&= K(B) e^{-B \inf_T T \inf_{<\mu,r>=1/T+c} H(\mu) dt} \qquad (10)
\end{aligned}$$

where we have used the convexity of $H(\mu)$ to get the last equality:

Lemma:

$$\inf_\phi \int_0^T \inf_{<\mu,r>=\phi'(t)+c} H(\mu) dt = T \inf_{<\mu,r>=1/T+c} H(\mu)$$

Proof: Consider the piecewise linear path ϕ defined by

$$\phi'(t) = \begin{cases} \frac{\epsilon}{T}, & t \in [0, \tau] \\ \frac{1-\epsilon}{T-\tau}, & t \in [\tau, T] \end{cases}$$

for some $\epsilon \in (0,1)$ and $\tau \in (0,T)$. Furthermore let

$$\begin{aligned}
\mu^{(1)} &= \text{arginf}_{<\mu,r>=\epsilon/T+c} H(\mu) \\
\mu^{(2)} &= \text{arginf}_{<\mu,r>=(1-\epsilon)/(T-\tau)+c} H(\mu) \\
\mu^{(0)} &= \frac{\tau}{T}\mu^{(1)} + \frac{T-\tau}{T}\mu^{(2)}
\end{aligned}$$

Then $< \mu^{(0)}, r >= \frac{1}{T} + c$, so

$$\begin{aligned}
\inf_{<\mu,r>=1/T+c} H(\mu) &\le H(\mu^{(0)}) \\
&= H\left(\frac{\tau}{T}\mu^{(1)} + \frac{T-\tau}{T}\mu^{(2)}\right) \\
&\le \frac{\tau}{T} H(\mu^{(1)}) + \frac{T-\tau}{T} H(\mu^{(2)}) \\
&= \int_0^T \inf_{<\mu,r>=\phi'(t)+c} H(\mu) dt
\end{aligned}$$

Approximating any path ϕ by a piecewise linear one we get the desired result.

Remark: Let $T_1 = \text{arginf}_T\{TH(\mu) :< \mu, r >= 1/T + c\}$. Then J_t will hit the boundary B at $T_B = BT_1$.

Examples:

(a) $E = \{1,2\}$. $H(\mu)$ is given in (6) in section 3. Define the constraint function

$$G(T,\mu) = <\mu, r> - \frac{1}{T} - c$$

and the Lagrangian function

$$F(T,\mu,f) = TH(\mu) - fG(T,\mu), \ f \in R$$

Since $\mu_2 = 1 - \mu_1$, the dependence on μ in these functions in through μ_1 only. We find optimal T and μ by solving the set of fixed point equations

$$\begin{cases} \frac{\partial F}{\partial T} = 0 \\ \frac{\partial F}{\partial \mu_1} = 0 \\ G(T,\mu) = 0 \end{cases}$$

That is

$$\begin{cases} T = \frac{1}{<\mu, r> - c} \\ f = T^2 H(\mu) \\ \frac{1 - 2\mu_1}{\sqrt{\mu_1(1-\mu_1)}} = \frac{T(q_1 - q_2) - f(r(1) - r(2))}{T\sqrt{q_1 q_2}} \end{cases} \tag{11}$$

Denoting α for the right hand side of the last equation, we can rewrite to

$$\mu_1 = \frac{1}{2} - \frac{\alpha}{2}\sqrt{\frac{1}{4+\alpha^2}}$$

5 Multi-input

Let $\{X_t^\nu : \nu = 1,2,\ldots,N\}$ be N identical independent Markov chains, each on the state space $E = \{1,2,\ldots,d\}$, with transition rates $\{q_{ij} : i,j \in E\}$ and with stationary distribution π. Each chain generates an input rate $r(X_t)$ into the fluid model described before. When there were no buffer boundaries, the contents at time t would be given by the free process

$$\bar{J}_t = \sum_{\nu=1}^{N} \int_0^t r(X_t^\nu)dt - ct$$

The actual process J_t is bounded below by 0 and above by B. A cycle of the process starts at 0 and ends either at 0 or at B (whichever comes first). During a cycle the two processes behave statistically the same. We will concentrate on cycles that overflow (i.e. end at B). Of course we assume the stability criterion $< \pi, r >< \frac{c}{N}$.

Let L_t^ν be the empirical distribution of chain ν and $\mu^\nu \in \mathcal{M}(E)$ (probability measures on E). Then according to large deviations

$$P(L_t^\nu \approx \mu^\nu) = K(t)e^{-tH(\mu^\nu)}$$

(for large t) with K and H described in section 3. Because of independence we get

$$\begin{aligned} P(L_t^\nu \approx \mu^\nu, \nu = 1,2,\ldots,N) &= K(t) \prod_{\nu=1}^{N} e^{-tH(\mu^\nu)} \\ &= K(t)e^{-t\sum_{\nu=1}^{N} H(\mu^\nu)} \end{aligned}$$

Write $\mu = (\mu^1, \ldots, \mu^N)$ and

$$H_N(\mu) = \sum_{\nu=1}^{N} H(\mu^\nu)$$

The total input upto time t is

$$\begin{aligned} S_t &= \sum_{\nu=1}^{N} \int_0^t r(X_t^\nu)dt \\ &= t\sum_{\nu=1}^{N} < L_t^\nu, r > \end{aligned}$$

So

$$\begin{aligned} & P(\frac{1}{t}S_t \approx s) \\ &= \sum_{\mu^1,\ldots,\mu^N : <\sum_{\nu=1}^{N} \mu^\nu, r> \approx s} P(L_t^\nu \approx \mu^\nu, \nu = 1,2,\ldots,N) \\ &= K(t)e^{-t\inf_\mu H_N(\mu)} \end{aligned}$$

where the infimum is over all probability measures $\mu = (\mu^1, \ldots, \mu^N)$ s.t.

$$< \sum_{\nu=1}^{N} \mu^\nu, r > \approx s$$

Following the same reasoning as is section 4 we come to the conclusion

$$\begin{aligned} & P(\{J_t\} \text{ hits } B \text{ before returning to } 0) \\ = \; & K(B)e^{-B \inf_T \inf_\mu T H_N(\mu)} \end{aligned} \tag{12}$$

where the first infimum is over $T > 0$ and the second over all probability measures $\mu = (\mu^1, \ldots, \mu^N)$ s.t.

$$< \sum_{\nu=1}^{N} \mu^\nu, r > \frac{1}{T} + c$$

The following lemma is needed to derive a scaling property for the overflow probability.

Lemma

$$\begin{aligned} & \inf\{\sum_{\nu=1}^{N} H(\mu^\nu) : \mu^1, \ldots, \mu^N \in \mathcal{M}(E), < \sum_{\nu=1}^{N} \mu^\nu, r >= \frac{1}{T} + c\} \\ = \; & \sum_{\nu=1}^{N} \inf\{H(\mu^\nu) : \mu^\nu \in \mathcal{M}(E), < \mu^\nu, r >= \frac{1}{NT} + \frac{c}{N}\}. \end{aligned} \tag{13}$$

Proof (i) LHS $\leq$ RHS is clear.

(ii) By the lemma of Fatou we have that

$$\begin{aligned} & \text{RHS} \\ \leq \; & \inf\{\sum_{\nu=1}^{N} H(\mu^\nu) : \mu^\nu \in \mathcal{M}(E), < \mu^\nu, r >= \frac{1}{NT} + \frac{c}{N}, \nu = 1, \ldots, N\}. \end{aligned} \tag{14}$$

Let $\rho = (\rho^1, \ldots, \rho^N) = \arg\inf$ LHS(13). Then there are numbers $\alpha^\nu \geq 0$ for $\nu = 1, 2, \ldots, N$ with $\sum_{\nu=1}^{N} \alpha^\nu = 1$ such that

$$< \rho^\nu, r >= \alpha^\nu(\frac{1}{N} + c).$$

Define $\mu = (\mu^1, \ldots, \mu^N)$ by

$$\mu^\nu = \frac{1}{N} \sum_{\nu=1}^{N} \rho^\nu \in \mathcal{M}(E)$$

for all ν. Because

$$< \mu^\nu, r >= \frac{1}{NT} + \frac{c}{N}$$

for all ν, μ is feasible for the RHS(14). Furthermore, because H is convex,

$$\frac{1}{N} \sum_{\nu=1}^{N} H(\mu^\nu) = H(\mu^1) = H(\sum_{\nu=1}^{N} \frac{1}{N} \rho^\nu) \leq \sum_{\nu=1}^{N} \frac{1}{N} H(\rho^\nu).$$

This inequality gives RHS(14) $\leq$ LHS(13).

Q.E.D.

We notice that the

$$\text{RHS}(13) \ = N \inf\{H(\mu) : \mu \in \mathcal{M}(E), < \mu, r >= \frac{1}{TN} + \frac{c}{N}\}.$$

Substituting into (12) we get

$$\begin{aligned} & Pr(\{J_t\} \text{ hits } B \text{ before returning to } 0) \\ = \ & K(B) e^{-B \inf_T TN \inf_{<\mu,r>=\frac{1}{TN}+\frac{c}{N}} H(\mu)} \\ = \ & e^{-B \inf_T T \inf_{<\mu,r>=\frac{1}{T}+\frac{c}{N}} H(\mu)}, \end{aligned} \tag{15}$$

by a change of variable.

6 Two different inputs

In this section we have 2 Markov independently generating inputs but we suppose that the two Markov chains ($\{X_t^1\}$ and $\{X_t^2\}$) may have different statistical characteristics. We do assume that they have both the same state space E as before and the same deterministic input rate function r. Write Q^ν for the matrix of transition rates of chain ν and π^ν for the associated stationary distribution. Assume the stability criterion $< \pi^1 + \pi^2, r >< c$.

Let H^ν be the level 2 large deviation rate function as derived in section 3 and expressed in (5), associated with the chain X_t^ν and set $H_2(\mu^1, \mu^2) = H^1(\mu^1) + H^2(\mu^2)$ for probability measures μ^1, μ^2 on E. Similar to the result in section 5 we can derive that

$$\begin{aligned} & P(\{J_t\} \text{ hits } B \text{ before returning to } 0) \\ = \ & K(B) e^{-B \inf_T \inf_{\mu^1,\mu^2} T H_2(\mu^1, \mu^2)} \end{aligned} \tag{16}$$

where the first infimum is over $T > 0$ and the second over all probability measures μ^1, μ^2 s.t.

$$< \mu^1 + \mu^2, r >= \frac{1}{T} + c$$

The scaling property as was derived in the case of similar inputs, is not present here.

Examples:

Consider the two-state Markov chains again: $E = \{1,2\}$. Just as in section 4 we can define the constraint function

$$G(T,\mu^1,\mu^2) = <\mu^1+\mu^2, r> - \frac{1}{T} - c$$

and the Lagrangian function

$$F(T,\mu^1,\mu^2,f) = TH_2(\mu^1,\mu^2) - fG(T,\mu^1,\mu^2)$$

where dependence on μ^ν is through μ_1^ν only. Setting the deravitives to zero we get the fixed point equations

$$\begin{cases} T = \dfrac{1}{<\mu^1+\mu^2,r> - c} \\ f = T^2 H(\mu) \\ \mu_1^1 = \frac{1}{2} - \frac{\alpha^1}{2}\sqrt{\dfrac{1}{4+(\alpha^1)^2}} \\ \mu_1^2 = \frac{1}{2} - \frac{\alpha^2}{2}\sqrt{\dfrac{1}{4+(\alpha^2)^2}} \end{cases} \tag{17}$$

where

$$\alpha^\nu = \frac{T(q_1^\nu - q_2^\nu) - f(r(1)-r(2))}{T\sqrt{q_1^\nu q_2^\nu}}$$

7 Numerical & simulation results

A) For one 2-state Markov generating source we have found that the probability of overflow in a cycle equals $\Phi = K(B)e^{-Bs_0}$, that the time to reach the buffer limit B in an overflow cycle is T_B while the chain behaves according to the empirical distribution μ. The relation is

$$<\mu, r> = \frac{1}{T_1} + c$$

Set

$$
\begin{aligned}
q_1 &= 10 \\
q_2 &= 30 \\
r_1 &= 400 \\
r_2 &= 4000 \\
c &= 1500
\end{aligned}
$$

State 2 is the bursty state. On the average there are 7.5 bursts per second. State 1 is a quiet state although some generation of fluid occurs. The average input per second $< \pi, r >= 1300$ (note that $\pi_1 = 0.75, \pi_2 = 0.25$), hence the load is 0.866667. When 1 unit of fluid stands for 1 message cell, the peak arrival rate (i.e. during bursts) is 4000 * 53 * 8 = 1.696 Mbits per second.

Applying section 2 we get

$$
\begin{aligned}
s_0 &= 2.909091 \;\; 10^{-3} \\
T_1 &= 5.0 \;\; 10^{-3} \\
\mu_1 &= 0.638889
\end{aligned}
$$

But section 4 leads to

$$
\begin{aligned}
s_0 &= 2.909091 \;\; 10^{-3} \\
T_1 &= 4.490909 \;\; 10^{-3} \\
\mu_1 &= 0.632591
\end{aligned}
$$

From simulations (95 % confidence intervals, runs for B = 2000, 2500 and 3000) we obtain

$$
\begin{aligned}
s_0 &= 3.236068 \;\; 10^{-3} \\
T_1 &= 3.529654 \;\; 10^{-3} \\
\mu_1 &= 0.612579
\end{aligned}
$$

Here the value of s_0 is not obtained directly but calculated from simulated values for Φ by setting $K(B) \equiv K$ and then solving for K and s_0.

Since these simulations require long computing time, because they involve a rare event, it would be interesting to find a quick simulation. [Cottrel] and [Parekh] describe how to do that, viz. using an exponential change of measure. Let F_ξ be the probability distribution function of the i.i.d. ξ_n's introduced in section 2. Construct i.i.d. $\tilde{\xi}_n$'s with distribution

$$F_{\tilde{\xi}}(x) = \int_{-\infty}^{x} e^{s_0 t} dF_\xi(t)$$

i.e. the mean $\mu_{\tilde{\xi}}$ equals the optimal slope $\frac{1}{T_0}$ (s_0 and T_0 as given in (3) in section 2). The Laplace transform satisfies

$$M_{\tilde{\xi}}(s) = M_\xi(s + s_0)$$

An implementation of these $\tilde{\xi}_n$'s is realized by a Markov fluid model where the 2-state Markov chain $\{\tilde{X}_t\}$ has transition rates

$$\begin{aligned}\tilde{q}_1 &= q_2 \frac{c - r(1)}{r(2) - c} \\ \tilde{q}_2 &= q_1 \frac{r(2) - c}{c - r(1)}\end{aligned}$$

The chain has stationary distribution $\tilde{\pi}$ for which $< \tilde{\pi}, r >> c$. The rate functions r and c are the same. When this chain would follow its stationary distribution, the buffer would reach level 1 at time

$$\tilde{T}_1 = \frac{1}{< \tilde{\pi}, r > - c}$$

It is not difficult to show that $\tilde{\pi}$ and $\tilde{T}_1$ satisfy the fixed point equations (11) in section 4, i.e. they solve the optimization of (10).

When we actually perform these quick simulations applied to the example given we find again a time T_1 to reach level 1 while the chain follows an empirical distribution μ:

$$\begin{aligned}T_1 &= 3.336556 \ 10^{-3} \\ \mu_1 &= 0.607485\end{aligned}$$

B) For 2 sources we have the following results. First 2 identical sources as the one given above, and with r as before, $c = 3000$ so that the load is again 0.866667. Then according to section 6

$$
\begin{aligned}
s_0 &= 2.909091 \ 10^{-3} \\
T_1 &= 2.245455 \ 10^{-3} \\
\mu_1^1 &= 0.632591 \\
\mu_1^2 &= 0.632591
\end{aligned}
$$

as expected from the results of section 5. Simulations yield

$$
\begin{aligned}
T_1 &= 1.800341 \ 10^{-3} \\
\mu_1^1 &= 0.614026 \\
\mu_1^2 &= 0.617043
\end{aligned}
$$

Secondly, two different sources. Suppose $\{X_t^1\}$ is a Markov chain as above and $\{X_t^2\}$ a chain (on $\{1,2\}$ with $q_1 = 1, q_2 = 3$. The rate functions r and c are the same as above. This second chain has on the average only 0.75 bursts per second. From the equations (17) we find

$$
\begin{aligned}
s_0 &= 4.989825 \ 10^{-4} \\
T_1 &= 2.188005 \ 10^{-3} \\
\mu_1^1 &= 0.732589 \\
\mu_1^2 &= 0.529346
\end{aligned}
$$

whereas simulations yield

$$
\begin{aligned}
T_1 &= 1.496032 \ 10^{-3} \\
\mu_1^1 &= 0.727979 \\
\mu_1^2 &= 0.472639
\end{aligned}
$$

Acknowledgement

I thank Jean Walrand for introducing me to the subject and guiding me by helpful discussions.

References

[Cottrel] M. Cottrel, J-C. Fort & G. Malgouyres, "Large deviations and rare events in the study of stochastic algorithms", IEEE Trans. Autom. Control (9), 1983, p. 907-920.

[Courcoubetis] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand & R. Weber, " Admission control and routing in ATM Networks using inferences from measured buffered occupancy", Presented at the ORSA/TIMS special interest meeting on Applied Probability, Monterey, Ca., January 9-11, 1991.

[Donsker] M.D. Donsker & S.R.S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time", part I: Comm. Pure Appl. Math. (28), 1975, p. 1-47.

[Ellis] R.S. Ellis, "Entropy, large deviations and statistical mechanics" (Springer), 1985.

[Ellis] R.S. Ellis, "Large deviations for a general class of random vectors", Ann. Prob. (12), 1988, p.1-12.

[Mitra] D. Mitra, D. Anick & M.M. Sondhi, "Stochastic thory of a data-handling system with multiple sources", Bell Sys. Tech. J. (61), 1982, p. 1871-1894.

[Parekh] S.Parekh & J. Walrand, "A quick simulation method for excessive backlog in network of queues", IEEE Trans. Autom. Control (34), 1989, p. 54-66.

[Varadhan] S.R.S. Varadhan, "Large deviations and applications", SIAM (1984).

[Ventsel] A.D. Ventsel, "Rough limit theorems on large deviations for Markov stochastic processes", part II: Theory Prob. Appl. (21), 1976, p. 499-512.

[Weiss] A. Weiss, "A new technique of analyzing large traffic systems", Adv. Appl. Prob. (18), 1986, p. 506-532.