Network Design and Control Using On-Off and Multi-level Source Traffic Models with Long-Tailed Distributions

N. G. Duffield and W. Whitt
AT&T Laboratories,
Room {A175, A117}, 180 Park Avenue, Florham Park, NJ 07932-0971, USA
E-mail {duffield,wow}@research.att.com

Abstract

A major challenge for designing and controlling emerging high-speed integrated-services communication networks is to develop methods for analyzing more realistic source traffic models that are consistent with recent traffic measurements. We consider the familiar on-off source traffic model, but we allow the on and off times to have long-tailed distributions such as the Pareto and Weibull distributions. We also consider a more general traffic model in which the required bandwidth (arrival rate) as a function of time for each source is represented as the sum of two stochastic processes: (1) a macroscopic (longer-time-scale) level process and (2) a microscopic (shorter-time-scale) within-level variation process. We let the level process be a finite-state semi-Markov process (SMP), allowing general (possibly long-tailed) level holding-time distributions, and we let the within-level variation process be a zero-mean piecewise-stationary process. However, the fine structure of the within-level variation process turns out not to matter in our analysis. We make design and control decisions based on the likelihood that aggregate demand (the input rate from a set of sources) will exceed capacity (the maximum possible output rate), using a specification of the sources and their source traffic models to predict demand. This approach to model analysis avoids the customary queueing detail (focus on buffer content and overflow).

We propose using transient analysis, exploiting asymptotics associated with multiplexing a large number of sources. A conditional law of large numbers supports approximating the future aggregate demand conditional on current state information by its conditional mean value, conditional on the levels and elapsed times in levels of the sources. The conditional aggregate mean can be expressed compactly in terms of its Laplace transform and efficiently calculated by numerical transform inversion. We supply further approximations which enable the rapid calculation of the conditional mean without using numerical transform inversion.

As an application in control, we formulate an integer program in which to evaluate the outcomes of decision within a given cost structure. As an application in design, we describe a simple approximate scheme in which a network link can be dimensioned to achieve multiplexing gain while keeping overdemand sufficiently rare and short.

Keywords: source traffic model, admission control, congestion control, overload control, transient analysis, deterministic fluid approximation, long-tailed distributions, Laplace transforms, numerical transform inversion, statistical multiplexing, value of information

1. Introduction

In order to help design and control the emerging high-speed communication networks, we want source traffic models that can be both realistically fit to data and successfully analyzed. Many recent traffic measurements have shown that network traffic is quite complex, exhibiting phenomena such as long-tailed probability distributions and long-range dependence; e.g., see Cáceres, Danzig, Jamin and Mitzel [3], Leland, Taqqu, Willinger, and Wilson [18], Paxson and Floyd [19], Crovella and Bestavros [6] and Feldmann [11]. In fact, the long-tailed distributions may be the cause of all these phenomena, because they tend to cause long-range dependence and (asymptotic) self-similarity. For example, the input and buffer content processes associated with an on-off source exhibit long-range dependence when the on and off times have long-tailed probability distributions; e.g. see Section 8. Long-tailed distributions are known to cause self-similarity in models of (asymptotically largely) aggregated traffic; see Willinger, Taqqu, Sherman and Wilson [22].

In this paper we propose a way to analyze the performance of a network with multiple on-off sources and more general multi-level sources in which the on-time, off-time, and level-holding-time distributions are allowed to have long tails. To do so we must go beyond the familiar Markovian analysis. To achieve the required analyzability with this added model complexity, we propose a simplified kind of analysis. In particular, we avoid the customary queueing detail (and its focus on buffer content and overflow) and instead concentrate on the probability that aggregate demand (the input rate from a collection of sources) exceeds capacity (the maximum possible output rate) at any time. (This approach also can generate approximations describing loss and delay with finite capacity; e.g., Section 5 of [10].)

Motivation for considering on-off and multi-level models as source models comes from traces of frame-sizes generated by certain video encoders; e.g., see Grasse, Frater and Arnold [14]. Shifts between levels in mean frame size appear to arise from scene changes in the video, with the distribution of scene durations long-tailed. Indeed, the expectation that scene durations will have long-tailed distributions in one of the motivations behind the Renegotiated Constant Bit Rate (RCBR) proposal of Grossglauser, Keshav and Tse [15].

A key to being able to analyze the system with such complex sources represented by our traffic model is exploiting asymptotics associated with multiplexing a large number of sources. The ever-increasing network bandwidth implies that more and more sources will be able to be multiplexed. This gain is generally possible, even in the presence of long-tailed distributions and more general long-range dependence; e.g. see Duffield [8, 9] for demonstration of the multiplexing gains available for long-range dependent traffic in shared buffers. As the scale increases, describing the detailed behavior of all sources becomes prohibitively difficult, but fortunately it becomes easier to describe the aggregate, because the large numbers produce statistical regularity. As the size increases, the aggregate demand can be well described by laws of large numbers, central limit theorems and large deviation principles.

We have in mind two problems: first, we want to do capacity planning and, second, we want to do real-time connection admission control and congestion control. In both cases, we want to determine whether any candidate capacity is adequate to meet the aggregate demand associated with a set of sources. In both cases, we represent the aggregate demand simply as the sum of the bandwidth requirements of all sources. In forming this sum, we regard the bandwidth processes of the different sources as probabilistically independent.

The performance analysis for capacity planning is coarser, involving a longer time scale, so that it may be appropriate to do a steady-state analysis. However, when we consider connection admission control and congestion control, we suggest focusing on a shorter time scale. We are still concerned with the relatively long time scale of connections, or scene times in video, instead of the shorter time scales of cells or bursts, but admission control and congestion control are sufficiently short-term that we propose focusing on the *transient* behavior of the aggregate demand process. In fact, even for capacity planning the transient analysis plays an important role. The transient analysis determines how long it takes to recover from rare congestion events. One application we have in mind is that of networks carrying rate-adaptive traffic. In this case the bandwidth process could represent the ideal demand of a source, even though it is able to function when allocated somewhat less bandwidth. So from the point of view of quality, excursion of aggregate bandwidth demand above available supply may be acceptable in the short term, but one would want to dimension the link so that such excursions are sufficiently short-lived. In this or other contexts, if the recovery time from overload is relatively long, then we may elect to provide extra capacity (or reduce demand) so that overload becomes less likely. The main contribution of this paper is to show how the transient analysis for design and control can be done.

Our approach is interesting for on-off and multi-level source models, but with little extra effort we can treat a wider class. The general model we consider has two components. The bandwidth demand of each source as a function of time $\{B(t) : t \geq 0\}$, is represented as the sum of two stochastic processes: (1) a macroscopic (longer-time-scale) *level process* $\{L(t) : t \geq 0\}$ and (2) a microscopic (shorter-time-scale) *within-level variation process* $\{W(t) : t \geq 0\}$, i.e.,

$$B(t) = L(t) + W(t) \ , \quad t \geq 0 \ . \tag{1.1}$$

We let the macroscopic level process $\{L(t) : t \geq 0\}$ be a *semi-Markov process* (SMP) as in Chapter 10 of Çinlar [5]; i.e., the level process is constant except for jumps, with the jump transitions governed by a Markov process, while the level holding times (times between jumps) are allowed to have general distributions depending on the originating level and the next level. Given a transition from level j to level k, the holding time in level j has cumulative distribution function (cdf) F_{jk}. Conditional on the sequence of successive levels, the holding times are mutually independent. To obtain models compatible with traffic measurements cited earlier, we allow the holding-time cdf's F_{jk} to have long tails.

We assume that the within-level variation process $\{W(t) : t \geq 0\}$ is a zero-mean piecewise-stationary process. During each holding-time interval in a level, the within-level variation process is an independent segment of a zero-mean stationary process, with the distribution of each segment being allowed to depend on the level. We allow the distribution of the stationary process segment to depend on the level, because it is natural for the variation about any level to vary from level to level.

We will require only a limited characterization of the within-level variation process; it turns out that the fine structure of the within-level variation process plays no role in our analysis. Indeed, that is one of our main conclusions. In several examples of processes which we envisage modeling by these methods, there will only be the level process. First, the level process may be some smoothed functional of a raw bandwidth process. This is the case with algorithms for smoothing stored video by converting into piecewise constant rate segments in some optimal manner subject to buffering and delay constraints; see Salehi, Zhang, Kurose and Towsley [20]. With such smoothing, the input rate will directly be a level process as we have defined it. Alternatively, the level process may stem from rate reservation over the period between level-shifts, rather than the bandwidth actually used. This would be the case for RCBR previously mentioned. In this situation we act as if the reservation level is the actual demand, and thus again have a level process.

The remainder of this paper is devoted to showing how to do transient analysis with the source traffic model. As in our previous paper [10], we suggest focusing on the future time-dependent mean conditional on the present state. The present state of each level process consists of the level and age (elapsed holding time in that level). Because of the anticipated large number of sources, the actual bandwidth process should be closely approximated by its mean, by the law of large numbers (LLN). As in [10], the conditional mean can be thought of as a deterministic fluid approximation; e.g., see Chen and Mandelbaum [4]. Since the within-level variation process has mean zero, the within-level variation process has no effect upon this conditional mean. Hence, the conditional mean of the aggregate bandwidth process is just the sum of the conditional means of the component level processes. Unlike the more elementary $M/G/\infty$ model considered in [10], however, the conditional mean here is not available in closed form.

The outline of the paper is as follows. In Section 2, we show that the Laplace transform of the mean of the transient conditional aggregate demand can be expressed concisely. This is the main enabling result for the remainder of the paper. The conditional mean itself can be very efficiently computed by numerically inverting its Laplace transform. To carry out the inversion, we use the Fourier-series method in Abate and Whitt [1] (the algorithm Euler exploiting Euler summation), although alternative methods could be used. The inversion algorithm is remarkably fast; computation for each time point corresponds simply to a sum of fifty terms. We provide numerical examples in Examples 3.1, 4.2 and 7.1. Example 7.1 is of special interest, because the level-holding-time distribution there is Pareto.

In Section 3 we show that in some cases we can avoid the inversion entirely and treat much larger models. We can avoid the inversion if we can assume that the level holding times are relatively long compared to the times of interest for control. Then we can apply a single-transition approximation, which amounts to assuming that the Markov chain is absorbing after one transition. Then the conditional mean is directly expressible in terms of the level holding-time distributions. In Section 4 we show the value of having more detailed state information, specifically the current ages of levels. With long-tailed distributions, a large elapsed holding time means that a large remaining holding time is very likely; e.g. see Section 8 of [10] for background, and Harchol-Balter and Downey [16] for an application in another setting. In Section 5 we illustrate the use of state information for network control. In Section 6 we turn to applications to capacity planning. The idea is to approximate the probability of an excursion in demand using Chernoff bounds and other large deviation approximations, then chart its recovery to a target acceptable level using the results on transience. Interestingly, the time to recover from excursions sufficiently close to the target level depends on the level durations essentially only through their mean. Correspondingly, the conditional mean demand relaxes linearly from it excursion, at least approximately so, for sufficiently small times. If the chance for a larger excursion is negligible (as determined by the large deviation approximation mentioned) then this simple description may suffice. An example is given in Section 7. In Section 8 we show how long-range dependence in the level process arises through long-tailed level-holding-time distributions.

2. Transient Analysis

Approximation by the conditional mean bandwidth. In this paper, the state information upon which we condition will be either the current level of each source or the current level and age (current time) in that level of each source. No state from the within level process is assumed. Conditional on that state information, we can compute the probability that each source will be in each possible level at any time in the future, from which we can calculate the conditional mean and variance of the aggregate required bandwidth by adding.

The Lindeberg-Feller central limit theorem (CLT) for non-identically-distributed summands can be applied to generate a normal approximation characterized by the conditional mean and conditional variance; see p. 262 of Feller [13]. For the normal approximation we must check that the aggregate is not dominated by only a few sources.

Let $B(t)$ denote the (random) aggregate required bandwidth at time t, and let $I(0)$ denote the (known deterministic) state information at time 0. Let $(B(t)|I(0))$ be a random variable with the conditional distribution of $B(t)$ given the information $I(0)$. By the CLT, the normalized random variable

$$\frac{(B(t)|I(0)) - E(B(t)|I(0))}{\sqrt{Var(B(t)|I(0))}} \tag{2.1}$$

is approximately normally distributed with mean 0 and variance 1 when the number of sources is suitably large.

Since the conditional mean alone tends to be very descriptive, we use the approximation

$$(B(t)|I(0)) \approx E(B(t)|I(0)) \ , \tag{2.2}$$

which could be justified by a (weaker) law of large numbers instead of the CLT. We will show that the conditional mean in (2.2) can be efficiently computed, so that it can be used for real-time control. From (2.1), we see that the error in the approximation (2.2) is approximately characterized by the conditional standard deviation $\sqrt{Var(B(t)|I(0))}$. We also will show how to compute this conditional standard deviation, although the required computation is more difficult. If there are n sources that have roughly equal rates, then the conditional standard deviation will be $O(\sqrt{n})$, while the conditional mean is $O(n)$.

Given that our approximation is the conditional mean, and given that our state information does not include the state of the within-level variation process, the within-level variation process plays no role because it has zero mean. Let i index the source. Since the required bandwidths need not have integer values, we index the level by the integer j, $1 \leq j \leq J_i$, and indicate the associated required bandwidths in the level by b_j^i. Hence, instead of (1.1), the required bandwidth for source i can be expressed as

$$B^i(t) = b^i_{L^i(t)} + W_{L_i(t)}(t) \ , \quad t \geq 0 \ . \tag{2.3}$$

Let $P^{(i)}_{jk}(t|x)$ be the probability that the source-i level process is in level k at time t given that at time 0 it was in level j and had been so for a period x (i.e., the age or elapsed level holding time at time 0 is x). If $\mathbf{j} \equiv (j_1, \ldots, j_n)$ and $\mathbf{x} \equiv (x_1, \ldots, x_n)$ are the vectors of levels and ages of the n source level processes at time 0, then the *state information* is $I(0) = (\mathbf{j}, \mathbf{x}) = (j_1, \ldots, j_n; x_1, \ldots, x_n)$ and the *conditional aggregate mean* is

$$E(B(t)|I(0)) \equiv M(t|\mathbf{j}, \mathbf{x}) = \sum_{i=1}^{n} \sum_{k_i=1}^{J_i} P^{(i)}_{j_i k_i}(t|x_i) b^i_{k_i} \ . \tag{2.4}$$

From (2.4), we see that we need to compute the conditional distribution of the level, i.e., the probabilities $P^{(i)}_{jk}(t|x)$, for each source i. In this section we show how to compute these conditional probabilities. We consider a single source and assume that its required bandwidth process is a semi-Markov process (SMP). (We now have no within-level variation process.) We now omit the superscript i. Let $L(t)$ and $B(t)$ be the level and required bandwidth, respectively, at time t as in (2.3). The process $\{L(t) : t \geq 0\}$ is assumed to be an SMP, while the process $\{B(t) : t \geq 0\}$ is a function of an SMP, i.e., $B(t) = b_{L(t)}$, where b_j is the required bandwidth in level j. If $b_j \neq b_k$ for $j \neq k$, then $\{B(t) : t \geq 0\}$ itself is an SMP, but if $b_j = b_k$ for some $j \neq k$, then in general $\{B(t) : t \geq 0\}$ is not an SMP.

Laplace transform analysis. Let $A(t)$ be the age of the level holding time at time t. We are interested in calculating

$$P_{jk}(t|x) \equiv P(L(t) = k | L(0) = j, A(0) = x) \tag{2.5}$$

as a function of j, k, x, and t. The state information at time 0 is the pair (j, x). Let P be the transition matrix of the DTMC governing level transitions and let $F_{jk}(t)$ be the holding-time cdf given that there is a transition from level j to level k. For simplicity, we assume that

$F^c_{jk}(t) = 1 - F_{jk}(t) > 0$ for all j, k, and t, so that all positive x can be level holding times. Let $P(t|x)$ be the matrix with elements $P_{jk}(t|x)$ and let $\hat{P}(s|x)$ be the Laplace transform (LT) of $P(t|x)$, i.e., the matrix with elements that are the Laplace transforms of $P_{jk}(t|x)$ with respect to time, i.e.,

$$\hat{P}_{jk}(s|x) = \int_0^\infty e^{-st} P_{jk}(t|x) dt. \tag{2.6}$$

We will derive an expression for $\hat{P}(s|x)$. For this purpose, let G_j be the holding-time cdf in level j, unconditional on the next level, i.e.,

$$G_j(x) = \sum_k P_{jk} F_{jk}(x) \ . \tag{2.7}$$

For any cdf G, let G^c be the complementary cdf, i.e. $G^c(x) = 1 - G(x)$. Also let

$$H_{jk}(t|x) = \frac{P_{jk} F_{jk}(t+x)}{G^c_j(x)} \quad \text{and} \quad G_j(t|x) = \sum_k H_{jk}(t|x) \tag{2.8}$$

for G_j in (2.7). Then let $\hat{h}_{jk}(s|x)$ and $\hat{g}_j(s|x)$ be the associated Laplace-Stieltjes transforms (LSTs), i.e.,

$$\hat{h}_{jk}(s|x) = \int_0^\infty e^{-st} dH_{jk}(t|x) \quad \text{and} \quad \hat{g}_j(s|x) = \int_0^\infty e^{-st} dG_j(t|x) \ . \tag{2.9}$$

Let $\hat{h}(s|x)$ be the matrix with elements $\hat{h}_{jk}(s|x)$. Let $\hat{q}(s)$ be the matrix with elements $\hat{Q}_{jk}(s)$, where

$$Q_{jk}(t) = P_{jk} F_{jk}(t) \quad \text{and} \quad \hat{q}_{jk}(s) = \int_0^\infty e^{-st} dQ_{jk}(t) \ . \tag{2.10}$$

Let $\hat{D}(s|x)$ and $\hat{D}(s)$ be the diagonal matrices with diagonal elements

$$\hat{D}_{jj}(s|x) \equiv [1 - \hat{g}_j(s|x)]/s, \qquad \hat{D}_{jj}(s) \equiv [1 - \hat{g}_j(s)]/s \ , \tag{2.11}$$

where $\hat{g}_j(s)$ is the LST of the cdf G_j in (2.7).

Theorem 2.1 *The transient probabilities for a single SMP source have the matrix of Laplace transforms*

$$\hat{P}(s|x) = \hat{D}(s|x) + \hat{h}(s|x)\hat{P}(s|0) \ , \tag{2.12}$$

where

$$\hat{P}(s|0) = (I - \hat{q}(s))^{-1} \hat{D}(s). \tag{2.13}$$

Proof. In the time domain, condition on the first transition. For $j \neq k$,

$$P_{jk}(t|x) = \sum_l \int_0^t dH_{jl}(u|x) P_{lk}(t-u|0) \ ,$$

so that

$$\hat{P}_{jk}(s|x) = \sum_l \hat{h}_{jl}(s|x) \hat{P}_{lk}(s|0),$$

while

$$P_{jj}(t|x) = G_j^c(t|x) + \sum_l \int_0^t dH_{jl}(u|x) P_{lj}(t-u|0) ,$$

so that

$$\hat{P}_{jj}(s|x) = \frac{1-\hat{g}_j(s|x)}{s} + \sum_l h_{jl}(s|x)\hat{P}_{lj}(s|0).$$

Hence, (2.12) holds. However, $P(t|0)$ satisfies a Markov renewal equation, as in Section 10.3 of Çinlar [5], i.e., for $j \neq k$,

$$P_{jk}(t|0) = \sum_l \int_0^t dQ_{jl}(u) P_{lk}(t-u|0), \quad \text{and} \quad P_{jj}(t|0) = G_j^c(t) + \sum_l \int_0^\infty dQ_{jl}(u) P_{lj}(t-u|0) ,$$

so that

$$P(t|0) = D(t) + Q(t) * P(t|0)$$

where $*$ denotes convolution, and (2.13) holds.

To compute the LT $\hat{P}(s|0)$, we only need the LSTs $\hat{f}_{jk}(s)$ and $\hat{g}_j(s)$ associated with the basic holding-time cdf's F_{jk} and G_j. However, to compute $\hat{P}(s|x)$, we also need to compute $\hat{D}(s|x)$ and $\hat{h}(s|x)$, which require computing the LSTs of the *conditional* cdf's $H_{jk}(t|x)$ and $G_j(t|x)$ in (2.8). The LSTs of these conditional cdf's are often easy to obtain because some cdf's inherit their structure upon conditioning. For example, this is true for phase-type, hyperexponential and Pareto distributions. Moreover, other cdf's can be approximated by hyperexponential or phase-type cdf's; see Asmussen, Nerman and Olsson [2] and Feldmann and Whitt [12]. If the number of levels is not too large, then it will not be difficult to compute the required matrix inverse $(I - q(s))^{-1}$ for all required s. Note that, because of the probability structure, the inverse is well defined for all complex s with $\mathrm{Re}(s) > 0$. To illustrate with an important simple example, we next give the explicit formula for an on-off source.

Example 2.1. Suppose that we have an on-off source, i.e., so that there are two states with transition probabilities $P_{12} = P_{21} = 1$ and holding time cdf's G_1 and G_2. From (2.8) or by direct calculation,

$$\begin{aligned} \hat{P}(s|0) \equiv \begin{pmatrix} \hat{P}_{11}(s|0) & \hat{P}_{12}(s|0) \\ \hat{P}_{21}(s|0) & \hat{P}_{22}(s|0) \end{pmatrix} &= (I - \hat{q}(s))^{-1}\hat{D}(s) \\ &= \frac{1}{s(1-\hat{g}_1(s)\hat{g}_2(s))} \begin{pmatrix} 1-\hat{g}_1(s) & \hat{g}_1(s)(1-\hat{g}_2(s)) \\ \hat{g}_2(s)(1-\hat{g}_1(s)) & 1-\hat{g}_2(s) \end{pmatrix} . \end{aligned} \tag{2.14}$$

Suppose that the levels are labeled so that the initial level is 1. Note that all transitions from level 1 are to level 2. Hence when considering the matrix $\hat{h}(s|x)$ in (2.9) it suffices to consider only the element $\hat{h}_{12}(s|x)$. Since

$$H_{12}^c(t|x) = G_1^c(t|x) = \frac{G_1^c(t+x)}{G_1^c(x)} , \quad \text{then} \quad \hat{h}_{12}(s|x) = \hat{g}_1(s|x) = \int_0^\infty e^{-st} dG_1(t|x) . \tag{2.15}$$

Since $P_{11}(t|x) = 1 - P_{12}(t|x)$, it suffices to calculate only $P_{12}(t|x)$. Hence, in this context

$$\hat{P}_{12}(s|x) = \frac{\hat{g}_1(s|x)(1-\hat{g}_2(s))}{s(1-\hat{g}_1(s)\hat{g}_2(s))} \ . \tag{2.16}$$

We now determine the mean, second moment, and variance of the bandwidth process of a general multi-level source as a function of time. It is elementary that

$$m_j(t|x) = E(B(t)|L(0)=j, A(0)=x) = \sum_k P_{jk}(t|x)b_k \tag{2.17}$$

$$s_j(t|x) = E(B(t)^2|L(0)=j, A(0)=x) = \sum_k P_{jk}(t|x)b_k^2 \tag{2.18}$$

$$v_j(t|x) = \mathrm{Var}(B(t)|L(0)=j, A(0)=x) = s_j(t|x) - m_j(t|x)^2 \ . \tag{2.19}$$

We can calculate $m_j(t|x)$ and $s_j(t|x)$ by a single inversion of their Laplace transforms, using

$$\hat{m}_j(s|x) \equiv \int_0^\infty e^{-st} m_j(t|x)dt = \sum_k P_{jk}(s|x)b_k \quad \text{and} \quad \hat{s}_j(s|x) = \sum_k \hat{P}_{jk}(s|x)b_k^2. \tag{2.20}$$

To properly account for the within-level variation process when it is present, we should add its variance in level j, say $w_j(t,x)$, to $v_j(t,x)$, but we need make no change to the mean $M_j(t,x)$. We anticipate that $W_j(t,x)$ will tend to be much less than $v_j(t,x)$ so that $w_j(t,x)$ can be omitted; but it could be included.

Finally, we consider the aggregate bandwidth associated with n sources. Again let a superscript i index the sources. The conditional aggregate mean and variance are

$$M(t|\mathbf{j},\mathbf{x}) \equiv E(B(t)|I(0)) = \sum_{i=1}^n m_{j_i}^i(t|x_i) \tag{2.21}$$

and

$$V(t|\mathbf{j},\mathbf{x}) \equiv Var(B(t)|I(0)) = \sum_{i=1}^n [v_{j_i}^i(t|x_i) + w_{j_i}^i(t|x_i)], \tag{2.22}$$

where $\mathbf{j} = (j_1, \ldots, j_n)$ is the vector of levels and $\mathbf{x} = (x_1, \ldots, x_n)$ is the vector of elapsed holding times for the n sources with the single-source means and variances as in (2.17) and (2.19).

It is significant that we can calculate the conditional aggregate mean at any time t by performing a single inversion. We summarize this elementary but important consequence as a theorem.

Theorem 2.2 *The Laplace transform of the n-source conditional mean aggregate required bandwidth as a function of time is*

$$\hat{M}(s|\mathbf{j},\mathbf{x}) \equiv \int_0^\infty e^{-st} M(t|\mathbf{j},\mathbf{x})dt = \sum_{i=1}^n \sum_{k_i=1}^{J_i} \hat{P}_{j_i k_i}^{(i)}(s|x_i)b_{k_i} \ , \tag{2.23}$$

where the single-source transform $\hat{P}_{j_i k_i}^{(i)}(s|x_i)$ is given in Theorem 2.1.

Unlike for the aggregate mean, for the aggregate variance we evidently need to perform n separate inversions to calculate $v^i_{j_i}(t|x_i)$ for each i and then add to calculate $V(t|\mathbf{j},\mathbf{x})$ in (2.22). (We assume that the within-level variances $w^i_{j_i}(t|x_i)$, if included, are specified directly). Hence, we suggest calculating only the conditional mean on line for control, and occasionally calculating the conditional variance off line to evaluate the accuracy of the conditional mean.

3. The One-Transition and Two-Transition Approximations

The most complicated part of the conditional aggregate mean transform $\hat{M}(s|\mathbf{j},\mathbf{x})$ in (2.23) is the matrix inverse $(I - \hat{q}(s))^{-1}$ in the transform of the single-source transition probability in (2.13). Since the matrix inverse calculation can be a computational burden when the number of levels is large, it is natural to seek approximations which avoid this matrix inverse. We describe such approximations in this section.

The matrix inverse $(I - q(s))^{-1}$ is a compact representation for the series $\sum_{n=0}^{\infty} q(s)^n$. For $P(t|x)$, it captures the possibility of any number of transitions up to time t. However, if the levels are relatively long in the time scale relevant for control, then the mean for times t of interest will only be significantly affected by a very few transitions. Indeed, often only a single transition need be considered.

The single-transition approximation is obtained by making the Markov chain absorbing after one transition. Hence, the single-transition approximation is simply

$$P_{jk}(t|x) \approx H_{jk}(t|x) \ , \quad j \neq k \ , \quad \text{and} \quad P_{jj}(t|x) \approx G^c_j(t|x) + H_{jj}(t|x) \tag{3.1}$$

for $H_{jk}(t|x)$ in (2.8) and $G_j(t|x)$ in (2.8). From (3.1) we see that no inversion is needed.

Alternatively, we can develop a two-transition approximation. (Extensions to higher numbers are straightforward.) Modifying the proof of Theorem 2.1 in a straightforward manner, we obtain

$$P_{jk}(t|x) = \int_0^t G^c_k(t-u)dH_{jk}(u|x) + \sum_{\ell} \int_0^t P_{\ell k} F_{lk}(t-u) dH_{j\ell}(u|x) \tag{3.2}$$

for $j \neq k$ and

$$P_{jj}(t|x) = G^c_j(t|x) + \sum_{\ell} \int_0^t P_{\ell j} F_{lj}(t-u) dH_{j\ell}(u|x) \ . \tag{3.3}$$

Expressed in the form of transforms, (3.2) and (3.3) become

$$\hat{P}_{jk}(s|x) = \hat{h}_{jk}(s|x)\frac{(1-\hat{g}_k(s))}{s} + \sum_{\ell} \hat{h}_{j\ell}(s|x) P_{\ell k} \frac{\hat{f}_{lk}(s)}{s} \tag{3.4}$$

for $j \neq k$ and

$$\hat{P}_{jj}(s|x) = \frac{1-\hat{g}_j(s|x)}{s} + \sum_{\ell} \hat{h}_{j\ell}(s|x) P_{lj} \frac{\hat{f}_{lj}(s)}{s} \ . \tag{3.5}$$

Numerical inversion can easily be applied with (3.4) and (3.5). However, since the time-domain formulas (3.2) and (3.3) involve single convolution integrals, numerical computation of (3.2) and

(3.3) in the time domain is also a feasible alternative. Moreover, if the underlying distributions have special structure, then the integrals in (3.2) and (3.3) can be calculated analytically. For example, analytical integration can easily be done when all holding-time distributions are hyperexponential.

Example 3.1. To illustrate how the two approximations compare to the exact conditional mean, we give a numerical example. We consider a single source with four levels. The transitions move cyclically through the levels: $P_{12} = P_{23} = P_{34} = P_{41} = 1$. The level holding-time ccdf's are:

$$\begin{aligned} G_1^c(t) &= 0.5e^{-10t} + 0.5e^{-0.1t} \;, \quad G_2^c(t) = e^{-0.1t} \;, \\ G_3^c(t) &= 0.9e^{-2t} + 0.1e^{-0.1t} \;, \quad G_4^c(t) = e^{-0.1t} \;, \; t \geq 0 \;. \end{aligned}$$

The level bandwidths are $b_1 = b_3 = 100$ and $b_2 = b_4 = 0$. Suppose that we start in level 1 with an age of 8. From the form of $G_1^c(t)$, we see that the conditional level-1 holding-time ccdf $G_1^c(t|x)$ is then approximately $e^{-0.1t}$. Hence the first two mean level holding times are approximately 10. Hence we might consider the one-transition and two-transition approximations in the interval $[0, 10]$. The two approximations are compared to the exact value of the conditional mean in Figure 1. (All are computed by numerical transform inversion.) The approximations are very good up to $t = 1$ or 2, but they start to degrade by $t = 10$. The two-transition approximation performs not so well for larger t because the actual holding time in level 3 is likely to be quite short. More generally, our experience is that the one-transition and two-transition approximations tend to perform quite satisfactorily if the mean level holding times in the first few levels are substantially *larger* than the times t of interest. In this example the approximations are quite good in the interval $[0, 1]$.

4. The Value of Information

We can use the source model to investigate the value of information. We can consider how prediction is improved when we condition on, first, only the level and, second, on both level and age. The reference case is the steady-state mean

$$M = \sum_{i=1}^{n} m^i \quad \text{and } m^i = \sum b_j^i p_j^i \;, \tag{4.1}$$

where p_j^i is the steady-state probability, i.e.,

$$p_j^i = \frac{\pi_j^i m(G_j^i)}{\sum_k \pi_k^i m(G_k^i)} \;, \tag{4.2}$$

with π^i the steady-state vector of the Markov chain P^i ($\pi^i = \pi^i P^i$) and $m(G_j^i)$ the mean of G_j^i for G_j in (2.7), for all sources i. With the steady-state mean, there is no conditioning. Section 2 gives the formula for conditioning on both level and age. Now we give the formulas conditioning only on the level; i.e., we condition on the level, assuming that we are in steady-state. We omit the i superscript. Then the age in level j has the stationary-excess cdf

$$G_{je}(t) = \frac{1}{m(G_j)} \int_0^t G_j^c(u)du \;, \quad t \geq 0 \;. \tag{4.3}$$

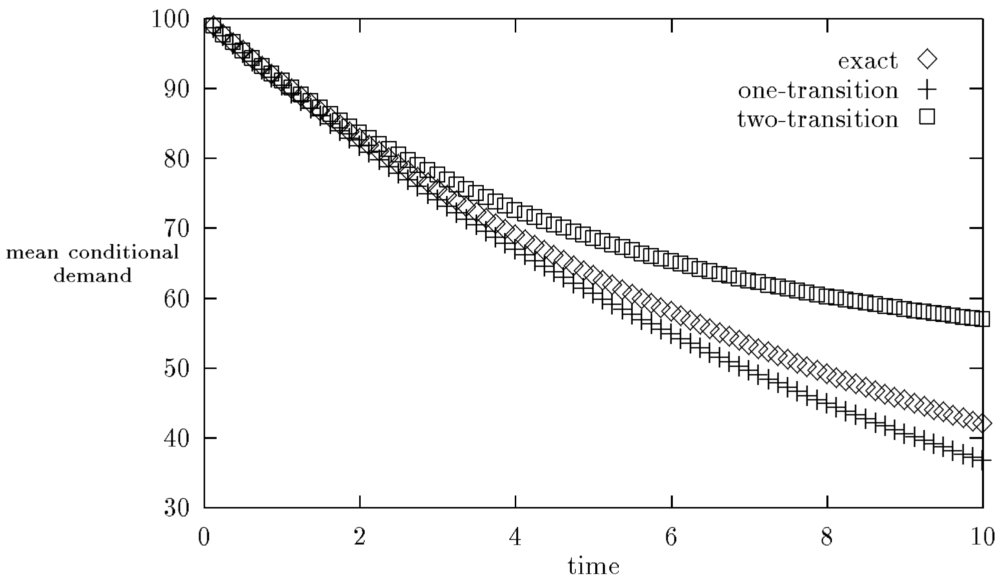

Figure 1: A comparison of the one-transition and two-transition approximations with the exact conditional mean aggregate demand in Example 3.1.

Let $P_{jk}(t)$ be the probability of being in level k at time t conditional on being in level j in steady state at time 0. Let $\hat{P}_{jk}(s)$ be its Laplace transform. Let $m_j(t)$ be the conditional steady-state mean given level j at time 0 and let $\hat{m}_j(s)$ be its Laplace transform: Clearly

$$m_j(t) = \sum_{k=1}^{J} P_{jk}(t) b_k \quad \text{and} \quad \hat{m}_j(s) = \textstyle\sum_{k=1}^{J} \hat{P}_{jk}(s) b_k \ . \tag{4.4}$$

Hence, it suffices to calculate $\hat{P}_{jk}(s)$.

Theorem 4.1 *Assume that the level-holding-time cdf depends only on the originating level, i.e., $F_{jk}(t) = G_j(t)$. The steady-state transition probabilities conditional on the level for a single SMP source have the matrix of Laplace transforms*

$$\hat{P}(s) = \hat{D}_e(s) + \hat{g}_e(s)\hat{P}(s|0) \ , \tag{4.5}$$

where $\hat{P}(s|0)$ is the matrix in (2.13), $\hat{g}_e(s)$ is the matrix with elements

$$\hat{g}_{ejk}(s) = P_{jk}\hat{g}_{je}(s) = P_{jk}\frac{(1-\hat{g}_j(s))}{sm(G_j)} \ , \tag{4.6}$$

$\hat{D}_e(s)$ is the diagonal matrix with diagonal elements

$$\hat{D}_{ejj}(s) \equiv \frac{1-\hat{g}_{je}(s)}{s} = \frac{sm(G_j) - 1 + \hat{g}_j(s)}{s^2 m(G_j)} \ , \tag{4.7}$$

$\hat{g}_j(s)$ is the level-j holding-time LST and $\hat{g}_{je}(s)$ is the LST of its stationary-excess cdf in (4.3).

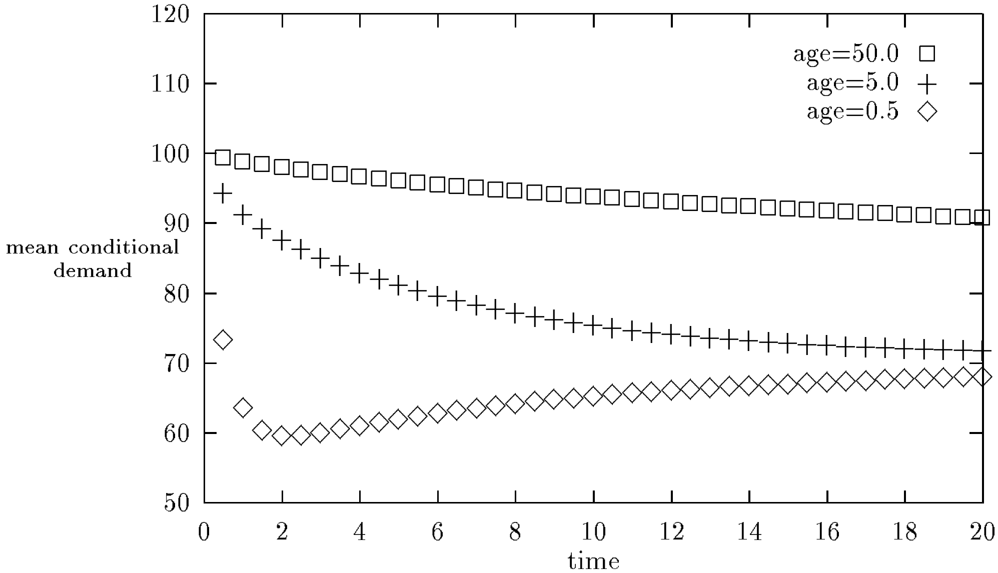

Figure 2: The conditional mean aggregate demand as a function of the age of the holding time in level 1 for Example 4.2.

Proof. Modify the proof of Theorem 2.1, inserting $P_{jl}G_{je}(t)$ for $H_{jl}(t|x)$ and $G^c_{je}(t)$ for $G^c_j(t|x)$.
Example 4.1. Consider the on-off source in Example 2.1. Paralleling (2.16), it suffices to calculate only $P_{12}(t)$. Its Laplace transform is

$$\hat{P}_{12}(s) = \frac{\hat{g}_{1e}(s)(1-\hat{g}_2(s))}{s(1-\hat{g}_1(s)\hat{g}_2(s))} \ . \tag{4.8}$$

Example 4.2. To show the value of knowing the age, consider an on-off source with holding-time ccdf's

$$G^c_1(t) = 0.01e^{-0.01t} + 0.1e^{-0.1t} + .89e^{-t}, \qquad G^c_2(t) = e^{-t} \ , \quad t \geq 0 \ . \tag{4.9}$$

Let the bandwidths be $b_1 = 100$ and $b_2 = 0$. Since $m(G_1) = 2.89$ and $m(G_2) = 1.00$, the steady-state mean is

$$EB(\infty) = \frac{100m(G_1)}{m(G_1)+m(G_2)} = 74.29 \ .$$

Let the initial level be 1. Since G_1 has an exponential component with mean 100, we anticipate the time to reach steady state to be between 100 and 1000. In Figure 2 we plot the conditional mean $m_1(t|x)$ for $x = 0.5$, 5.0 and 50.0, computed by numerical transform inversion. Figure 2 shows that the age plays a very important role.

5. Network Control

The source model makes it possible to investigate several different kinds of controls. As in Sections 4 and 9 of [10], we can consider rejecting new sources or removing existing sources. We can also consider changing the levels of existing sources. For example, on-off sources in the on state might be turned off. We also can consider rate controls corresponding to changing the bandwidths assigned to the levels. With each of these controls, we can calculate the resulting conditional mean aggregate bandwidth to evaluate the performance of the control.

To illustrate, suppose that we wish to consider which of n candidate sources to serve over the time interval $[0,T]$. Suppose that source i earns a fixed revenue R_i plus a revenue rate r_i per unit of bandwidth per time. Also suppose that we want to keep the demand below a capacity c at all times. Then we can solve the following integer program. Let y_i be the decision variable, with $y_i = 1$ if source i is served and $y_i = 0$ if not. The integer program can be formulated as:

$$\max \sum_{i=1}^{n} y_i (R_i + r_i \int_0^T m_{j_i}^i (t|x_i) dt) \tag{5.1}$$

subject to

$$\sum_{i=1}^{n} y_i m_{j_i}^i (t_k|x_i) \le c \ , \quad 0 \le k \le K \ , \tag{5.2}$$

for a set of time points $0 = t_0 < t_1 < \ldots < t_K = T$. In (5.1) and (5.2) the level j_i and age x_i of source i are included because these are presumed to be known.

6. Recovery from Congestion in Steady State

For capacity planning, it is useful to consider the time required to recover from a high-congestion event, as well as the likelihood of the high-congestion event. The likelihood of a high-congestion event in steady state can be estimated using a large deviation principle (LDP) approximation. The well-known Chernoff bound (e.g. see [7]) gives an upper bound to the stationary tail probabilities of the aggregate level process, even for finitely many sources. By Chebychev's inequality, for all $\theta > 0$,

$$\mathbf{P}[L(t) \ge x] \le e^{-\theta x} \mathbf{E}[e^{\theta L(t)}] = e^{-\theta x} \prod_i \mathbf{E}[e^{\theta L_i(t)}] = e^{-\theta x} \prod_i \sum_j p_j^i e^{\theta b_j^i}, \tag{6.1}$$

where b_j^i is the required bandwidth and p_j^i is the steady-state probability of level j in source i, as in (4.2). Thus,

$$\mathbf{P}[L(t) \ge x] \le e^{-I(x)}, \quad \text{where} \quad I(x) = \sup_{\theta > 0} \left(\theta x - \sum_i \log \sum_j p_j^i e^{\theta b_j^i} \right) \ . \tag{6.2}$$

It can be shown [7] that such bounds are asymptotically tight (have a *large deviation* limit) as the number of sources increases provided the spectrum of behavior of individual sources is sufficiently regular, yielding the exponential approximation

$$P(L(t) \ge x) \approx e^{-I(x)} \ . \tag{6.3}$$

Finding the rate function I will in general require numerical solution of the variational expression (6.2). It can be shown that the RHS of (6.2) is a concave function of θ, and under mild conditions it is differentiable also. Hence the supremum is achieved at the unique solution θ to the Euler-Lagrange equation

$$x = \sum_i \left(\frac{\sum_j b^i_j p^i_j e^{\theta b^i_j}}{\sum_{j'} p^i_{j'} e^{\theta b^i_{j'}}} \right) . \tag{6.4}$$

Generally, it is not difficult to numerically determine the supremum in (6.2) by location of the solution to (6.4).

Example 6.1 In special cases the variational problem can be solved explicitly. This is possible in the case of n homogeneous two-level sources. Here we have $b^i_j = b_j$ with $j \in \{1,2\}$, $0 \le b_1 < b_2$ and $p_1 + p_2 = 1$. For this case,

$$\begin{aligned} I(x) &= n \sup_{\theta > 0} \left(\theta x - \log(p_1 e^{\theta b_1} + p_2 e^{\theta b_2}) \right) & (6.5) \\ &= n \left(\frac{x}{b_2 - b_1} \log y(x) - \log \left(p_1 y(x)^{b_1/(b_2-b_1)} + p_2 y(x)^{b_2/(b_2-b_1)} \right) \right) & (6.6) \end{aligned}$$

with $y(x) = p_1(x - b_1)/(p_2(b_2 - x))$ for $b_1 \le x \le b_2$ and $I(x) = +\infty$ elsewhere.

We now show how to estimate the time to recover from the high-congestion event, where the high-congestion event is a large initial bandwidth x. We understand recovery to occur when the aggregate bandwidth is again less than or equal to the capacity c. In applications, we suggest examining the function of aggregate bandwidth giving *both* the probability of reaching that level and the recovery time from that level to assess whether or not capacity is adequate to meet demand.

We assume that recovery occurs when the aggregate bandwidth drops below a level c, where $x > c > m$, with x being the initial level and m being the steady-state mean. Given that we know the current level of each level process, we know that the remaining holding time (and also the age) is distributed according to the level-holding-time stationary-excess distribution in (4.3). We use the LDP to approximate the conditional distribution of the level process for each source (in steady state). The idea is to perform the appropriate change of measure (tilting) corresponding to the rare event.

Given that $P(\sum_{i=1}^n B^i \ge x) \approx e^{-I(x)}$ for $I(x)$ in (6.2), the LDP approximation is

$$P(B^i = b^i_j | \sum_{i=1}^n B^i \ge x) \approx \bar{p}^i_j \equiv \frac{p^i_j e^{b^i_j \theta^*}}{\sum_{k=1}^{J_i} p^i_k e^{b^i_k \theta^*}} , \tag{6.7}$$

where θ^* yields the supremum in (6.2). Put another way, comparing (6.7) with (6.4) we see that θ^* is chosen to make the expectation of $\sum_i B_i$ equal to x under the distribution $\bar{p}$. In the homogeneous case, equality in (6.7) in the limit as the number of sources increases is due to the conditional limit theorem of Van Campenhout and Cover [21]. The limit can be extended to cover suitably regular heterogeneity in the b^i_j, e.g., finitely many types of source.

We thus approximate the conditional bandwidth process by

$$(B(t)|B(0) = x) \quad \approx \quad E(B(t)|B(0) = x) = \sum_{i=1}^n \sum_{j=1}^{J_i} \bar{p}^i_j \int_0^\infty E(B^i_j(t)|B^i_j(0) = y) dG^i_{je}(y)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{J_i} \bar{p}_j^i \sum_{k=1}^{J_i} b_k^i P_{jk}(t) \ , \tag{6.8}$$

for $\bar{p}_j^i$ in (4.3) which has Laplace transform

$$\sum_{i=1}^{n} \sum_{j=1}^{J_i} \bar{p}_j^i \sum_{k=1}^{J_i} b_k^i \hat{P}_{jk}(s) \ . \tag{6.9}$$

The Laplace transform $\hat{P}_{jk}(s)$ in (6.9) was derived in Theorem 4.1. We can numerically invert it to calculate the conditional mean as a function of time. We then can determine when $E(B(t)|B(0) = x)$ first falls below c. In general, this conditional mean need not be a decreasing function, so that care is needed in the definition, but we expect it to be decreasing for suitably small t because the initial point $B(0)$ is unusually high.

7. A Linear Approximation

Assuming that the relevant time is not too large, we might approximate the conditional mean bandwidth using a Taylor-series approximation

$$E(B(t)|B(0) = x) = x + tr(x) \ , \tag{7.1}$$

where

$$r(x) \ \equiv \ E(B'(0)|B(0) = x) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \bar{p}_j^i \sum_{k=1}^{J_i} b_k^i P'_{jk}(0) = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \bar{p}_j^i \sum_{k=1}^{J_i} P_{jk}(b_k^i - b_j^i)/m(G_j) \tag{7.2}$$

which has the advantage that no numerical inversion is required.

Suppose the service capacity is $c > E(B(0))$ and we condition on $B(0) > c$. If $r(B(0)) < 0$, we can use (7.1) to approximate the first time to return to c, the *recovery time*, by

$$\tau = (x - c)/r(x) \ . \tag{7.3}$$

Suppose in addition that B is reversible; this will happen if the matrix P is reversible. Then since both the residual lifetime and the current age have distribution F_{jke}, $E(B(-t) \mid B(0)) = E(B(t) \mid B(0))$. Consequently $B(0) = x$ is a local maximum, at $t = 0$, of $E(B(t) \mid B(0))$.

Now suppose that there are n independent sources. Then as in [10], it follows by use of an appropriate functional law of large numbers that, as $n \to \infty$ under regularity conditions, the stochastic paths of the B process converge to this mean path. Thus we can identify, asymptotically as $n \to \infty$, $t = 0$ as a hitting time for the level x. Thus, we can use (6.3) to approximate the probability of this hitting time and τ in (7.3) to approximate the associated recovery time.

Example 7.1. Consider homogeneous two-level sources, i.e., $j \in \{1, 2\}$, $0 \leq b_1 < b_2$, $p_1 + p_2 = 1$ with mean lifetimes m_1, m_2, and $P_{11} = P_{22} = 1 - P_{12} = 1 - P_{21} = 0$. With n sources and

# sources initially in each level		initial total demand	steady-state probability of x	recovery time τ		
				linear approx.	inversion	
					exponential duration	Pareto duration
n_1	n_2	x	$e^{-I(x)}$			
25	25	150	7.5×10^{-4}	0	0	0
22	28	162	1.9×10^{-5}	0.15	0.16	0.19
19	31	174	2.4×10^{-7}	0.24	0.29	0.41
16	34	186	1.4×10^{-9}	0.31	0.41	0.64
13	37	198	3.5×10^{-12}	0.37	0.50	0.91

Table 1: Homogeneous two-level sources. Approximate hitting probabilities of aggregate demand x, together with recovery time τ of mean from x, by linear approximation, and exact for (i) exponential duration, and (ii) Pareto duration of higher level; see Example 7.1.

$B(0) = x$ we can calculate the $\bar{p}_j$ in (6.7) directly from the relation $x = n(\bar{p}_1 b_1 + \bar{p}_2 b_2)$ for $x \in \{nb_1, (n-1)b_1 + b_2, \ldots, b_1 + (n-1)b_2, nb_2\}$. Then

$$r(x) \quad = \quad n\bar{p}_1(b_2 - b_1)/m_1 + \bar{p}_2(b_1 - b_2)/m_2 = (nb_2 - x)/m_1 - (x - nb_1)/m_2 \ . \qquad (7.4)$$

As a concrete example, we let $b_1 = 1$, $b_2 = 5$, $m_1 = 3$, $m_2 = 1$, giving a mean bandwidth per source of $(m_1 b_1 + m_2 b_2)/(m_1 + m_2) = 2$. We also let $n = 50$ and $c = 150$. The parameters of the example were chosen as a caricature of video traffic on an OC3 link: take b_i, x, c in Mb/sec, m_i in seconds.

We present in Table 1 some values of n_1, n_2, the number of sources in each level for a given x, the approximate probability $e^{-I(x)}$ of demand exceeding x (using I from (6.5)), and τ. For comparison we give also the *exact* recovery time for the mean, calculated by using numerical transform inversion methods [1], for particular models of level durations with the same means: (i) both level durations exponentially distributed; (ii) lower level exponential, upper level duration Pareto with exponent 1.5, and hence ccdf $G^c(x) = (1 + 2x)^{-3/2}$ in order to give mean $m_2 = 1$. The Pareto density $g(x) = a(1+x)^{-(1+a)}$ has Laplace transform $ae^s s^a \Gamma(-a, s)$, where $\Gamma(a, z)$ is the incomplete gamma function

$$\Gamma(a, z) = \int_z^\infty t^{a-1} e^{-t} dt \ . \qquad (7.5)$$

Hence, the required transform values for the Pareto distribution are readily computable.

In Figure 3 we display the evolution of the conditioned mean in the linear approximation, and for the two distributions above with the same mean. As should be expected, the linear approximation is more accurate when the initial level x is closer to the capacity c. The linear approximation also behaves worse for the Pareto high-level durations than for the exponential high-level durations. The linear approximation tends to consistently provide a lower bound on the true recovery time for the mean. Even though the linear-approximation estimate of the recovery time diverges from the true mean computed by numerical inversion as the hitting level x increases, the probability of such high x can be very small. Even the largest errors in predicted recovery times in Table 1 are

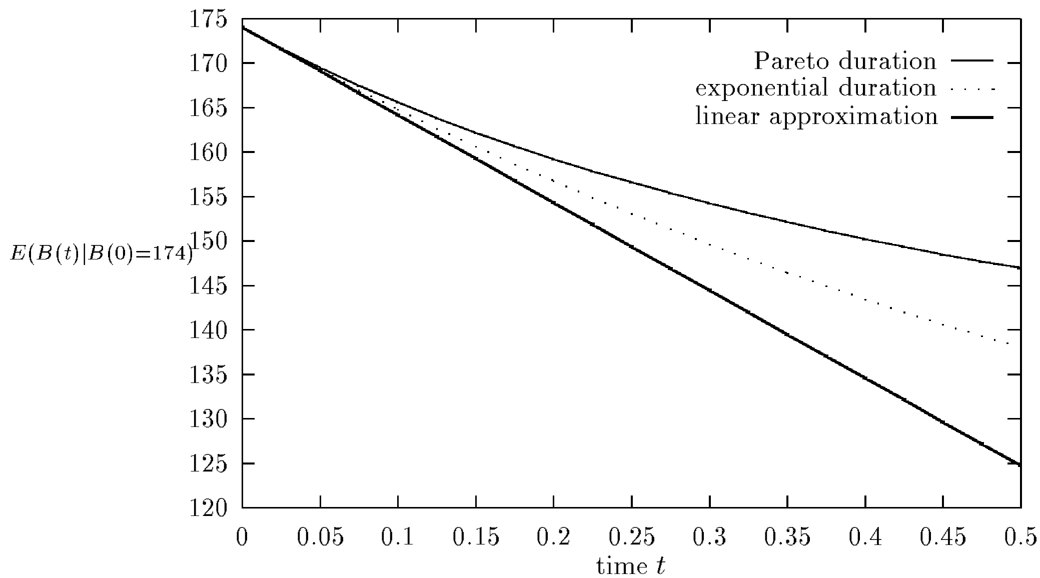

Figure 3: Recovery curves for two-level sources: linear approximation and numerical transform inversion with exponential and Pareto durations in the higher level; see Example 7.1.

within one order of magnitude, and so might be regarded as suitable approximations. ¿From our experiments, we conclude that the linear approximation is a convenient rough approximation, but that the numerical inversion yields greater accuracy.

Simple criteria for link dimensioning. A key point is the two-dimensional characterization of rare congestion events in terms of likelihood and recovery time. To further show how this perspective can be exploited, we plot in Figure 4, for various offered loads, the approximate probability $e^{-I(x(\tau))}$ of a demand at least $x(\tau)$ as a function of τ, where $\tau = (c - x(\tau))/r(x(\tau))$; i.e., $x(\tau)$ is the demand from which the recovery time to the level c is τ, using the linear approximation. Figure 4 shows that the two criteria together impose more constraints on what sets of sources are acceptable. Expressed differently, for the *same* probability of occurrence, rare congestion events can have very *different* recovery times.

8. Covariance Structure

Useful characterizations of the aggregate and single-source bandwidth processes are their (auto)covariance functions. The covariance function may help in evaluating the fitting. We now show that we can effectively compute the covariance function for our traffic source model.

Let $\{B(t) : t \geq 0\}$ and $\{B^i(t) : t \geq 0\}$ be stationary versions of the aggregate and source-i bandwidth processes, respectively. Assuming that the single-source bandwidth processes are mutually independent, the covariance function of the aggregate bandwidth process is the sum of

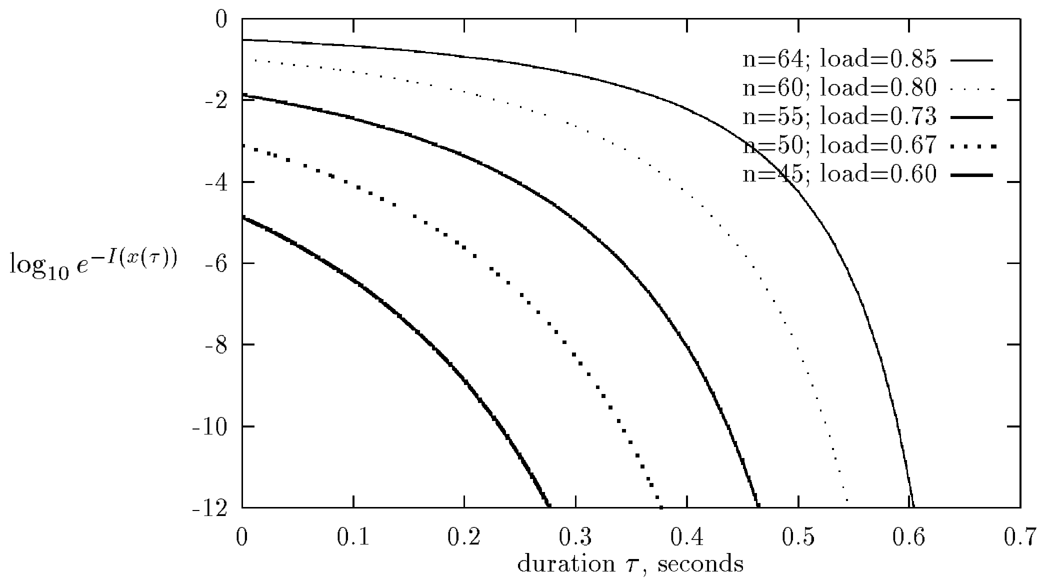

Figure 4: Design criteria: estimated probability of overdemand of at least duration τ, for various offered loads; see Section 7.

the single-source covariance functions, i.e.,

$$R(t) \equiv \mathrm{Cov}(B(0), B(t)) = \sum_{i=1}^{n} \mathrm{Cov}(B^i(0), B^i(t)) \ . \tag{8.1}$$

Hence, it suffices to focus on a single source, and we do, henceforth dropping the superscript i.

In general,

$$R(t) = S(t) - m^2, \tag{8.2}$$

where the steady-state mean m is as in (4.1) and (4.2) and

$$S(t) \equiv EB(0)B(t) = \int_0^\infty dx \sum_{j=1}^{J} b_j p_j g_{je}(x) \sum_{k=1}^{J} b_k P_{jk}(t|x) \tag{8.3}$$

$$+ \sum_{j=1}^{J} p_j G_{je}^c(t) Cov(W_j(0), W_j(t)) \ , \tag{8.4}$$

where $g_{je}(x) = G_j^c(x)/m(G_j)$ is the density of G_{je}, and the second term captures the effect of the within-level variation process. In (8.3) b_j is the bandwidth in level j, p_j is the steady-state probability of level j, $g_{je}(x) = G_j^c(x)/m(G_j)$ with G_j the level-j holding-time cdf and $m(G_j)$ its mean, and $P_{jk}(t|x)$ is the transition probability, whose matrix of Laplace transforms is given in Theorem 3.1 of [10]. We can thus calculate $S(t)$ by numerically inverting its Laplace transform

$$\hat{S}(s) \equiv \int_0^\infty e^{-st} S(t) dt = \sum_{j=1}^{J} \frac{b_j p_j}{m(G_j)} \int_0^\infty G_j^c(x) \sum_{k=1}^{J} b_k \hat{P}_{jk}(s|x) dx$$

$$+\sum_{j=1}^{J} p_j \int_0^\infty e^{-st} G_{je}^c(t) Cov(W_j(0), W_j(t)) dt \ . \tag{8.5}$$

To treat the second term in (8.5), we can assume an approximate functional form for the covariance of the within-level variation process $W(t)$. For example, if

$$Cov(W_j(0), W_j(t)) \approx \sigma_j^2 e^{-\eta_j t} \ , \quad t \geq 0 \ , \tag{8.6}$$

then

$$\begin{aligned}
&\sum_{j=1}^{J} p_j \int_0^\infty e^{-st} G_{je}^c(t) Cov(W_j(0), W_j(t)) dt \\
&\qquad \approx \sum_{j=1}^{J} p_j \sigma_j^2 \int_0^\infty e^{-(s+\eta_j)t} G_{je}^c(t) dt \\
&\qquad = \sum_{j=1}^{J} p_j \sigma_j^2 \frac{(1-\hat{g}_j(s+\eta_j))}{(s+\eta_j) m(G_j)} \ .
\end{aligned} \tag{8.7}$$

Thus, with approximation (8.6), we have a closed-form expression for the second term of the transform $\hat{S}(s)$ in (8.5). For each required s in $\hat{S}(s)$, we need to perform one numerical integration in the first term of (8.5), after calculating the integrand as a function of x.

A major role is played by the asymptotic variance $\int_0^\infty R(t)dt$. For example, the heavy-traffic approximation for the workload process in a queue with arrival process $\int_0^t B(u)du, \ t \geq 0$, depends on the process $\{B(t) : t \geq 0\}$ only through its rate $EB(0)$ and its asymptotic variance; see Iglehart and Whitt [17]. The input process is said to exhibit long-range dependence when this integral is infinite. The source traffic model shows that long-range dependence stems from level holding-time distributions with infinite variance.

Theorem 8.1 *If a level holding-time cdf G_j has infinite variance, then the source bandwidth process exhibits long-range dependence, i.e.,*

$$\int_0^\infty R(t)dt = \infty \ .$$

Proof. In (8.4) we have the component

$$\int_0^\infty g_{je}(x) P_{jj}(t|x) dx \ ,$$

which in turn has the component

$$\int_0^\infty g_{je}(x) G_j^c(t|x) = G_{je}^c(t) \ ,$$

but

$$\int_0^\infty G_{je}^c(t) dt = \infty$$

if G_j has infinite variance. (As can be seen using integration by parts, the integral is the mean of G_{je}; see p. 150 of Feller [13]. In general, G_{je} has $k^{\rm th}$ moment $m_{k+1}(G_j)/(k+1)m_1(G_j)$, where $m_k(G_j)$ is the $k^{\rm th}$ moment of G_j.)

Note that if approximation (8.6) holds, then the level process contributes to long-range dependence, but the within-level variation process does not, because

$$\sum_{j=1}^{J} p_j \int_0^\infty G_{je}^c(t) Cov(W_j(0), W_n(t)) dt$$

$$\approx \sum_{j=1}^{J} p_j \sigma_j^2 \frac{(1-\hat{g}_j(\eta_j))}{\eta_j m(G_j)} < \infty \ .$$

9. Conclusions

We have shown how transient analysis for network control (Section 5) and design (Sections 6 & 7) can be carried out for multi-level sources with general, possibly long-tailed, level-holding-time distributions. In Section 2 we analyzed the transient behavior of a general source traffic model composed of a semi-Markov level process and a zero-mean piecewise-stationary within-level variation process. We approximated the conditional aggregate demand from many sources given system state information by the conditional aggregate mean given level values and ages. The within-level variation process plays no role in this approximation. We showed that the conditional mean can be effectively computed using numerical transform inversion (Section 2) and developed several approximations to it (Sections 3 & 7). We showed how the model can be exploited to study the value of information (Section 4). We applied our techniques to examples in network control (Section 5) and design (Section 7).

Even though our approach is to focus on offered load, unaltered by loss and delay associated with finite capacity, we can apply the conditional mean approximation in Section 2 to develop an approximation to describe loss and delay from a finite-capacity system, just as described in Section 5 of [10] for the M/G/∞ arrival process.

References

[1] J. Abate and W. Whitt, "Numerical inversion of Laplace transforms of probability distributions," *ORSA J. on Computing*, vol. 7, pp. 36–43, 1995.

[2] S. Asmussen, O. Nerman and M. Olsson, "Fitting phase-type distributions via the EM algorithm," *Scand. J. Statist.*, vol. 23, pp. 419–414, 1996.

[3] R. Cáceres, P. G. Danzig, S. Jamin and D. J. Mitzel, "Characteristics of wide-area TCP/IP conversations," *Computer Communications Review*, vol. 21, pp. 101–112, 1991.

[4] H. Chen and A. Mandelbaum, "Discrete flow networks: bottleneck analysis and fluid approximations," *Math. Oper. Res.*, vol. 16, pp. 408–446, 1991.

[5] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[6] M. E. Crovella and A. Bestavros "Self-similarity in World Wide Web traffic – evidence and possible causes," *Proc. ACM Sigmetrics '96*, pp. 160–169, 1996.

[7] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, Jones and Bartlett, Boston, 1993.

[8] N. G. Duffield, "Economies of scale in queues with sources having power-law large deviation scalings", *J. Appl. Prob.* vol. 33, pp. 840–857, 1996.

[9] N. G. Duffield, "Queueing at large resources driven by long-tailed $M/G/\infty$-modulated processes," preprint, AT&T Laboratories, 1996.

[10] N. G. Duffield and W. Whitt, "Control and recovery from rare congestion events in a large multi-server system," *Queueing Systems*, vol. 26, pp. 69–104, 1997. Shorter version "Recovery from congestion in large multi-server systems" *Proceedings ITC-15*, Washington, DC, USA, 22-27 June, 1997, pp. 371–380.

[11] A. Feldmann, "Modeling characteristics of TCP connections," AT&T Laboratories, 1996.

[12] A. Feldmann and W. Whitt, "Fitting mixtures of exponentials to long-tail distributions to analyze network performance models," *Proc. IEEE Infocom '97*, pp. 1098–1106.

[13] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II, sec. ed., Wiley, New York, 1971.

[14] M. Grasse, M.R. Frater and J.F. Arnold, "Origins of long-range dependence in variable bit rate video traffic," in *Proceedings ITC-15,* Washington, DC, USA, 22-27 June, 1997, pp. 1379-1388.

[15] M. Grossglauser, S. Keshav, and D. Tse, "RCBR: A simple and efficient service for multiple time-scale traffic," in *Proc. ACM SIGCOMM95*, pp. 219-230, 1995.

[16] M. Harchol-Balter and A. Downey. "Exploiting Process Lifetime Distributions for Dynamic Load Balancing," in Proceedings of ACM Sigmetrics '96 Conference on Measurement and Modeling of Computer Systems , Philadelphia, PA, pp. 13–24, 1996.

[17] D. L. Iglehart and W. Whitt, "Multiple channel queues in heavy traffic, II: sequences, networks and batches," *Adv. Appl. Prob.* vol. 2, pp. 355–369, 1970.

[18] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1–15, 1994.

[19] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, pp. 226–244, 1995.

[20] J. Salehi, Z. Zhang, J. Kurose, and D. Towsley, "Supporting stored video: Reducing rate variability and end-to-end resource requirements through optimal smoothing," in *Proc. ACM Sigmetrics*, pp. 222-231, 1996.

[21] J. M. Van Campenhout and T. M. Cover, "Maximal entropy and conditional probability," *IEEE Trans. Inform. Theory*, vol. 4, pp. 183–189, 1981.

[22] W. Willinger, M. S. Taqqu, R. Sherman and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71–86, 1997.