Limit theorems relating to OBS networks

Damon Wischik

2 July 2004

Abstract

In this report we describe a selection of probabilistic limit theorems, and discuss which are relevant for modelling OBS networks. This report accompanies "Mathematical modelling of Optical Burst-Switched (OBS) Networks" [7].

Contents

Introduction

The Central Limit theorem is a typical example of the sort of limit theorem we are interested in. It says that, if $(\xi_1, \xi_2, \dots)$ is a sequence of independent identically distributed random variables with mean μ and variance σ^2 then (under mild conditions on the distribution)

$$\mathbb{P}\Big[\frac{\xi_1 + \cdots + \xi_L - L\mu}{\sqrt{L}} > x\Big] \to \mathbb{P}\big[N(0,\sigma^2) > x\big] \quad \text{as } L \to \infty. \tag{1}$$

For an example use of this, suppose we have an OBS buffer which empties every T time units, fed by L independent identically distributed traffic flows, and X_i is the amount of work from flow i that arrives in the interval $(0,T)$. Then we would set $\xi_i = X_i$ and use (1) as the approximation

$$\mathbb{P}\big[X_1 + \cdots + X_L > L\mu + x\sqrt{L}\big] \approx \mathbb{P}\big[N(0,\sigma^2) > x\big], \tag{2}$$

and be confident that this approximation is good when L is large. Suppose now that we do not know L, we only know that the sum $Y = X_1 + \cdots + X_L$ is the aggregate of many independent workloads. Suppose also that we do not know x, we only know the total buffer size $B = L\mu + x\sqrt{L}$. Rearranging (2) we get the naive approximation

$$\begin{aligned}\mathbb{P}\big[Y > B\big] &\approx \mathbb{P}\big[N(0, L^{-1}\operatorname{Var} Y) > L^{-1/2}(B - \mathbb{E}Y)\big]\\ &= \mathbb{P}\big[N(\mathbb{E}Y, \operatorname{Var} Y) > B\big].\end{aligned} \tag{3}$$

This should be accurate if Y is the aggregate of many independent flows, and if the mean buffer utilization $B/\mathbb{E}Y$ is of order $1 - L^{-1/2}$. It is incorrect to say that "Y is approximately normal"; it is correct to say that "If the system parameters are scaled in a certain fashion, then for the purposes of estimating the probability of overflow Y may be approximated by a normal". Example 3 gives a case where the incorrect statement can be seriously misleading.

There are many different ways to scale the system parameters, leading to different sorts of limit theorem, and in this report we will summarize a number of them which are useful (or which have been used) to describe OBS networks. Table 1 summarizes them. Here is some notation used in that table, and throughout this report. Let $X(0,t)$ be the amount of work arriving in a traffic flow in the interval $(0,t)$. Write X for the entire arrival process. Write $X^{\oplus L}$ for the sum of L independent copies of X, and write $X^{\otimes L}$ for the speeded-up version $X^{\otimes L}(0,t) = X(0, Lt)$. (These are referred to as the *many-flows scaling* and the *fast-time scaling* of X).

We will present limit theorems like (1), and omit working through to approximations like (3). This is to stress the point that the limit theorems are only relevant when system parameters are scaled in a suitable way.

1 Central limit theorem

We have already stated the central limit theorem in (1). This limit holds if the ξ_i are independent and have finite mean and variance. It also holds when they are not identical, and only nearly independent, under various conditions [2, Section 27].

The corresponding result which can tell us about the loss ratio is

$$\mathbb{E}\Big[\frac{\xi_1 + \cdots + \xi_L - L\mu}{\sqrt{n}} - x\Big]^+ \to \mathbb{E}\big[N(0,\sigma^2) - x\big]^+ \quad \text{as } L \to \infty.$$

CLT- $\oplus$ (1.1)	$\mathbb{P}\big[X^{\oplus L}(0,t) > L\mu t + x\sqrt{L}\big] \approx \mathbb{P}\big[N(0,\sigma_t^2) > x\big]$
CLT- $\otimes$ (1.2)	$\mathbb{P}\big[X^{\otimes L}(0,t) > L\mu t + x\sqrt{L}\big] \approx \mathbb{P}\big[N(0,\sigma^2) > x\big]$
LD- $\oplus$ (2.1)	$\frac{1}{L}\log\mathbb{P}\big[X^{\oplus L}(0,t) > Lx\big] \approx -\sup_{\theta\in\mathbb{R}}\big\{\theta x - \Lambda_t(\theta)\big\}$
LD- $\otimes$ (2.2)	$\frac{1}{L}\log\mathbb{P}\big[X^{\otimes L}(0,t) > Lx\big] \approx -\sup_{\theta\in\mathbb{R}}\big\{\theta x - \Lambda(\theta)\big\}$
LD-LRD- $\otimes$ (2.3)	$\frac{1}{L^{2(1-H)}}\log\mathbb{P}\big[X^{\otimes L}(0,t) \geq Lx\big] \approx -\sup_{\theta\in\mathbb{R}}\big\{\theta x - t^{2(1-H)}\Lambda(\theta t^{-(1-2H)})\big\}$
MD- $\oplus$ (3.1)	$\frac{1}{L^\beta}\log\mathbb{P}\big[X^{\oplus L}(0,t) > L\mu + L^{(1+\beta)/2}x\big] \approx -x^2/2\sigma_t^2$
Poisson- $\oplus$ (4)	$\mathbb{P}\big[X^{\oplus L}(0,L^{-1}t) \geq x\big] \approx \mathbb{P}\big[\text{Poisson}(\mu t) \geq x\big]$

Table 1: The different limiting results described in this report. The meanings of the various symbols are given in the relevant section.

1.1 CLT- $\oplus$

We have explained in (2) one way in which the central limit theorem can apply to OBS networks. That equation can be thought of as referring to $X^{\oplus L}(0,t)$, with $\mu = t^{-1}\mathbb{E}X(0,t)$ and $\sigma^2 = \operatorname{Var} X(0,t)$.

Example 1
For example, let X be an M/G/∞ source: that is, X consists of jobs which arrive as a Poisson process of rate λ, and which transmit at rate 1 for their duration, and that durations are independent and identical. Suppose the duration D is Pareto, that is, $\mathbb{P}(D > t) = (t+1)^{-\alpha}$ for some $\alpha > 1$. The formulae for μ and σ^2 are

$$\mu = \lambda t\mathbb{E}D$$
$$\sigma^2 = A + Bt + Ct^{3-\alpha}$$

where the quantities A, B and C are as given in [1]. ◇

1.2 CLT- $\otimes$

Let X be an arrival process with mean rate $\mu = t^{-1}\mathbb{E}X(0,t)$. For a wide range of arrival processes X (including Markov chains, though not including any that are long-range dependent) it can be shown that

$$\mathbb{P}\Big[L^{1/2}\big(L^{-1}X^{\otimes L}(0,t) - \mu t\big) \geq x\Big] \to \mathbb{P}\big[N(0,\sigma^2) \geq x\big], \tag{4}$$

where $\sigma^2 = \lim_{t\to\infty} t^{-1}\operatorname{Var} X(0,t)$. This is the limit at the foundation of heavy traffic queueing theory [12, Chapter 5].

This approximation makes sense when L is large, i.e. when we are interested in $X(0, Lt)$ for large L, sufficiently large that $\operatorname{Var} X(0, Lt) \approx Lt\sigma^2$, i.e. that correlations in the input traffic are negligible. What are the timescales of correlations in Internet traffic? TCP is self-clocked, and its feedback delay is equal to the round-trip time, so TCP traffic has correlations at timescales up to its round-trip time (and longer). Now, a TCP flow which passes through a burst assembler with assembly time T necessarily has round-trip time greater than T. So, at timescale T, correlations in the input traffic will not be negligible. We conclude that one needs to be very cautious in applying CLT-$\otimes$ to describe burst assemblers.

A necessary condition for (4) is that the variance $\operatorname{Var} X(0, t)$ grow linearly in t for large t. For long-range dependent flows, such as typical Internet flows, this does not hold. This has led to an awful lot of fuss. In some cases, however, an analogue does hold, which we now describe.

1.3 CLT-LRD-$\otimes$

For example let X be an on-off source with Pareto on and off times, with mean rate μ, and let $\tilde{X}$ be given the CLT-$\oplus$ limit

$$\tilde{X}(0, t) = \lim_{M \to \infty} M^{1/2}\big(M^{-1} X^{\oplus M}(0, t) - M\mu\big).$$

Then $\tilde{X}^{\otimes L}$ satisfies a modified CLT-$\otimes$ limit: for suitable σ,

$$\mathbb{P}\Big[L^{1-H}\big(L^{-1}\tilde{X}^{\otimes L}(0, t)]\big) > x\Big] \to \mathbb{P}\big[N(0, \sigma^2) > x\big]. \tag{5}$$

Note that $L^{1/2}$ has been replaced by L^{1-H}, reflecting the fact that the variance of $\tilde{X}(0, t]$ grows like t^{2H}. (We do not have to subtract the mean in this equation, because $\tilde{X}$ has already been normalized to have mean rate zero.) This result is described more fully by Willinger et al. [13]. See also [6], where it is shown that (5) holds for a wide class of Gaussian processes.

We do not believe that this limit is useful for modelling OBS networks, for the reasons given in Section 2.3 below.

2 Large deviations

Let ξ be a random variable, and let ξ^L be the sum of L independent copies of ξ. It is a standard result known as Cramér's theorem (see e.g. [5]) that

$$\frac{1}{L} \log \mathbb{P}\big[\xi^L > Lx\big] \to -I(x) \tag{6}$$

where

$$I(x) = \sup_{\theta \in \mathbb{R}} \big\{\theta x - \Lambda(\theta)\big\} \quad \text{and} \quad \Lambda(\theta) = \log \mathbb{E} \exp(\theta \xi) \tag{7}$$

if the log moment generating function $\Lambda(\theta)$ is finite in a neighbourhood of the origin). There is a refined version of this estimate, known as the Bahadur-Rao improvement:

$$\mathbb{P}\big[\xi^L > Lx\big] = \frac{1}{\hat{\theta}\sqrt{2\pi L\sigma^2(\hat{\theta})}} e^{-LI(x)}\big(1+o(1)\big) \tag{8}$$

where $\hat{\theta}$ is the optimizing parameter in $I(x)$ and $\sigma^2(\theta) = \Lambda''(\theta)$.

The same expression (6) holds for $L^{-1}\log\mathbb{E}(\xi^L - Lx)^+$, and its refinement is given in [11] as

$$\mathbb{E}\big[\xi^L - Lx\big]^+ = \frac{1}{\hat{\theta}^2\sqrt{2\pi L\sigma^2(\hat{\theta})}} e^{-LI(x)}\big(1+o(1)\big) \tag{9}$$

Equation (6) concerns the log-probability of an event, as opposed to the central limit equation (1) which concerns the actual probability. Log-probabilities are governed by the principle of the largest term, which may heuristically be expressed as

$$\log\mathbb{P}(\xi \in A) \approx \sup_{a\in A}\log\mathbb{P}(\xi \approx a).$$

This makes it very easy to manipulate expressions! Expressions such as (3) on the other hand can be hard to manipulate, as they involve integrals. One should think of LD as giving a first-order approximation to a probability, based only on the most likely way for an event to occur: if it tells us that two probabilities are different, they are substantially different. CLT is more refined, and may distinguish between the probabilities of events which are indistinguishable at a large deviations level.

The unrefined estimates can be quite far off, and one way to remedy this is with the assistance of simulation. The trouble with small probabilities is that it takes a very very long simulation to estimate them well (and it is small probabilities that large deviations is concerned with). There is, happily, a branch of simulation theory which deals with this, called *fast simulation*. The idea behind fast simulation is that we can use knowledge of a large deviations result to speed up the simulation, by biasing our random number generator: the bias is given by the optimizing $\hat{\theta}$. This is a large field, which we will not go into. See [3].

2.1 LD- $\oplus$

Let $X^{\oplus L}$ be the aggregate of L independent copies of an arrival process X. It is straightforward to apply the basic limit theorem above to approximate $\mathbb{P}[X^{\oplus L}(0,T) > B]$. The approximation we get should be good when L is large, when B is comparable to $\mathbb{E}X^{\oplus L}(0,T)$ (e.g. $B = 1.3\,\mathbb{E}X^{\oplus L}(0,T)$), and the loss probability is $\exp(-L\ O(1))$. In other words, the large deviations limit is suitable for describing regimes in which there is lower loss and also lower mean buffer utilization than in the central limit.

2.2 LD- $\otimes$

Large deviations also tells us about the fast-time scaling. It turns out that (6) also holds when ξ^L is the sum of random variables that are nearly independent in the sense that

$$\Lambda(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E} \exp \theta\xi^L \tag{10}$$

exists and is suitably smooth. This is known as the Gärtner-Ellis theorem [5]. (The refined expressions no longer hold.)

If we let $\xi^L = X^{\otimes L}(0,t)$ then this result tells us about

$$\lim_{L\to\infty} \frac{1}{L} \log \mathbb{P}\big[X^{\otimes L}(0,T) > B\big]$$

when B is of order L, i.e. when we are considering very large buffers. The smoothness condition requires that the correlations between say $X(0,Lt_1)$ and $X(Lt_1,Lt_2)$ become negligible for large L so that $X^{\otimes L}$ has asymptotically independent increments.

This result is not useful for modelling OBS networks, for same reason that CLT- $\otimes$ is not useful. It is also not useful for the simpler reason that (10) does not hold when we are considering traffic flows which are long-range dependent. For such flows we need a modified result, which we now describe.

2.3 LD-LRD- $\otimes$

The modified version of the LD- $\otimes$ estimate for long-range dependent traffic says that

$$\frac{1}{L^{2(1-H)}} \log \mathbb{P}\big[L^{-1}X^{\otimes L}(0,t) \geq x\big] \to -\sup_{\theta\in\mathbb{R}} \Big\{\theta x - t^{2(1-H)}\Lambda\Big(\frac{\theta}{t^{(1-2H)}}\Big)\Big\}$$

where

$$\Lambda(\theta) = \lim_{t\to\infty} \frac{1}{t^{2(1-H)}} \log \mathbb{E} \exp\big(\theta t^{1-2H} X(0,t]\big). \tag{11}$$

(This is in fact just a restatement of the Gärtner-Ellis result.) If we use this to approximate $\mathbb{P}[X^{\otimes L}(0,T) > B]$, then the approximation will be valid to the extent that the limit (11) is approached, i.e. to the extent that the traffic is well-described by its power-law scaling.

We argued in Section 1.2 that the problem with CLT- $\otimes$ is that it ignores correlations in the traffic. But the LD-LRD- $\otimes$ limit *does* allow for correlations over long timescales—indeed this limit is most commonly used for describing the impact of long-range dependence [8, 14]. The important question here is *what sort of correlations?* It is widely accepted that, over long timescales, Internet traffic is well-described by power-law scaling. This power-law scaling derives perhaps from the fact that the files to be transferred are heavy-tailed. However it is also well-known that, over short timescales, one needs a richer description. Feldmann, Gilbert, and Willinger [9], for example, suggest a multifractal description. However one characterizes it, the distinctive short-timescale behaviour of Internet traffic seems to arise from short-timescale causes such as the TCP control loop. As we argued in Section 1.2, the burst assembly time will be shorter than the round trip time of a TCP flow, and therefore the power-law scaling does not adequately describe traffic over the timescales we are interested in. We acknowledge, of course, that traffic in optical networks will display long-range dependence, but this only captures a minor part of what matters—what matters being the traffic characteristics over the intrinsic timescale of burst formation.

Example 2

Consider the 'chunked fractional Brownian motion' traffic source X, defined as follows. Let $Y(t)$ be a fractional Brownian motion with Hurst parameter H, drift μ and variance parameter σ^2 (so that $\mathbb{E}Y(t) = \mu t$ and $\operatorname{Var} Y(t) = \sigma^2 t^{2H}$). Pick some period $U > 0$, let ϕ be a random variable uniform in $[0, U]$, and set

$$X(0,t) = Y\big(U\lfloor U^{-1}(t + U - \phi)\rfloor\big).$$

Then

$$\begin{aligned}\log \mathbb{E}e^{\theta A(0,t)} &= \theta\mu t + \tfrac{1}{2}\theta^2\sigma^2 t^{2H} \\ &\quad + \log\Big\{(1-\delta)\exp\Big[-\theta\mu\delta U - \tfrac{1}{2}\theta^2\sigma^2 U^{2H}\big((n+\delta)^{2H} - n^{2H}\big)\Big] \\ &\quad + \delta\exp\Big[\theta\mu(1-\delta)U + \tfrac{1}{2}\theta^2\sigma^2 U^{2H}\big((n+1)^{2H} - (n+\delta)^{2H}\big)\Big]\Big\}\end{aligned} \tag{12}$$

and

$$\operatorname{Var} A(0,t) = \mu^2 U^2 \delta(1-\delta) + \sigma^2 t^{2H}\Big[(1-\delta)\Big(\frac{n}{n+\delta}\Big)^{2H} + \delta\Big(\frac{n+1}{n+\delta}\Big)^{2H}\Big]$$

and so

$$\Lambda(\theta) = \theta\mu + \tfrac{1}{2}\theta^2\sigma^2.$$

This will be a reasonable approximation if T is significantly larger than U. If T is close to the period U, then n will be in the range $0, 1, 2$, and so the correction term in (12) will be significant. ◇

3 Moderate deviations

The central limit theorem tells us about common deviations from mean behaviour, which are governed by mean and variance, and have $O(1)$ probability. Large deviations tells us about large deviations from mean behaviour, which depend on the entire distribution, and have $e^{-L\,O(1)}$ probability.

It turns out that mean and variance are sufficient to describe not only common events but also moderately rare events. Let ξ^L be the sum of L independent copies of a random variable ξ. Let $\mu = \mathbb{E}\xi$ and $\sigma^2 = \operatorname{Var}\xi$. It can be shown that, under a minor technical condition on the distribution of ξ, for any $\beta \in (0, 1)$

$$\frac{1}{L^\beta}\log\mathbb{P}\Big(L^{(1-\beta)/2}\big(L^{-1}\xi^L - \mu\big) \geq x\Big) \to -\sup_{\theta\in\mathbb{R}}\big\{\theta x - (\theta\mu + \tfrac{1}{2}\theta^2\sigma^2)\big\} \tag{13}$$

$$= -x^2/2\sigma^2. \tag{14}$$

This is like the large deviations result (6), but where the log moment generating function Λ has been replaced by its second-order approximation $\Lambda(\theta) \approx \theta\mu + \frac{1}{2}\theta^2\sigma^2$, making it simple to calculate the supremum.

(In fact (13) is a special case of the Gärtner-Ellis theorem, as explained for example in [10, Chapter 9].)

CLT is useful because it relies on a parsimonious description of the distribution (it uses only mean and variance); LD is useful because of the principle of the largest term (which says we can replace integrals by supremums). Often it will turn out that both CLT and LD are nevertheless analytically intractable: the one because we need integrals to work out probabilities, the other because the full statistical characteristics of a traffic flow are unwieldy. Then one can use MD. Moderate deviations theory should be thought of as a first-order approximation of probabilities (like LD) combined with a second-order approximation of distributions (like CLT). It combines the parsimony (and inaccuracy) of CLT with the tractability (and inaccuracy) of LD.

Practically, MD means that easily-measured quantities like mean and variance can be used to predict hard-to-measure quantities like the (hopefully rare) probability of overflow. CLT is not appropriate for this purpose, since it is only appropriate for estimating probabilities which are $O(1)$. Nor is LD appropriate, since to use it we need to know the full log moment generating function Λ, yet it can be as hard to measure $\Lambda(\theta)$ for large θ as it is to simulate the event of interest.

We conjecture that the same expression (14) holds for expected loss, and that refined expressions corresponding to the large deviations refined expressions (8) and (9) also hold, when $\Lambda(\theta)$ is replaced by its second-order approximation.

3.1 MD- $\oplus$

It is straightforward to apply moderate deviations to approximate $\mathbb{P}[X^{\oplus L}(0,T) > B]$. The approximation should be good when mean buffer utilization is $O(1 - L^{-(1-\beta)/2})$ and when the loss probability is $\exp(-L^\beta O(1))$, for some $0 < \beta < 1$. These are intermediate between the corresponding magnitudes for CLT- $\oplus$ ($\beta = 0$) and LD- $\oplus$ ($\beta = 1$).

The following example shows that one cannot wantonly make the approximation $\Lambda(\theta) \approx \theta\mu + \frac{1}{2}\theta^2\sigma^2$. It only works when the system parameters are scaled properly.

Example 3 (How things can go wrong)
Let X be as specified in Example 2. Suppose that $\sigma^2 = 0$, to make the following calculations tractable. This means that X is a pure periodic source. Let $T = (n+\delta)U$. By CLT- $\oplus$,

$$\mathbb{P}\big[X^{\oplus L}(0,T) > B\big] \approx \mathbb{P}\big[N(\mu T, \sigma_T^2)^{\oplus L} > B\big] \tag{15}$$

$$= \mathbb{P}\Big[\mathrm{N}(0,1) > \frac{B - L\mu T}{L\sigma_T}\Big]. \tag{16}$$

By LD- $\oplus$,

$$\log \mathbb{P}\big[X^{\oplus L}(0,T) > B\big] \approx -\sup_{\theta\in\mathbb{R}}\Big\{\theta B - L\log \mathbb{E}e^{\theta X(0,T)}\Big\} \tag{17}$$

where the log moment generating function is given by (12). It can be shown that this expression is

$$= \begin{cases} L\log\delta & \text{for } B = L\mu(t+U\delta) \\ -\infty & \text{for } B > L\mu(t+U\delta) \end{cases} \tag{18}$$

These two statements (16) and (17) are indubitably correct, insofar as the limits allow us to make the approximations. But we run into trouble when we wantonly combine them. For example, applying the LD- $\oplus$ approximation to (15),

$$\begin{aligned} \log \mathbb{P}\big[X^{\oplus L}(0,T] > B\big] &\approx -\sup_{\theta\in\mathbb{R}}\Big\{\theta B - L(\theta\mu T + \tfrac{1}{2}\theta^2\sigma_t^2)\Big\} \\ &= -\frac{(B-L\mu T)^2}{2L\sigma_T^2} \end{aligned} \tag{19}$$

This contradicts (18). The problem here is that we need to be more subtle in interpreting approximations like (16) and (17). Each only holds for high levels of aggregation, and for B of a certain scale: (15) only holds when the buffer utilization $L\mu T/B$ is $O(1-L^{-1/2})$, and (17) only holds when it is $O(1)$. As aggregation increases *these two scales diverge*.

In fact, the MD- $\oplus$ scaling shows us that there is a scale of B in which the approximation (19) holds: this scale is $B = O(1-L^{-(1-\beta)/2})$ for any $0<\beta<1$. For buffers of that scale, it is legitimate to mix LD- $\oplus$ and CLT- $\oplus$. $\diamond$

3.2 MD- $\otimes$

There is also a family of moderate deviations results in between CLT- $\otimes$ and LD- $\otimes$: under suitable conditions, for $\beta\in(0,1)$

$$\frac{1}{L^\beta}\log\mathbb{P}\Big(L^{(1-\beta)/2}\big(L^{-1}X^{\otimes L}(0,t)-\mu t\big)\geq x\Big) \to -x^2/2\sigma^2,$$

where $\sigma^2 = \lim_{t\to\infty} t^{-1}\operatorname{Var} X(0,t)$, with related results. We believe this is not appropriate for modelling OBS networks for the same reasons that neither CLT- $\otimes$ nor LD- $\otimes$ is appropriate.

4 Poisson- $\oplus$ process limit

Here is another results about $\oplus$ scalings. Let X be a point process with mean arrival rate μ. Then $X^{\oplus L}$ converges to a Poisson process, over short timescales, and

$$X^{\oplus L}(0, L^{-1}t) \to \text{Poisson}(\mu t) \tag{20}$$

where convergence is in distribution (which means we can replace the left hand side by the right in estimating both the probability of overflow and the expected loss). This result relies on the limit

$$\lim_{t\to 0}\frac{1}{t}\log\mathbb{E}\exp\theta X(0,t)=\mu(e^{\theta}-1). \tag{21}$$

(The right-hand side is the log moment generating function for a Poisson random variable.) For details on how this limit can be applied to conventional queueing problems see [4].

In words, as aggregation increases, and as the timescale of interest decreases in proportion, the aggregate traffic converges to a Poisson process. This should be accurate over timescales t for which the limit (21) is good, that is, timescales over which the number of packets from a single flow is approximately Poisson. If t is a significant part of the round trip time, as it would be in a Type II OBS network, then (as we argued in Sections 1 and 2) we would expect non-trivial correlation structure over timescale t, and therefore the limit (21) will not be good. If t is much shorter than the round trip time, as it would be in a Type I OBS network, the approximation should be reasonable. We conclude that one must be cautious in using this Poisson limit to model burst accumulators.

Note also that this limit concerns point processes. To use it in traffic modelling, we would say that a point corresponds to a packet; we then need to decide what size packet. It is mathematically awkward to have analyse models with packets of different sizes.

We have already remarked that it is dangerous to combine different limits. One must pay close attention to exactly what sort of scaling each limit entails for the system parameters, if one is to produce meaningful answers. Here is another example of what can go wrong.

Example 4 (How things can go wrong)
Let X be as described in Example 2. Let $\sigma^2=0$, to make the calculations tractable. So X is just a periodic point process, of random phase, with μ units of work arriving every U time units. Let $T=(n+\delta)U$.

It is easy to see that

$$\mathbb{P}\big[X^{\oplus L}(0,T)>B\big]=\mathbb{P}\big[\mathrm{Bin}(L,\delta)>B-L\mu Un\big]. \tag{22}$$

According to the Poisson-$\oplus$ limit, $X^{\oplus L}$ converges to a Poisson process of rate $L\mu$. Now, according to the CLT-$\otimes$ limit, if $X^{\oplus L}$ were a Poisson process of rate $L\mu$ then for large T and B

$$\mathbb{P}\big[X^{\oplus L}(0,T)>B\big]\approx\mathbb{P}\Big[N(0,1)>\frac{B-TL\mu}{\sqrt{TL\mu}}\Big]$$

This has a totally different form to the true answer (22). (Consider how the two formulae vary with δ.)

The trouble is that the Poisson-$\oplus$ limit is concerned with short timescales, and the CLT-$\otimes$ limit is concerned with long timescales. It is hard to see how the two limits can be used together coherently. $\diamond$

References

[1] Ron Addie, Petteri Mannersalo, and Ilkka Norros. Most probable paths and performance formulae for buffers with Gaussian input traffic. *European Transactions on Telecommunications*, 2002.

[2] Patrick Billingsley. *Probability and Measure*. Wiley, 3rd edition edition, 1995.

[3] James A. Bucklew. *Large Deviation techniques in decision, simulation, and estimation*. Wiley, 1990.

[4] Jin Cao and Kavita Ramanan. A Poisson limit for buffer overflow probabilities. In *Proceedings of IEEE Infocom*, 2002. URL `http://www.ieee-infocom.org/2002/papers/655.pdf`.

[5] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2 edition, 1998.

[6] A. B. Dieker. Conditional limit theorems for queues with Gaussian input, a weak convergence approach. URL `http://homepages.cwi.nl/~ton/publications/`. Submitted for publication. Available as CWI technical report PNA-E0306, 2003.

[7] Michael Dueser, Polina Bayvel, and Damon Wischik. Mathematical modelling of optical burst-switched (OBS) networks. Submitted, 2004.

[8] N. G. Duffield and Neil O'Connell. Large deviations and overflow probabilities for the general single-server queue, with applications. *Math. Proc. Camb. Phil. Soc.*, 1995.

[9] Anja Feldmann, Anna Gilbert, and Walter Willinger. Data networks as cascades: investigating the multifractal nature of Internet WAN traffic. In *Proceedings of ACM Sigcomm*, 1998. URL `http://www.research.att.com/~anja/feldmann/sigcomm98_cascade.abs.html`.

[10] Ayalvadi Ganesh, Neil O'Connell, and Damon Wischik. *Big Queues*. Springer, 2004.

[11] N. Likhanov and R. R. Mazumdar. Cell loss asymptotics for buffers fed with a large number of independent stationary sources. *Journal of Applied Probability*, 36:86–96, 1999.

[12] Ward Whitt. *Stochastic-process limits*. Springer, 2001.

[13] W. Willinger, V. Paxson, and M. S. Taqqu. Self-similarity and heavy tails: structural modelling of network traffic. In R. Adler, R. Feldman, and M. S. Taqqu, editors, *A practical guide to heavy tails: statistical techniques for analysing heavy tailed distributions*. Birkhauser, 1998. URL `http://math.bu.edu/people/murad/pub/tails-w16-posted.ps`.

[14] D. J. Wischik and A. J. Ganesh. The calculus of Hurstiness. URL http://www.statslab.cam.ac.uk/~djw1005/Stats/Research/hurstiness.html. Submitted, 2004.