Title	The superposition of variable bit rate sources in an ATM multiplexer
Author(s)	Norros, Ilkka; Roberts, James W.; Simonian, Alain; Virtamo, Jorma T.
Citation	IEEE Journal on Selected Areas in Communications vol. 9(1991):3, pp. 378-387
Date	1991
URL	http://dx.doi.org/10.1109/49.76636
Rights	Copyright © (1991) IEEE. Reprinted from IEEE Journal on Selected Areas in Communications. This article may be downloaded for personal use only

The Superposition of Variable Bit Rate Sources in an ATM Multiplexer

Ilkka Norros, James W. Roberts, Alain Simonian, and Jorma T. Virtamo

***Abstract*—When variable bit rate sources are multiplexed in an ATM network, there arise queues with a particular form of correlated arrival process. We analyze such queues by exploiting a result expressing the distribution of work in system of the $G/G/1$ queue originally derived by Beneš. We provide a simple alternative demonstration of this result and extend it to the case of fluid input systems. The result is applied first to a queue where the arrival process is a superposition of periodic sources (the $\Sigma D_i/D/1$ queue), and then to a variable input rate constant output rate fluid system. The latter is shown to model the so-called burst component of the considered superposition queueing process. The difference between this and the real queue, the cell component, can be evaluated by means of the results obtained for the $\Sigma D_i/D/1$ queue. The relative importance of these two components is explored with reference to the particular case of a superposition of on/off sources.**

I. Introduction

The ATM technique allows digital communications of any type to share common transmission links and switching devices on a statistical multiplexing basis. Information is transmitted in the form of constant length cells, and the different types of communication are distinguished by the way their sources produce cells. Constant bit rate sources emit cells periodically at a frequency determined by their bit rate. On/off sources emit cells periodically during activity periods, or "bursts," of variable length alternating with silences, also of variable length. More generally, the cell emission rate might vary continuously, in the sense that the interval between successive cells varies gradually, or discontinuously, with the rate changing at random instants between different constant values. When these sources share the same network resources, there arise queueing systems of a particular nature. In the present paper, we consider such a queue which would typically occur in the buffer of an ATM multiplexer. Our objective is to derive analytical tools allowing buffer dimensioning for very low overflow probabilities.

The superposition of on/off sources has been studied, notably, in the context of packetized speech and it has long been recognized that the convenient device of assuming that the superposition of a large number of independent sources yields a Poisson arrival process can lead to quite inaccurate results [7], [35], [14]. More accurate queueing models must take account of the correlated nature of the cell arrival process [35], [26]. Two kinds of correlation can be identified:

- negative correlation between successive interarrival times due to the periodic nature of cell emissions by active sources;
- positive correlation between the average arrival intensities in successive periods of length greater than the inter-cell time of the multiplexed sources.

Various modeling approaches proposed in the literature attempt to account for these dependence effects.

The superposition arrival process is modeled as a renewal process in [35]. Correlations of the second kind are approximately accounted for by the choice of the second moment of the interarrival time distribution. An approach which has proved more popular is to approximate the arrival process by a Markov-modulated Poisson process (MMPP): the arrival rate is governed by the evolution of a discrete-space Markov process; when in state i, cells are generated according to a Poisson process of rate λ_i. System state is then Markovian, and equilibrium state equations can be solved algorithmically (e.g., using matrix geometric methods). A two-state MMPP is proposed in [14] where the four parameters (state transition rates and the two arrival intensities) are chosen to match four arrival process characteristics. Different possibilities are available for the choice of these four characteristics [27], [24]. In [15], the author models a superposition of N on/off sources as an N-state MMPP where the arrival rate is simply proportional to the number of active sources. Clearly, in any MMPP model, when the number of states increases, the numerical problem of solving the state equations to determine the performance measures of interest is compounded. Use of spectral expansion techniques appears to reduce the computational complexity [11].

The use of point process models, such as the MMPP, can be criticized on two counts:

- they do not accurately represent short-term correlation effects [26];
- performance evaluation remains complex.

Simpler models, which also retain the long-term correlation characteristics of the arrival process, are obtained with the so-called fluid approximation: the arrival of (discrete) cells is assimilated to the (continuous) arrival of a liquid. This appears as a reasonable approximation when the cell interarrival times are small compared to the time between arrival rate changes.

Kosten was perhaps the first to employ this approximation for a superposition of exponentially distributed bursts starting at the epochs of a Poisson process (infinite source model) [18]. The superposition of a finite number of on/off fluid sources is considered in [1]. Like the MMPP, the arrival rate is modulated according to a Markov process. Nonexponentially distributed burst and silence lengths (e.g., Erlang or hyperexponential distributions) can be accounted for by multiplying the number of states of this underlying process [19]–[22]. The model in [1] is generalized in [36] to allow a buffer of finite capacity. A superposition of on/off sources of potentially different bit rates is analyzed in [16] while the model in [32] allows sources with two types of on/off behavior. Discrete-time models with correlated input described in [23] are also members of the family of fluid approximations. A continuously varying intensity input

Manuscript received April 26, 1990; revised November 9, 1990. This paper was supported by the European cooperative project COST 224 and Telecom Finland.

I. Norros and J. T. Virtamo are with the Technical Research Centre of Finland, SF-02150 Espoo, Finland.

J. W. Roberts and A. Simonian are with the Centre National d'Etudes des Telecommunications, 92131 Issy les Moulinex, France.

IEEE Log Number 9042042.

0733-8716/91/0400-0378$01.00 © 1991 IEEE

process is modeled as a fluid model with a diffusion input in [31]. The same system is considered in [17] as an asymptotic limit for a superposition of a large number of on/off sources.

The negative correlation between cell interarrival times is a local phenomenon occurring while the composition of active sources remains constant. When the overall arrival rate remains below multiplex capacity, the system behaves like the so-called $\Sigma\ D_i/D/1$ queue: a superposition of independent periodic sources of possibly different periods and random phase is offered to a deterministic server. In the case where the sources all have the same period, this queue was solved in [9] and revisited more recently in [13], [28], [29], [4], [38], [30], [25] where alternative formulas and calculation procedures are proposed. The more general superposition of sources of different periods is considered in [29] and [37], where accurate approximate formulas for the queue length distribution are derived.

The focus of the present paper is on the relationship between the real queueing system and the equivalent fluid queue. We see this fluid queue not as an approximation but as the exact expression of one component of the real queue, to which we must add a further component due to local fluctuations of the cell arrival rate about the fluid average. The first component accounts for the long-term positive correlations in the arrival rate process, while we can introduce the effects of the negative correlation between cell interarrival times in the second.

We first establish, in Section II, some basic queueing results useful in analyzing both discrete and fluid arrival queues. These results are applied, in Section III, to the $\Sigma\ D_i/D/1$ queue and, in Section IV, to queues with fluid arrivals. In the latter case, the derived relations allow an approach to more general arrival processes than those amenable to analysis by the usual Markovian methods. In Section V, we establish the relationship between the real queue and its burst component derived from the equivalent fluid model. In Section VI, we consider some numerical examples, illustrating the decomposition of queue behavior into burst and cell components.

II. General Results for Queueing Analysis

After a succinct description of the type of arrival process to be considered, we introduce some general tools allowing us to analyze the queue arising in an ATM multiplexer.

A. Arrival Process

It is an essential feature of an ATM-based network that the bit rate of sources can be variable. This variability may take different forms. In Fig. 1, we distinguish:

- on/off sources;
- more general piecewise constant rate sources;
- continuously varying rate surces.

We expect many forms of data- and image-based communication will exhibit output of the first kind while the latter two may be more typical of multi-media and VBR video communications. Bit rate variability is manifested in the network by the changing frequency of cell arrivals. However, at source, data may be considered to be generated as a continuous bit stream which is "packetized" into cells at the network input. It is essentially this bit stream which is depicted in Fig. 1.

In considering the queue arising when a superposition of such variable bit rate sources is offered to an ATM multiplex, it is useful to distinguish two time scales:

- a "cell scale" where we consider the congestion due to simultaneous cell arrivals occurring in a time span equivalent to a source inter-cell time;

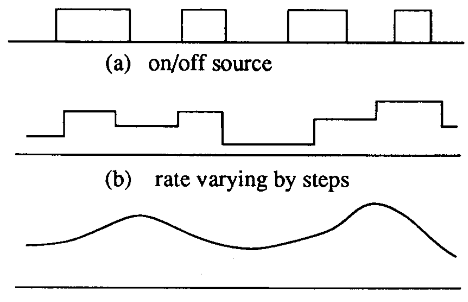


Fig. 1. Variable bit rate sources.

- a "burst scale" where congestion occurs when the total arrival rate, averaged over a period greater than an inter-cell time, is greater than the multiplex capacity.

To study burst scale congestion, it is convenient to ignore the discrete nature of the cell arrival process and revert to the notion of a continuous bit stream. This allows us to use fluid models to evaluate what we refer to as the "burst component" of the multiplex queue. This appears somewhat simpler than analyzing the real (cell) queue directly. To correctly evaluate the latter, we must also account for a "cell component" resulting from the discrete (i.e., packetized) nature of the arrival process. In the remainder of this section, we establish some general tools for studying both cell and burst components.

B. Virtual Waiting Time Distribution

The general result for constant service-time single server queues proved in [29] and [37] turns out to be a special case of a theorem due to Beneš expressing the virtual waiting time distribution of a $G/G/1$ queue [2]. The results we quote below can also be rigorously proved along the lines of the demonstrations in [2] (see also [5]) but we prefer to give a more intuitive justification.

Consider a system where a single constant rate server is submitted work according to a random process. The work may arrive in jumps (customers arriving to a queue) or in a continuous flow (fluid arriving to a reservoir). The server has the capacity to accomplish 1 unit of work per unit of time and has a waiting room of unlimited capacity. We assume the arrival process is such that the system is stable and we consider the state of the system at an arbitrary instant in time which we take for time zero.

Let $W(t)$ $(t \geq 0)$ be the amount of work arriving to the system in the interval $(-t, 0)$ and let V_t be the work still in the system at time $-t$: in queueing terms, V_t is the virtual waiting time at $-t$. Define $X(t) = W(t) - t$ to be the excess work arriving in $(-t, 0)$.

It is easy to see that:

$$V_0 = \sup_{t \geq 0} \{X(t)\}. \tag{2.1}$$

The virtual waiting time can thus be obtained as the maximum of a stochastic process.

Now, $X(t)$ may increase by jumps but decreases always occur continuously. Furthermore, assuming the system is stable, $X(\infty) = -\infty$ and thus $X(t)$, for $t \geq 0$, attains all values in the interval $(-\infty, V_0]$. It follows that, if $V_0 \geq x$, then there exists a unique instant $-u$ $(u \geq 0)$ such that $X(u) = x$ and $X(w) <$

x for all $w > u$. The converse is obviously true and we have,

$$\{V_0 \geq x\} \Leftrightarrow \{\exists \text{ unique } u \text{ s.t. } X(u) = x \text{ and } X(w) < x \text{ for } w > u\}.$$

The instant u is the greatest value of t such that $X(t) = x$ (see Fig. 2). It defines a partition of all trajectories of the process $X(t)$ $(t \geq 0)$ leading to a value $V_0 \geq x$. Let $\nu(x)$ be the complementary distribution of V_0: $\nu(x) = \Pr\{V_0 > x\}$ (we write $\nu(x-)$ for $\Pr\{V_0 \geq x\}$). Noting that $u > 0$ when $V_0 > 0$, we deduce the relation:

$$\nu(x-) = \int_{u>0} \Pr\{X(u + du) < x \leq X(u) \text{ and } X(w) < x \text{ for } w > u\}. \tag{2.2}$$

Now, applying relation (2.1) at the point u (i.e., u takes the role of the arbitrary instant 0), we deduce the equivalence:

$$\{V_u = 0\} \Leftrightarrow \{X(w) \leq X(u) \text{ for } w > u\}.$$

Except for some very strange arrival processes, therefore, the instant u is almost surely uniquely defined by $\{X(u) = x$ and $V_u = 0\}$. A sufficient condition is that the event $\{x$ is a local maximum of $(X(t))_{t \geq 0}\}$ has zero probability. With this condition, we also have $\nu(x) = \nu(x-)$ and we can write:

$$\nu(x) = \int_{u>0} \Pr\{X(u + du) < x \leq X(u) \text{ and } V_u = 0\}. \tag{2.3}$$

1) The G/D/1 Queue: In a constant service-time queueing system, work arrives discontinuously in quanta of equal size equivalent to the service requirement of one customer. Take this requirement as the unit of work and let $N(t)$ be the number of customer arrivals in $(-t, 0)$. We then have $X(t) = N(t) - t$ and the probability in the right-hand side of (2.3) is concentrated on the values of u such that $x + u$ is an integer. We can thus replace the integral by a summation:

$$\nu(x) = \sum_{n>x} \Pr\{N(n - x) = n \text{ and } V_{n-x} = 0\}. \tag{2.4}$$

2) Fluid Arrival Process: If $W(t)$ is a continuous function, the considered system behaves like a reservoir with constant output whenever its content V_t is non-zero. Let Λ_t be the arrival rate at time $-t$:

$$\Lambda_t = \frac{dW(t)}{dt} = 1 + \frac{dX(t)}{dt}.$$

Writing $X(u + du) = X(u) + (\Lambda_u - 1)\,du$ in (2.3), we deduce:

$$\nu(x) = \int_{u>0} \Pr\{x \leq X(u) < x + (1 - \Lambda_u)\,du \text{ and } V_u = 0\}.$$

Summing over all possible values of Λ_u (the system is certainly not empty at u if the arrival rate is greater than 1), we deduce:

$$\nu(x) = \int_{u>0} \int_{0 \leq \lambda \leq 1} \Pr\{x \leq X(u) < x + (1 - \lambda)\,du \quad \lambda \leq \Lambda_u < \lambda + d\lambda \text{ and } V_u = 0\}$$

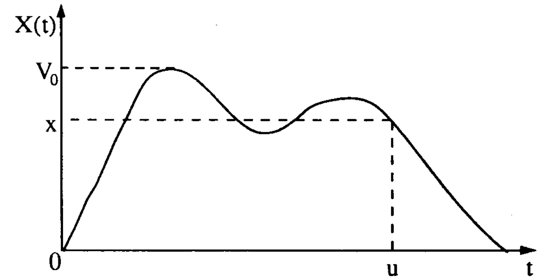


Fig. 2. A realization of the process $X(t)$.

which we can rewrite as:

$$\nu(x) = \int_{u>0} \int_{0 \leq \lambda \leq 1} (1 - \lambda) \frac{d^2}{dx\, d\lambda} \cdot \Pr\{X(u) \leq x, \Lambda_u \leq \lambda \text{ and } V_u = 0\}\, d\lambda\, du. \tag{2.5}$$

See [33] for an alternative derivation of this result.

III. A Superposition of Periodic Sources

When constant bit rate sources are multiplexed in an ATM network, the cell arrival process appears as a superposition of periodic streams. We consider here the queue arising when S independent periodic streams are superposed in a multiplexer with unlimited waiting room. The time between two cell emissions of stream i is equal to D_i, expressed in units of the multiplex cell transmission time, and its phase with respect to a common time origin is chosen at random between 0 and D_i. The multiplex load, $\Sigma\, 1/D_i$, is assumed to be less than one. We refer to this system as the $\Sigma\, D_i/D/1$ queue. Its evaluation has application to the superposition of variable bit rate sources, as explained in Section V below.

A. Bounds and Approximations

Bounds and accurate approximations for the queue length distribution of the $\Sigma\, D_i/D/1$ queue were derived in [29] and [37]. We reformulate the principal results here to give expressions for the virtual waiting time distribution $\nu(x)$.

To use expression (2.4), we need the joint probability of n cell arrivals in an interval $(-t, 0)$ and an empty system at time $-t$. Let the number of cell arrivals from source i be $N_i(t)$ and let $d_i(t)$ and $r_i(t)$ represent the integer and fractional part of the quotient t/D_i, respectively. We then have:

$$N_i(t) = \begin{cases} d_i(t) \text{ with probability } 1 - r_i(t) \\ d_i(t) + 1 \text{ with probability } r_i(t). \end{cases}$$

Writing $K_i = N_i - d_i$ $(\in \{0, 1\})$, we can expand (2.4) by conditioning on the possible values of the K_i contributing to the event $N(n - x) = n$:

$$\nu(x) = \sum_{n>x} \left\{ \sum_{\{\Sigma k_i = n - \Sigma d_i\}} \left(\prod_i r_i^{k_i}(1 - r_i)^{1-k_i} \right) \cdot \Pr\{V_{n-x} = 0 \mid K_i = k_i, i = 1, \cdots, S\} \right\}$$

where, for brevity, we omit the argument of d_i, r_i, and K_i. We conjecture an upper bound for $\nu(x)$ by replacing the conditional empty system probability by $(1 - \rho_c)^+$ where ρ_c is the conditional arrival intensity at time $n - x$:

$$\rho_c = \sum_{i=1}^{S} \frac{1 - k_i}{D_i(1 - r_i)}.$$

This should overestimate the empty system probability since, in a rough average sense, the arrival intensity is greater than ρ_c for $t > n - x$. We deduce the bound:

$$\nu(x) \le \sum_{n > x} \left\{ \sum_{\{\Sigma k_i = n - \Sigma d_i\}} \left(\prod_i r_i^{k_i}(1 - r_i)^{1-k_i} \right) \cdot \left(1 - \sum_i \frac{1 - k_i}{D_i(1 - r_i)} \right)^+ \right\} \tag{3.1}$$

Expression (3.1) is particularly complex to evaluate numerically (less complex but looser bounds are derived in [29]) but a probability change argument applied in [37] leads to an accurate approximation allowing rapid calculation. Following the same procedure detailed in [37], we deduce

$$\nu(x) \le \sum_{n > x} \left\{ \frac{\prod_i (1 + r_i z_n - r_i)}{z_n^{n - \Sigma d_i}} \frac{1}{\sqrt{2\pi}\sigma_n} \cdot \left(1 - \sum_i \frac{1}{D_i(1 + r_i z_n - r_i)} \right) \right\} \tag{3.2}$$

where z_n is the root of

$$\sum_i \frac{r_i z}{1 + r_i z - r_i} = n - \sum_i d_i$$

and

$$\sigma_n^2 = \sum_i \frac{r_i(1 - r_i) z_n}{(1 + r_i z_n - r_i)^2}.$$

We refer to [37] for the rather involved derivation of (3.2) whose final justification lies in the excellent numerical results obtained. The approximation technique, consisting in an application of the central limit theorem after a probability change, is used in a different context in Section IV-D.

B. Characterizing the Superposition Process

It would be convenient for the performance evaluation and dimensioning of ATM systems to be able to succinctly characterize a superposition of periodic sources in terms of a small number of parameters. To gain some idea of what parameters are important, we have performed a large number of numerical experiments. The following parameters appear to be among the most significant: multiplex load ($\Sigma\, 1/D_i$); number of multiplexed streams (S); period of the lowest rate stream (max $\{D_i\}$).

In Fig. 3, we compare the 10^{-10} percentile of the virtual waiting time distribution $\nu(x)$ for a variety of different source mixes. (This percentile is assimilated to the required buffer dimension). In the figure, three kinds of mix are differentiated depending on the value of the greatest period: all mixes are composed of a number of streams of period 6.67 and 33.3, together with streams of one other period; this period is 200 for the first kind, 500 for the second, and 1000 for the third. By varying the proportions of streams of the different periods, we vary the total number of multiplexed streams while keeping the multiplex load fixed at 0.9. The figure plots the 10^{-10} percentile against the number of streams. Also plotted is the 10^{-10} percentile obtained when all streams have the same period.

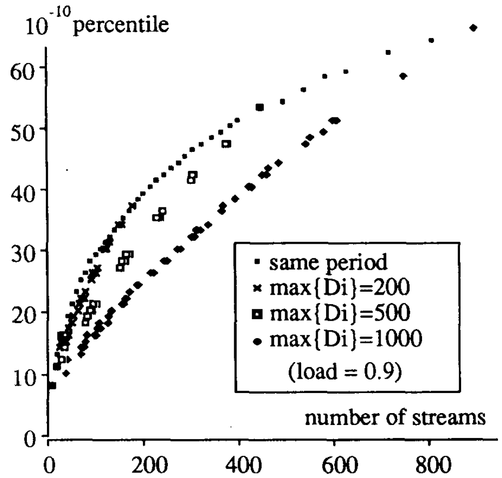


Fig. 3. Buffer dimension (10^{-10} percentile) against number of streams.

We note from these results that, for a given greatest period, the percentile lies roughly on the same curve even though the mix can change quite significantly between neighboring points; buffer requirements increase with the bit rate of the lowest rate streams; a superposition of homogeneous streams yields the greatest buffer requirement.

C. Superposition of Homogeneous Sources

When all streams have the same period ($D_i = D$), we can derive an exact expression for the virtual waiting time distribution using the general relation (2.4) on making the following observation. If $S < D$, there is necessarily some instant in $(-D, 0)$ at which the system is empty and the value of V_0 depends only on the arrivals occurring after such an instant. V_0 thus has the same distribution as a queueing system in which the arrival process consists uniquely of the S arrivals uniformly distributed over $(-D, 0)$. We apply (2.4) to this system writing:

$$\nu(x) = \sum_{x < n \le S} \Pr\left\{N(n - x) = n\right\} \cdot \Pr\left\{V_{n-x} = 0 \mid N(n - x) = n\right\}.$$

The first probability is binomial:

$$\Pr\left\{N(n - x) = n\right\} = \binom{S}{n} \left(\frac{n - x}{D}\right)^n \left(1 - \frac{n - x}{D}\right)^{S-n}.$$

To calculate the second, note that, given $N(n - x) = n$, V_{n-x} depends only on the $S - n$ uniformly distributed arrivals in the interval $(-D, -n + x)$. Reversing the above argument, the distribution of V_{n-x} is the same as in a queue with a superposition of $S - n$ periodic arrival processes of period $D - n + x$. As $-(n - x)$ is an arbitrary instant with respect to this process, we have $\Pr\{V_{n-x} = 0\} = 1 - \rho$ where $\rho = (S - n)/(D - n + x)$ is the server load in the modified queue. We deduce:

$$\nu(x) = \sum_{x < n \le S} \binom{S}{n} \left(\frac{n - x}{D}\right)^n \left(1 - \frac{n - x}{D}\right)^{S-n} \frac{D - S + x}{D - n + x}. \tag{3.3}$$

D. Brownian Bridge Approximation

A convenient analytic approximation is obtained by approximating the process $N(t) - (t/D)S$ by $\sqrt{S} \cdot B(t/D)$ where $B(\cdot)$ is a Brownian bridge, i.e., a continuous Gaussian process on $[0, 1]$ with zero mean and covariance function: $E\{B(u) \cdot B(v)\} = u(1 - v)$, [8]. It is well known that:

$$\Pr\left\{\sup_{0 \le u \le 1} \left\{B(u) - au\right\} > x\right\} = \exp\left(-2x(x + a)\right)$$

[see [34], exercise 2.2.4 and formula (2.2.8)]. Substituting the parameters of the present model and applying (2.1) yields:

$$\nu(x) \approx \exp\left\{-2x\left(\frac{x}{S} + \frac{1-\rho}{\rho}\right)\right\} \quad (3.4)$$

where $\rho = S/D$. For the limiting case $S = D$, we deduce the formulas:

$$E\{V_0\} \approx \sqrt{\frac{\pi S}{8}}; \quad \text{Var}\{V_0\} \approx \frac{4 - \pi}{8} S. \quad (3.5)$$

Sengupta has reported independent and more thorough studies of this approximation and, in particular, proves a heavy traffic limit theorem establishing the accuracy of (3.4) in certain limit conditions [30].

In Table I, we compare exact and asymptotic results for a superposition of streams of period 100 for loads of 1 ($S = 100$) and 0.8 ($S = 80$). The Brownian bridge approximation is good for heavy traffic and has a pleasing simple expression. It may, however, be too inaccurate for dimensioning purposes at moderate loads.

TABLE I
Accuracy of Brownian Bridge Approximation

	$\rho = 1.0$		$\rho = 0.8$	
x	Exact	Approx.	Exact	Approx.
1	9.73e-01	9.80e-01	5.39e-01	5.92e-01
10	1.27e-01	1.35e-01	1.21e-03	5.53e-04
20	2.78e-04	3.36e-04	1.74e-08	2.06e-09
30	8.86e-09	1.52e-08	1.66e-15	5.18e-17

IV. Burst Scale Fluid Models

Fluid models appear particularly attractive for modeling the queueing process in an ATM multiplex when the cell arrival rate varies and can momentarily exceed link capacity. Cell scale queues, such as those considered in the previous section, can then become negligible compared to the queue produced whenever the arrival rate exceeds capacity for any length of time. The behavior of this queue is largely independent of how the cells arrive: it is convenient to suppose they arrive as a fluid.

A. Bounds for General Fluid Input

Relation (2.5) provides a tool for analyzing an infinite capacity queue with fluid input. Let $\psi_t(x, \lambda)$ denote the joint density of $X(t)$ and Λ_t:

$$\psi_t(x, \lambda) = \frac{d^2}{dx\, d\lambda} \Pr\left\{X(t) \le x, \Lambda_t \le \lambda\right\}.$$

Relation (5) can then be written:

$$\nu(x) = \int_{u>0} \int_{\lambda} (1 - \lambda)\psi_u(x, \lambda) \cdot \Pr\left\{V_u = 0 \mid X(u) = x, \Lambda_u = \lambda\right\} d\lambda\, du.$$

We are generally unable to calculate the conditional probability in the above integral. However, since this is certainly less than 1, we have an immediate upper bound:

$$\nu(x) \le \int_{u>0} \int_{\lambda} (1 - \lambda)\psi_u(x, \lambda)\, d\lambda\, du. \quad (4.1)$$

The quality of this bound as an approximation depends on the value of $\Pr\{V_u = 0 \mid X(u) = x, \Lambda_u = \lambda\}$. In most practically interesting cases, we maintain that this will be very close to one. In particular, when we have a high capacity link multiplexing a large number of sources, the system behaves rather like a multiserver system for which we know the empty queue probability to be very much closer to 1 than $1 - \rho$.

Bound (4.1) has been evaluated explicitly when the input rate varies continuously as an Ornstein–Uhlenbeck process [33]. We consider below how it might be applied to a superposition of on/off sources.

B. Generally Distributed On/Off Sources

Consider the particular fluid input process consisting of a superposition of S independent statistically identical on/off sources. In this case, Λ_t takes only discrete values. Let the source input rate, when on, be γ and express (4.1) as:

$$\nu(x) \le \int_{u>0} \sum_{0 \le n < 1/\gamma} (1 - n\gamma)\phi_u(x + u, n)\, du$$

where

$$\phi_t(w, n) = \frac{d}{dw} \Pr\left\{W(t) \le w \text{ and } \Lambda_t = n\gamma\right\} \text{ for } 0 \le n \le S.$$

This integral can readily be evaluated numerically if we can calculate the function $\phi_t(w, n)$. To this end, we distinguish the contribution to the work $W(t)$ arriving in $(-t, 0)$ of sources which are on at $-t$ and that of sources which are off at $-t$. Let $W_i(t)$ be the contribution of source i and define

$$\alpha_t(w) = \frac{d}{dw} \Pr\left\{W_i(t) \le w \text{ and source } i \text{ on at } -t\right\}$$

$$\beta_t(w) = \frac{d}{dw} \Pr\left\{W_i(t) \le w \text{ and source } i \text{ off at } -t\right\}.$$

Thus, $W(t) = \Sigma\, W_i(t)$ and we can express $\phi_t(w, n)$ as:

$$\phi_t(w, n) = \binom{S}{n} \alpha_t^{*(n)} * \beta_t^{*(S-n)}(w) \quad (4.2)$$

where * denotes convolution.

To derive the densities α_t and β_t we must, of course, make some more assumptions about the individual source processes. For example, the case where the succession of on and off periods constitutes an alternating renewal process is considered in [3].

C. Shifted Normal Approximation

In fact, whatever the source input process, we will have trouble exactly evaluating the convolution in (4.2). Fortunately, the shifted (or tilted) normal approximation (see [12], p. 188), successfully used to derive expression (3.2) for the $\Sigma\, D_i/D/1$

queue [37], proves particularly convenient here. We present this technique with reference to the on/off source model of Section IV-C.

First, introduce the Laplace transforms $\phi_t^*(s, n)$, $\alpha_t^*(s)$, and $\beta_t^*(s)$:

$$\phi_t^*(s, n) = \int \phi_t(w, n) e^{-sw}\, dw, \text{ etc.}$$

Taking transforms in (4.2), we deduce

$$\phi_t^*(s, n) = \binom{S}{n} \alpha_t^*(s)^n\, \beta_t^*(s)^{S-n}. \tag{4.3}$$

Now, define the shifted density

$$\phi_t^{(s)}(w, n) = \frac{\phi_t(w, n) e^{-sw}}{\phi_t^*(s,n)} \tag{4.4}$$

which is the convolution of similarly transformed densities $\alpha_t^{(s)}$ (n times) and $\beta_t^{(s)}$ ($S - n$ times). The shifted normal approximation technique consists in replacing $\phi_t^{(s)}$ by a central limit approximation and inverting (4.4) to provide an estimate of ϕ_t. As the Gaussian approximation is most accurate about the distribution mean, to estimate $\phi_t(w, n)$, we choose s so that w is precisely the mean of $\phi_t^{(s)}(w, n)$, i.e., s is the root of

$$w = \frac{\int u \phi_t(u, n) e^{-su}\, du}{\phi_t^*(s, n)}$$

which may be written:

$$w = -\left(n \frac{d\alpha^*}{ds}\Big/\alpha^* + (S - n) \frac{d\beta^*}{ds}\Big/\beta^* \right).$$

The normal approximation for $\phi_t^{(s)}$ is then:

$$\phi_t^{(s)}(w, n) \approx \frac{1}{\sqrt{2\pi}\sigma_s}$$

where σ_s^2 is the variance of $\phi_t^{(s)}$

$$\sigma_s^2 = n\left(\frac{d^2\alpha^*}{ds^2}\Big/\alpha^* - \left(\frac{d\alpha^*}{ds}\Big/\alpha^* \right)^2 \right) + (S - n)\left(\frac{d^2\beta^*}{ds^2}\Big/\beta^* - \left(\frac{d\beta^*}{ds}\Big/\beta^* \right)^2 \right).$$

Finally then, we have an approximation depending only on the Laplace transforms:

$$\phi_t(w, n) \approx \frac{e^{sw}\phi_t^*(s, n)}{\sqrt{2\pi}\sigma_s}.$$

Results of implementing this method are presented in [3].

V. Integrating Burst and Cell Scale Congestion

The fluid model was introduced in Section IV, as an approximation, valid when congestion is mainly due to the burst scale arrival rate exceeding multiplex capacity. It is more satisfying to recognize the fluid queue models as a means for evaluating a "burst component" of the real queue. To evaluate the exact multiplexer performance, we must add to this a further component due to cell scale fluctuations about the burst scale average arrival rate.

A. Coupling Cell Process and Fluid Process

From the point of view of the multiplex, the only arrival process is the discrete cell arrival process. With respect to this process, it is possible to define any number of "equivalent fluid processes." The nature of this ambiguity is clarified with reference to the following definition.

A *stationary conservative packetization scheme* is a triple $\langle (\Gamma_t), (T_n), U \rangle$ where

i) (Γ_t) is a nonnegative stationary process on $(-\infty, \infty)$;
ii) (T_n) is a stationary point process on $(-\infty, \infty)$;
iii) U is a random variable with values in $[0, 1]$;
iv) (Γ_t), (T_n), and U are defined on the same probability space and are coupled by the relations:

$$T_0 = \inf\left\{ t \geq 0: U + \int_0^t \Gamma_s\, ds > 1 \right\}, \text{ and for all } n > 0,$$

$$T_{n+1} = \inf\left\{ t > T_n: \int_{T_n}^t \Gamma_s\, ds > 1 \right\},$$

$$T_{-n-1} = \sup\left\{ t < T_{-n}: \int_t^{T_{-n}} \Gamma_s\, ds \geq 1 \right\}.$$

The stationarity requirement allows us to consider 0 as a random time point for both processes simultaneously. The random variable U may be interpreted as the fullness of the packetizer buffer at time 0. By "conservative," we imply that the scheme preserves the volume of work produced in the sense that:

$$\int_{T_n}^{T_{n+1}} \Gamma_t\, dt = 1, \text{ for all } n.$$

The random variable U determines the *phase* of (T_n) with respect to (Γ_t).

The following examples illustrate this definition with reference to the relation between a cell stream and a fluid process.

1) Random phase coupling: if U is uniform on $[0, 1]$ and independent of (Γ_t), we say (Γ_t) and (T_n) are coupled with random phase. (It may readily be shown that, given a stationary fluid process (Γ_t), it is always possible to construct a stationary equivalent point process (T_n) by random phase coupling).

2) General synchronous coupling: given (T_n), a stationary fluid process can always be defined by:

$$\Gamma_t = \frac{1}{T_{n+1} - T_n} \text{ for } t \in [T_n, T_{n+1}).$$

3) Synchronous coupling of on/off sources: consider the important case where the cell process consists of bursts with constant bit rate; the distance between consecutive cells is D, say, inside a burst and greater than D between bursts; the most natural fluid process is $\Gamma_t = 1/D$ if $t \in [T_n, T_n + D)$ for some n and $\Gamma_t = 0$ otherwise. In this case, U depends on (Γ_t) and, in particular, $U = 1$ if $\Gamma_0 = 0$.

The difference between two packetizations of the same fluid stream over any time interval (difference determined by the values of U) is at most one cell so that, from a practical point of view, we can choose the scheme which is most convenient for the purpose at hand. If we are given a cell process, it seems most natural to define an associated fluid stream by 2) or 3) above. If, on the other hand, we wish to define a cell process for a given fluid process, it would be more appropriate to assume random phasing.

B. Burst Component and Cell Component

Consider the queue arising in a multiplexer due to the superposition of S cell arrival streams, each resulting from an independent stationary conservative packetization scheme $\langle (\Gamma_t^i), (T_n^i), U^i \rangle$. Let $\Lambda_t = \Sigma \Gamma_{-t}^i$ be the fluid arrival rate at time $-t$ and define the fluid excess with respect to time 0:

$$X_b(t) = \int_0^t \Lambda_u \, du - t.$$

Divide the virtual waiting time V_t into burst and cell components, B_t and C_t, where B_t is the virtual waiting time obtained with the fluid arrival process Λ_t and C_t is defined simply as the difference $C_t = V_t - B_t$. Consider the value of V_t and its components at the arbitrary instant 0. We assume it is possible to determine the distribution of B_0 (by the methods of Section IV, for example) and seek to characterize the additional term C_0.

First note that C_0 is the difference between the maxima of two interrelated processes: $X(t)$ and $X_b(t)$ (see Fig. 4).

We have the identity, for $t > 0$,

$$V_0 = B_0 + C_0 = B_t + C_t + X(t) + I(t) \tag{5.1}$$

where $I(t)$ is the cumulative idle time in $(-t, 0)$:

$$I(t) = \int_0^t 1\{V_u = 0\} \, du.$$

Let τ be the smallest t such that $X_b(\tau)$ realizes the maximum of the process $X_b(t)$. We then have the properties (see figure):

i) $B_0 = X_b(\tau)$;

ii) $B_\tau = 0$ and $B_t > 0$ for $0 \le t < \tau$;

iii) $\Lambda_\tau \le 1 \le \Lambda_{\tau-}$.

Substituting τ for t in (5.1), we deduce:

$$C_0 = C_\tau + X(\tau) - X_b(\tau) + I(\tau). \tag{5.2}$$

Case $B_0 > 0$: In the case $B_0 > 0$, we have $\tau > 0$. Now, $I(\tau)$ is not necessarily zero, although the burst component is always positive in $(-\tau, 0)$ but we may reasonably expect it to be negligible. C_τ is the virtual waiting time at the moment when the burst component busy period began. In the particular case of on/off sources, it may be recognized that C_τ is the virtual waiting time of a $\Sigma\, D_i/D/1$ queue with load equal or close to one (by property iii above). For sources with output of forms b) or c) in Fig. 1, this is also approximately true.

The difference $X(\tau) - X_b(\tau)$ is not independent of the value of C_τ and we can say little about their joint distribution. If, however, we assume a random phase coupling between cell and burst processes, we can conclude that $X(\tau) - X_b(\tau)$ has zero expectation for any of the source characteristics depicted in Fig. 1. Let $N^i(t)$ be the number of cell emissions of stream i in $(-t, 0)$:

$$N^i(t) = \sup\{n \ge 0: T_{-n}^i > -t\} = \left[\int_{-t}^0 \Gamma_s^i \, ds + 1 - U\right]$$

where $[x]$ denotes the integer part of x.

With random phasing (i.e., U uniform on $[0, 1]$), we thus have

$$\mathrm{E}\{N^i(t)\} = \mathrm{E}\left\{\int_{-t}^0 \Gamma_s^i \, ds\right\}.$$

That $\mathrm{E}\{X(t)\} = \mathrm{E}\{X_b(t)\}$ follows on noting that $X(t) = \Sigma N^i(t) - t$. For the particular instant τ, we also have $\mathrm{E}\{X(\tau)\} = \mathrm{E}\{X_b(\tau)\}$ since τ is independent of the U^i.

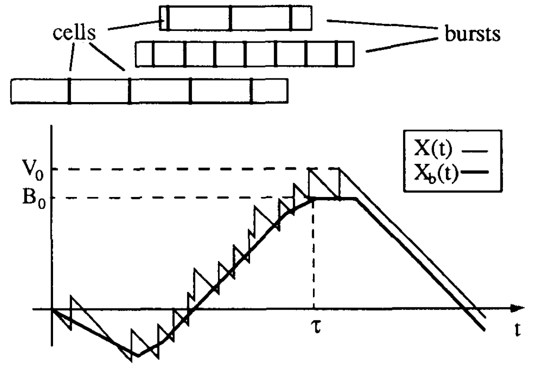


Fig. 4. A realization of processes $X(t)$ and $X_b(t)$.

We conclude from (5.2), therefore, that

$$\mathrm{E}\{C_0 | B_0 > 0\} \approx \mathrm{E}\{C_\tau | B_0 > 0\} \approx \mathrm{E}\{C_t | \Lambda_t = 1\} \tag{5.3}$$

i.e., the expected value of the cell component is approximately equal to the mean of a $\Sigma\, D_i/D/1$ queue of load 1. This is, in general, rather small compared to the value of B_0 as illustrated in the numerical example in Section VI-B below. Its variance is also rather small: in the case of homogeneous on/off sources, (3.5) provides the estimate $\mathrm{Var}\{C_\tau | B_0 > 0\} \approx (4 - \pi)S/8$; it may also be shown that, with random phase coupling, $X(t) - X_b(t)$ has a variance of $1/6$ times the number of sources active in $(-t, 0)$, i.e., $\mathrm{Var}\{X(t) - X_b(t)\} < S/6$.

Case $B_0 = 0$: When the burst component is zero, the behavior of the multiplex queue is qualitatively different, being determined by cell emissions in a relatively short interval before considered time 0. This behavior is most clearly defined in the case of homogeneous on/off sources where the virtual waiting time at time 0 depends only on the cell emissions within the interval $(-D, 0)$. Let M_0 be the number of sources which emit a cell in this interval. The conditional virtual waiting time distribution can be determined using the formulas of Sections III-C and III-D above. Of course, the distribution of M_0, conditioned on $B_0 = 0$, remains to be determined.

In the case of other types of variable bit rate sources, the queueing process is approximately that of a $\Sigma\, D_i/D/1$ system where the joint distribution of the D_i is conditioned on the fact that $B_0 = 0$. In most cases, for practical purposes, the condition $B_0 = 0$ can be replaced by the simpler condition $\Lambda_0 < 1$.

VI. Application to a Superposition of On/Off Sources

The general formula:

$$\nu(x) = \Pr\{C_0 > x | B_0 = 0\} \cdot \Pr\{B_0 = 0\} + \Pr\{B_0 + C_0 > x | B_0 > 0\} \cdot \Pr\{B_0 > 0\} \tag{6.1}$$

is a useful decomposition of the queueing process. However, it is clear that, in practice, it is unlikely that we need to take account of congestion in both burst and cell time scales at the same time. We first evaluate the queue length distribution when the probability of burst congestion is negligible, and then consider the case of packetized voice where the relative significance of the two components depends on the number of active sources.

A. Negligible Burst Component

We consider a superposition of S homogeneous on/off sources of period D, assuming the burst component of the multiplex queue is negligible: $\Pr\{B_0 > 0\} < \epsilon$ where ϵ is an order of magnitude smaller than some cell loss rate requirement (e.g., $\epsilon = 10^{-9}$). Assume for present purposes that cells and bursts are related as in example 3 of Section V-A, i.e., each burst begins with a cell emission and ends D time units after the last cell emission. This means that bursts in progress at time 0, and only these, emit a cell in the interval $(-D, 0)$. We denote this number by M_0. Let p be the ratio of the average burst length to the average burst silence cycle. Then p is also the probability a burst is on at 0 and we have:

$$\Pr\{M_0 = m\} = \binom{S}{m} p^m (1-p)^m.$$

By (6.1), we can write: $\nu(x) < \Pr\{C_0 > x \text{ and } B_0 = 0\} + \epsilon$. As noted above, $\Pr\{C_0 > x \text{ and } B_0 = 0\} = \Pr\{C_0' > x \text{ and } B_0 = 0\}$ where C_0' is the virtual waiting time due solely to the M_0 arrivals in $(-D, 0)$. Furthermore, since $\{M_0 \le D\}$ includes $\{B_0 = 0\}$, we can write:

$$\Pr\{C_0' > x \text{ and } B_0 = 0\} < \Pr\{C_0' > x \text{ and } M_0 \le D\}$$
$$= \sum_m \Pr\{C_0' > x \mid M_0 = m\} \binom{S}{m} p^m (1-p)^m.$$

With the above conditional probability given by the results of Section III-C, we deduce an approximation for $\nu(x)$ which is an upper bound for probabilities an order of magnitude greater than ϵ:

$$\nu(x) \approx \sum_{x < m \le D} \binom{S}{m} p^m (1-p)^{S-m} \sum_{x < n \le m} \binom{m}{n} \cdot \left(\frac{n-x}{D}\right)^n \left(1 - \frac{n-x}{D}\right)^{m-n} \frac{D-m+x}{D-n+x}. \tag{6.2}$$

It is interesting to verify how close to this distribution is the virtual waiting time distribution of an $M/D/1$ queue with equivalent load. Given D and p, we have calculated S so that $\Pr\{M_0 \ge D\} < 10^{-9}$ ($\Pr\{B_0 > 0\}$ is thus somewhat greater than 10^{-9}) and estimated $\nu(x)$ by (6.2). Results for $D = 100$ and $p = 0.5$ and 0.1 are shown in Fig. 5.

It is apparent that, with the condition of negligible burst scale congestion, the assumption of Poisson arrivals overestimates the virtual waiting time but constitutes a good approximation for a superposition of very bursty sources ($p = 0.1$). It should, in general, lead to good estimations of the first moments of the virtual waiting time distribution.

If $S < D$, we have $B_0 \equiv 0$ (and M_0 identically less than D) and expression (6.2) is exact (interpreting the binomial coefficients to be zero when $m > S$). Changing the order of summation and simplifying, we deduce a simpler expression:

$$\nu(x) = \sum_{x < n \le S} \binom{S}{n} \left(\frac{n-x}{D} p\right)^n \left(1 - \frac{n-x}{D} p\right)^{S-n} \cdot \frac{D - p(S+x)}{D - p(n+x)}. \tag{6.3}$$

Note that this relation is the same as (3.3) with D replaced by D/p. It is independent of any assumptions about the distribution of burst and silence lengths other than the requirement that the intercell interval be at least equal to D. This observation could, in fact, have been deduced from a more general result (Corollary 2.4.5) proved in [10].

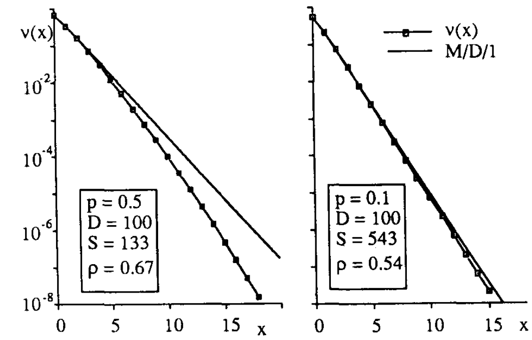


Fig. 5. Distribution $\nu(x)$ for negligible burst congestion.

B. Packetized Voice

The particular example of a superposition of packetized voice sources, introduced in [35], constitutes a useful benchmark for evaluating the accuracy of different approaches to modeling queues with a correlated arrival process (see [14], [15]). A 1.536 Mb/s link multiplexes voice sources coded at 32 Kb/s and packetized with silence suppression, each packet containing 512 bits. Mean talkspurt length is 352 ms and mean silence length 650 ms. The number of cells in a talkspurt is geometrically distributed with mean 22 and the silence has an exponential distribution. The number of multiplexed sources is a model parameter, less than or equal to 136 to ensure stability. The link is momentarily saturated when 48 or more sources are simultaneously active.

To estimate the mean virtual delay V_0 using the present approach, we estimate separately the mean burst component and the mean cell component. For the former, we have used the fluid model of [1] assuming, therefore, that talkspurt length has the exponential distribution. Note that the approach described in Section IV above provides a good estimate of the queue length distribution tail but is not suitable here for estimating mean delays.

The expected cell component is given by:

$$E\{C_0\} = E\{C_0 \mid B_0 = 0\} \cdot \Pr\{B_0 = 0\} + E\{C_0 \mid B_0 > 0\} \cdot \Pr\{B_0 > 0\}.$$

Applying the results of the previous sections, we approximate this by:

$$E\{C_0\} \approx \sum_{0 \le m \le 48} E\{C_0 \mid M_0 = m\} \cdot \Pr\{M_0 = m\} + E\{C_0 \mid M_0 = 48\} \cdot \Pr\{M_0 > 48\}.$$

The condition $M_0 = 48$ in the second term refers, of course, to the system state at the onset of the current burst scale busy period.

Results of applying this method are compared in Table II to the results of simulations taken from [35]. We have converted our virtual waiting time results to real waiting times by calculating the mean queue length in cells (the queue length distribution being determined from $\nu(x)$ at integer arguments) and

TABLE II
MEAN DELAY FOR A SUPERPOSITION OF VOICE SOURCES

	Delay Components		Mean Delay	
Sources	E { C_0 }	E { B_0 }	E { $C_0 + B_0$ }	Simul ± 95%
80	0.22	0.00	0.22	0.22 ± .007
110	0.55	0.35	0.90	0.89 ± .14
125	0.82	9.77	10.59	10.4 ± 1.3
132	0.93	56.59	57.52	52.1 ± 7.5
134	0.95	121.9	122.85	110 ± 21.4

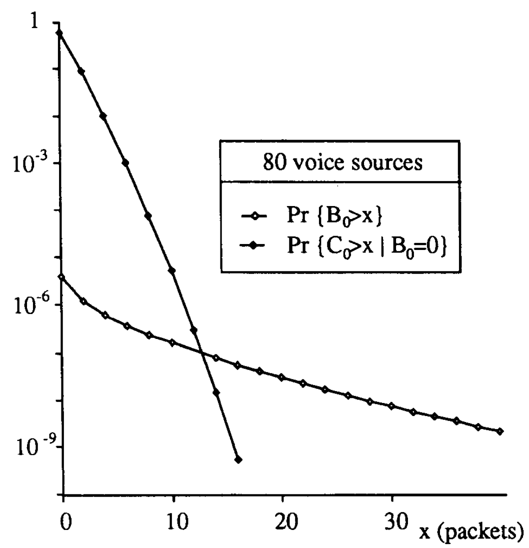


Fig. 6. Burst and cell components of virtual waiting time distribution.

applying Little's result. Waiting times are expressed in milliseconds. Results are visibly in good agreement.

The relative significance of the two components C_0 and B_0 is further illustrated in Fig. 6. This figure depicts the distribution of B_0 (calculated by the results of [1]) and the conditional distribution of C_0, given $B_0 = 0$, [approximated by (6.2)] in the case of 80 voice sources. The virtual waiting time distribution is practically equal to the maximum of these two distributions (we should, however, add a cell component of mean 4.2 cells in the case $B_0 > 0$). The importance of the burst component in calculating the low percentiles necessary for buffer dimensioning is evident, despite the very low probability of a positive burst component ($\Pr \{B_0 > 0\} < 10^{-5}$).

VII. CONCLUSIONS

To account for the correlations arising in the cell arrival process when variable bit rate sources are multiplexed in an ATM network, we have made use of a general result for the $G/G/1$ queue due to Beneš. We have given an intuitive demonstration of this result and applied it to determine the distribution of a multiplex queue first, when the arrival process is a superposition of periodic sources (the $\Sigma D_i/D/1$ queue) and second, when the arrival process is assimilated to a random intensity fluid input.

The fluid approximation for the superposition queue can be more satisfactorily viewed as a way to calculate the "burst component" of the real queue, to which must be added a further "cell component" depending on local fluctuations in the arrival rate about the fluid average. The cell component is shown to behave like a $\Sigma D_i/D/1$ queue when the burst component is zero, and to constitute a small positive bias when the latter is positive. The expected value of this bias is approximately equal to the mean of a $\Sigma D_i/D/1$ queue with load equal to 1 (this being finite and generally much smaller than the burst component).

The decomposition into cell and burst components clearly shows the relative importance of correlations at cell scale (due to locally periodic cell emissions by active sources) and at burst scale (due to the slowly varying longer-term arrival rate). In particular, we observe that the first moments of the delay distribution may depend significantly on the cell component while the low percentiles necessary for buffer dimensioning are given essentially by the burst component alone.

REFERENCES

[1] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data handling system with multiple resources," *Bell Syst. Tech. J.*, vol. 61, no. 8, pp. 1871-1894, 1982.
[2] V. E. Beneš, *General Stochastic Processes in the Theory of Queues*. Reading, MA: Addison Wesley, 1963.
[3] B. Bensaou, J. Guibert, and J. W. Roberts, "Fluid approximation for a superposition of on/off sources," in *Proc. ITC Sem.*, Morristown, NJ, Oct. 1990.
[4] A. Bhargava, P. Humblet, and M. G. Hluchyj, "Queueing analysis of continuous bit stream transport in packet networks," presented at GLOBECOM '89, Dallas, TX, Oct. 1989.
[5] A. A. Borovkov, *Stochastic Processes in Queueing Theory*. New York: Springer Verlag, 1976.
[6] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. Englewood Cliffs, NJ: McGraw-Hill, 1978.
[7] J. N. Daigle and J. D. Langford, "Models for analysis of packet voice communications systems," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 856-868, Sept. 1986.
[8] M. D. Donsker, "Justification and extension of Doob's heuristic approach to Kolmogorov-Smirnov theorems," *Ann. Math. Statist.*, vol. 23, pp. 277-281, 1952.
[9] A. E. Eckberg, "The single server queue with periodic arrival process and deterministic service time," *IEEE Trans. Commun.*, vol. 27, pp. 556-562, Mar. 1979.
[10] —, "Response time analysis for pipelining jobs in a tree network of processors," in *Applied Probability-Computer Science: The Interface Vol. 1*, R. L. Disney and T. J. Ott, Eds. Boston, MA: Birkhäuser, 1982.
[11] A. I. Elwalid, D. Mitra, and T. E. Stern, "A theory of statistical multiplexing of Markovian sources: Spectral expansions and algorithms," in *Proc. 1st Int. Workshop Numer. Solut. Markov Chains*, Raleigh, NC, Jan. 1990.
[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[13] A. Gravey, "Temps d'attente et nombre de clients dans une file nD/D/1," *Annales de l'Institut H. Poincaré-Probabilities et Statistiques*, vol. 20, no. 1, pp. 53-73, 1984.
[14] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 856-868, Sept. 1986.
[15] I. Ide, "Superposition of interrupted Poisson processes and its application to packetized voice multiplexers," in *Proc. Int. Teletraffic Cong. 12*, Turin, Italy, May 1988, paper 3.1B2.
[16] S. B. Jacobsen and L. Dittmann, "A fluid flow queueing model for heterogeneous on/off traffic," submitted for publication.
[17] C. Knessl and J. A. Morrison, "Heavy traffic analysis of a data handling system with many sources," submitted for publication.
[18] L. Kosten, "Stochastic theory of a multi-entry buffer (I)," Delft Progress Rep. Series F, vol. 1, pp. 10-18, 1974.

[19] ——, "Stochastic theory of a multi-entry buffer (II)," Delft Progress Rep., Series F, vol. 1, pp. 44-50, 1974.
[20] L. Kosten and O. Vreize, "Stochastic theory of a multi-entry buffer (III)," Delft Progress Rep, Series F, vol. 1, pp. 103-115, 1975.
[21] L. Kosten, "Stochastic theory of data handling systems with groups of multiple sources," in *Proc. Int. Symp. Perform. Comput. Commun. Syst.*, IFIP, Zürich, Mar. 1984.
[22] ——, "Liquid models for a type of information buffer problems," Delft Progress Rep., vol. 11, pp. 71-86, 1986.
[23] S. Q. Li, "Performance of a non-blocking space division packet switch with correlated input traffic," presented at GLOBECOM '89, Dallas, TX, Oct. 1989.
[24] K. Q. Liao and L. G. Mason, "A discrete time single server queue with a two-level modulated input and its applications," presented at GLOBECOM '89, Dallas, TX, Oct. 1989.
[25] T. J. Ott and J. G. Shanthikumar, "On a buffer problem for packetized voice with an N-periodic strongly interchangeable input process," Bellcore Rep., July 1989.
[26] V. Ramaswami, "Traffic performance modeling for packet communication: Whence, where and whither," presented at the Third Australian Teletraffic Sem., Melbourne, Australia, Nov. 1988, keynote address.
[27] M. H. Rossiter, "A switched Poisson model for data traffic," *Aus. Tele. Rev.*, vol. 21, no. 1, 1987.
[28] G. Ramamurthy and B. Sengupta, "Delay analysis of a packet voice multiplexer by the $\Sigma D_i/D/1$ queue," Bell Labs Rep., 1988.
[29] J. W. Roberts and J. T. Virtamo, "The superposition of periodic cell arrival processes in an ATM multiplexer," *IEEE Trans. Commun.*, Jan. 1991.
[30] B. Sengupta, "The $\Sigma D_i/G/1$ queue and an asymptotic analysis," Bell Labs Tech. Memo, Oct. 1989.
[31] A. Simonian, "Stationary analysis of a fluid queue with input rate varying as an Ornstein-Uhlenbeck process," to appear in *SIAM J. Appl. Math.*
[32] P. Sen, B. Maglaris, N-E. Rikli, and D. Anastassiou, "Models for packet switching of variable bit rate sources," *IEEE J. Select. Areas Commun.*, vol. 7, no. 5, pp. 865-869, June 1989.
[33] A. Simonian and J. T. Virtamo, "Transient and stationary distributions for fluid queues and input processes with a density," to appear in *SIAM J. Appl. Math.*
[34] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*. New York: Wiley, 1986.
[35] K. Sriram and W. Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE J. Select. Areas Commun.*, vol. 4, no. 6, pp. 833-846, Sept. 1986.
[36] R. C. Tucker, "Accurate method for analysis of a packet speech multiplexer with limited delay," *IEEE Trans. Commun.*, vol. 36, no. 4, Apr. 1988.
[37] J. T. Virtamo and J. W. Roberts, "Evaluating buffer requirements in an ATM multiplexer," presented at GLOBECOM '89, Dallas, TX, Oct. 1989.
[38] A. K. Wong, "Queueing analysis for ATM switching for continuous bit rate oriented traffic—A recursion method," Bell Labs Rep., 1989.

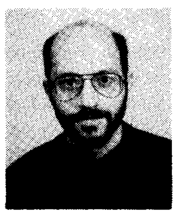
Ilkka Norros received the Ph.D. degree from the University of Helsinki, Helsinki, Finland, in 1986.

He was an Assistant in the Department of Mathematics, University of Helsinki, from 1977 to 1988. His research involved the application of martingale techniques to reliability theory. Since 1988, he has been employed as a Senior Research Scientist in the Telecommunications Laboratory, Technical Research Centre of Finland. His current research interests are stochastic processes in telecommunication systems, information theory, and reliability theory.

James W. Roberts received the B.Sc. degree in mathematics from the University of Surrey, Surrey, UK, in 1969 and received the Ph.D. degree from the Université Pierre et Marie Curie, Paris, France, in 1987.

Since graduating, he has worked in the field of teletraffic engineering and performance evaluation, first with the British Post Office from 1970 to 1975, with the French company SOCOTEL from 1975 to 1977, and with the Centre National d'Etudes des Télécommunications since 1978. His current research activity concerns the design and performance evaluation of broadband multiservice networks.

Dr. Roberts is the Chairman of the European project COST 224 on multiservice network design and performance evaluation.



Alain Simonian received degrees from the Ecole Polytechnique, Paris, France, in 1982 and from the Ecole Nationale Supérieure des Télécommunications in 1984.

Since 1984, he has been working in the Centre National d'Etudes des Télécommunications in the field of telecommunications network performance evaluation. His current research is concerned with the application of large deviation theory and stochastic analysis to the performance evaluation of broadband multiservice networks.

Jorma T. Virtamo received the D.Tech. degree in theoretical physics from the Helsinki University of Technology, Helsinki, Finland, in 1976.

In 1979-1980, he was a Visiting Scientist at the Max Planck Institute für Plasmaphysik, West Germany. In 1980, he joined Neste Oy, a Finnish industrial corporation, where he was engaged in exploratory R&D work. Since 1986, he has been with the Telecommunications Laboratory, Technical Research Centre of Finland, where he has performed research on video coding, data encryption, and teletraffic theory.