Characterizing the Variability of Arrival Processes with Indices of Dispersion

Riccardo Gusella

International Computer Science Institute
1947 Center Street, Suite 600
Berkeley, California 94704

TR-90-051
September 26, 1990

ABSTRACT

We propose to characterize the burstiness of packet arrival processes with indices of dispersion for intervals and for counts. These indices, which are functions of the variance of intervals and counts, are relatively straightforward to estimate and convey much more information than simpler indices, such as the coefficient of variation, that are often used to describe burstiness quantitatively.

We define and evaluate the indices of dispersion for some of the simple analytical models that are frequently used to represent highly variable processes. We then estimate the indices for a number of measured point processes that were generated by workstations communicating to file servers over a local-area network.

We show that nonstationary components in the measured packet arrival data distort the shape of the indices and propose ways to handle nonstationary data. Finally, to show how to incorporate measures of variability into analytical models and to offer an example of how to model our measured packet arrival processes, we describe a fitting procedure based on the index of dispersion for counts for the Markov-modulated Poisson process.

I. INTRODUCTION

Since the first analyses of computer traffic in the mid- and late 1970s [10, 19], which showed that packet arrival processes are highly variable, researchers have frequently described computer communication patterns as ''bursty''. Yet few have bothered to define burstiness. Most seem to invoke the term bursty when confronted with processes whose interarrival time distributions show greater variability than Poisson processes. (Exceptions do exist; see for instance [13], in which a distribution with coefficient of variation 0.3 is termed bursty.) The vagueness surrounding the concept of burstiness stems both from its use to denote different types of variability in many disparate situations and from the difficulty of characterizing in meaningful ways the capricious nature of packet arrivals.

The variability of packet arrival processes is strikingly manifested in the following Figures 1 and 2, which represent respectively the times between subsequent arrivals and the times between every four arrivals for the messages sent by a single-user workstation to its file server over a local-area network. In each figure, the logarithm of the interarrival time is on the ordinate and the serial number on the abscissa, and artificial zero lines have been placed at the median values of the series. This graphical arrangement allows us to view short interarrival times, which would otherwise be obscured by long ones: the former are printed toward the bottom of the graph and the latter toward the top.

FIGURE 1. TIMES BETWEEN SUCCESSIVE ARRIVALS

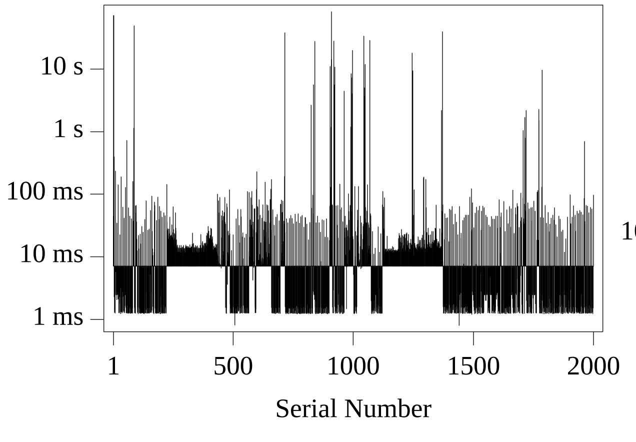
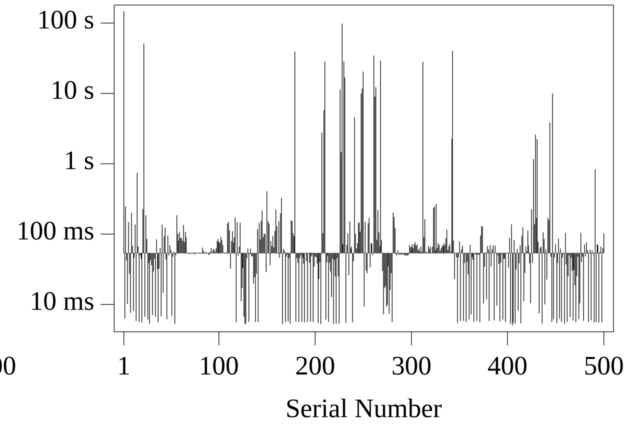


FIGURE 2. TIMES BETWEEN EVERY FOUR ARRIVALS



The variance of the interarrival times in Figure 2 is about six times larger than the variance of those in Figure 1. Clearly, the larger dispersion of values in the second figure stems from the clustering of small and large interarrival times in separate groups in the first figure. This bunching is caused by protocol features, such as fragmentation of large messages, that generate short interarrivals in batches. Similarly, a sequence of related remote procedure calls may generate a group of longer interarrivals because, for each request, the destination has to execute a procedure before returning the answer, which in turn will trigger the next request. Thus, the increase in variance is related to the temporal structure in the data and is not captured by simple burstiness indices such as the coefficient of variation, the peak-to-average ratio, an so on. In this paper we will use the term variability to refer explicitly to changes in the variance of the sum of consecutive interarrivals or in the variance of arrival counts over larger and larger time intervals.

Variability in packet arrivals has been connected to the queuing delays packets are subject to: the general rule is that more variability corresponds to longer delays. However, except in a few simple cases, the precise relationship between variability and queuing delays is difficult to represent analytically. Several attempts to resolve this issue have been made; in particular, Fendik and Whitt's approach [5], which uses a statistical index that models the variability of the arrival process but also captures the dependency between the interarrival and the service times, is worth mentioning.

In this paper we take a more narrow, focused approach: we characterize the variability of measured packet arrival processes with indices of dispersion functions and discuss the merits of these indices as well as the pitfalls of their indiscriminate use. Indices of dispersion have long been known in the statistics community as a powerful tool in the analysis of the second-order properties of point processes [1, 3], but, despite the flourishing in recent years of measurements and analyses of computer traffic data (for a survey see [16]), they have been rarely, if ever, applied to computer traffic measurements. Here, we demonstrate that indices of dispersion are valuable and valid tools for characterizing the variability of packet arrival processes. We also discuss how standard analytical models should be fitted to traffic measurements in order to take into account the variability of the data.

This paper is organized into three major parts. In Section II, we define the index of dispersion for intervals and the index of dispersion for counts, and review their basic properties. We then calculate one of these two indices for each of three classes of analytical models that are often used to represent bursty point processes: renewal models with hyperexponential interarrival times, batch Poisson processes, and Markov-modulated Poisson processes. Although the results are in most cases not new, the exercise serves the important purpose of clarifying the meaning and use of indices of dispersion. In Section III, we estimate the indices of dispersion for several measured packet arrival processes generated by single-user workstations communicating with file servers over a local-area network. We show how nonstationary data introduce difficulties, and suggest that semi-Markov models may model accurately both short- and long-term variability. Finally, in Section IV, as an example of how standard arrival models can incorporate the variability that we have analyzed, we develop a procedure to fit a Markov-modulated Poisson process to our arrival processes.

II. Indices of Dispersion

A. The Index of Dispersion for Intervals

Let us consider first describing point processes in terms of the lengths of the intervals between subsequent arrivals. For packet-arrival processes, these intervals, which we will call *interarrival times*, are defined as the length of time between the beginning of the transmission of a given packet and the beginning of the transmission of the previous packet. (Cox and Miller [2, page 339] define the same quantity more esoterically as *backward recurrence-time*.) Notice that under this definition the transmission time of the previous packet is included in the interarrival time; thus, for networks that allow variable packet sizes, this definition introduces a source of dependency between intervals and packet lengths. (A constant packet size imposes a lower bound on interarrival times, but not the dependency.)

The variance of the sum of two random variables depends on the covariance between them, and, if they have common variance, is given by

$$\mathrm{var}(X_{i+1} + X_{i+2}) = 2\mathrm{var}(X) + 2\mathrm{cov}(X_{i+1}, X_{i+2}) ,$$

and, in general, for the sum of n variables we have

$$\mathrm{var}(X_{i+1} + \cdots + X_{i+n}) = n\ \mathrm{var}(X) + 2\sum_{j=1}^{n-1}\sum_{k=1}^{j}\mathrm{cov}(X_j, X_{j+k}) . \qquad (1)$$

We have indicated with $\mathrm{var}(X)$ the common variance of the X_i (we will also write $E(X)$ for the common mean), and thus have assumed implicitly that the processes under consideration are at least weakly stationary, i.e., that their first and second moments are time invariant, and that the autocovariance series depends only on the distance k, the *lag*, between samples: $\mathrm{cov}(X_i, X_{i+k}) = \mathrm{cov}(X_j, X_{j+k})$, for all i, j, and k.

It is the dependency on the autocovariance, or, equivalently, on the autocorrelation, that makes the variance of the sum of intervals useful in describing arrival processes. In fact, in situations like those of Figure 1 and Figure 2, in which interarrivals smaller than the mean as well as interarrivals larger than the mean are grouped together, the covariance will assume positive values.

We will use the variance above, normalized by the factor $n\,E^2(X)$, as a measure of the variability of packet arrival processes. The sequence of values

$$J_n = \frac{\mathrm{var}(X_{i+1} + \cdots + X_{i+n})}{n\,E^2(X)} , \qquad (2)$$

with $n = 1, 2, \cdots,$ is called *index of dispersion for intervals* (IDI).

Notice that J_1 is $C_J^2 = \mathrm{var}(X)/E^2(X)$, the squared coefficient of variation for intervals. As a result of the normalization, for a Poisson process J_n has constant value 1 for all n; for a renewal process, whose interarrival times are identical and independently distributed (i.i.d.), J_n is also a constant in n of value C_J^2. Using equation (1) and the definition $\rho_n = \mathrm{cov}(X_i, X_{i+n}) / \mathrm{var}(X)$, we can express relation (2) in terms of the autocorrelation coefficients at lag n:

$$J_n = C_J^2 \left[1 + 2\sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)\rho_j \right] , \qquad (3)$$

which shows that point processes with positive correlation coefficients have monotonically increasing

IDI curves. Notice also that the limit of equation (3), when it exists, is proportional to the sum of all correlation coefficients (plus 1), that is,

$$\lim_{n \to \infty} J_n = C_J^2 \left[1 + 2 \sum_{j=1}^{\infty} \rho_j \right] . \tag{4}$$

The asymptote in the limit above depends on the sum (integral) of all the correlation coefficients. Since the interarrival times typically become statistically independent as the lag increases (the causation effects triggered by a packet transmission, such as additional queuing and increased disk activity, diminish as time increases), making their autocorrelation coefficients decrease to 0, for practical purposes the limit will be reached for a finite value of j. As noted above, packet-arrival processes normally have positive autocorrelation coefficients since both interarrivals shorter than the mean interarrival time and those longer than the mean interarrival time tend to occur in separate bursts. In packet-arrival processes, we would thus expect the IDI sequence to increase with n. Notice that, if the data are not stationary, we can still compute an estimate of J_n; however, equations (1), (3), and (4) are no longer generally valid.

B. The Index of Dispersion for Counts

We can also analyze point processes from the perspective of packet counts–the number of packets in an interval. We can define for packet counts a function similar to the index of dispersion for intervals. The *index of dispersion for counts* (IDC) is the variance of the number of arrivals in an interval of length t divided by the mean number of arrivals in t:

$$I_t = \frac{\mathrm{var}(N_t)}{E(N_t)} , \tag{5}$$

where N_t indicates the number of arrivals in an interval of length t. The IDC has been so defined in order that for a Poisson process the value of the IDC is 1, for all t.

In estimating the IDC of measured arrival processes, we will only consider the time at discrete, equally spaced instants τ_i ($i \geq 0$). Indicating with c_i the number of arrivals in $\tau_i - \tau_{i-1}$, we have

$$I_\tau = \frac{\mathrm{var}(\sum_{i=1}^{n} c_i)}{E(\sum_{i=1}^{n} c_i)} = \frac{\mathrm{var}(c_\tau)}{E(c_\tau)} \left[1 + 2 \sum_{j=1}^{n-1} (1 - \frac{j}{n}) \xi_j \right] , \tag{6}$$

where $\mathrm{var}(c_\tau)$ and $E(c_\tau)$ are the common variance and mean of the c_i's, and ξ_j is the autocorrelation coefficient of the c_i's at lag j.

Notice that, in general, I_t will not be constant for renewal processes, in which counts in disjoint intervals are correlated, save for some notable cases such as the Poisson process. However, observing that the sum of the counts in an interval of size t is less than or equal to k if and only if the sum of k interarrival times is larger than t (assuming that the process has a point at the origin), $\Pr(\sum_{i=1}^{n} c_i \le k) = \Pr(\sum_{j=1}^{k} X_j > t)$, it can be proved that the limits of the IDI and IDC are equal: $\lim_{n\to\infty} J_n = \lim_{t\to\infty} I_t$ [4]. While we can always estimate $var(N_t)/E(N_t)$, the representation on the righthand side of equation (6) is valid only if the data are stationary. Finally, we observe that both the IDC and the IDI are dimensionless quantities: they do not depend on the dimensions of the variables used in their estimation.

In the next three subsections, we will calculate the IDI for the class of renewal models with hyperexponential interarrival times, the IDC for batch Poisson processes, and present Heffes and Lucantoni's derivation of the IDC for Markov-modulated Poisson processes [9].

C. IDI for Processes with Hyperexponential Interarrival Time Distributions

The hyperexponential distribution of order k, H_k, is the weighted sum (mixture) of k exponential distributions:

$$F_{H_k}(t) = \Pr(H_k \le t) = \sum_{i=1}^{k} \alpha_i (1 - e^{-\lambda_i t})$$

with weights $\alpha_i > 0$, satisfying $\sum_{i=1}^{k} \alpha_i = 1$, and rates of the exponential distributions $\lambda_i > 0$. Because it is characterized by a coefficient of variation greater than 1, the hyperexponential distribution is often used to approximate the interarrival-time distribution of bursty processes. In the remainder of this section we will only consider H_2.

The mean of a H_2 distribution is

$$E(H_2) = \mu_1 = \frac{\alpha\lambda_2 + (1-\alpha)\lambda_1}{\lambda_1\lambda_2}$$

and the variance

$$\mathrm{var}(H_2) = \frac{2(1-\alpha)\lambda_1^2 + 2\alpha\lambda_2^2 - ((1-\alpha)\lambda_1 + \alpha\lambda_2)^2}{\lambda_1^2\lambda_2^2} ,$$

in which we have set $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$.

It is interesting to study the range of the squared coefficient of variation of intervals for the hyperexponential distribution. This is the constant value of the IDI of a renewal process whose interarrival time distribution is H_2. The coefficient of variation depends on three quantities: α, λ_1, and λ_2; thus, if we choose a value for $\mu_1 = \hat{\mu}$, which we keep constant, to derive

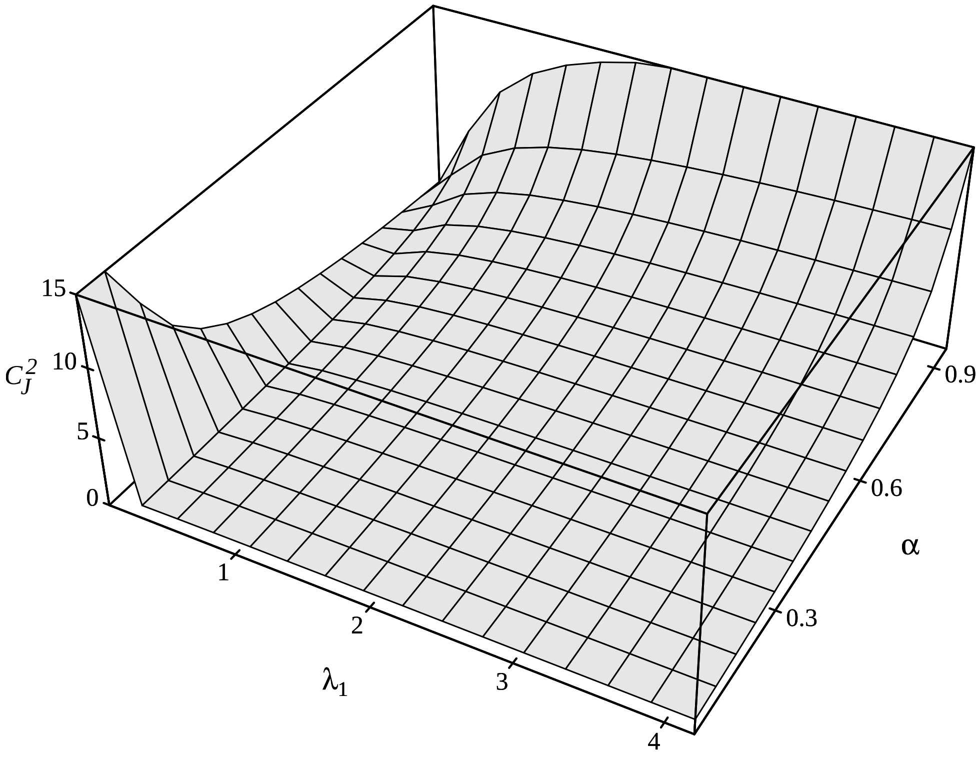
$$\lambda_2 = \frac{(1-\alpha)\lambda_1}{\hat{\mu}\lambda_1 - \alpha} ,$$

with the constraint $\lambda_1 > \alpha/\hat{\mu}$, we can obtain a formula for the coefficient of variation that depends only on α and λ_1:

$$C_J^2(\alpha,\ \lambda_1) = \frac{(1+\alpha)\hat{\mu}^2\lambda_1^2 - 4\alpha\hat{\mu}\lambda_1 + 2\alpha}{(1-\alpha)\hat{\mu}^2\lambda_1^2}\ .$$

In Figure 3 we show a three-dimensional plot of C_J^2 for $\hat{\mu} = 5$, which shows that the squared coefficient of variation increases both as α increases to 1 and (when λ_1 approaches $\alpha/\hat{\mu}$) as α decreases to 0.

FIGURE 3. SQUARED COEFFICIENT OF VARIATION FOR H_2



Indeed, we have

$$\lim_{\alpha\to 1} C_J^2(\alpha,\ \lambda_1{=}const) = \infty\ , \quad \lim_{\lambda_1\to\infty} C_J^2(\alpha{=}const,\ \lambda_1) = \frac{1+\alpha}{1-\alpha}\ , \quad \text{and} \quad \lim_{\alpha\to 0,\ \lambda_1\to\frac{\alpha}{\hat{\mu}}} C_J^2(\alpha,\ \lambda_1) = \infty\ ,$$

which proves that, for a given value of the mean arrival rate, the variability increases when α approaches the limits of its domain: 0 or 1. But this is inappropriate when the purpose of an approximation based on hyperexponential interarrival times is to model arrival processes markedly different from a Poisson process and with a large coefficient of variation. When α is nearly 0 or nearly 1, for ‘‘most of the time’’, the approximation process has interarrival time exponentially distributed with rate λ_2 or λ_1 respectively.

D. IDC for Batch Poisson Processes

The batch Poisson process is a generalization of the Poisson process in which a random number of simultaneous arrivals, p_i, replaces the original single arrival. The p_i's are i.i.d. and the total number of arrivals in an interval of duration t is $p(t) = \sum_{i=1}^{N(t)} p_i$, where $N(t)$ is the number of original Poisson arrivals. Since the p_i's are independent of $N(t)$, the mean number of arrivals in an interval of size t is clearly $E\big(N(t)\big)E(p_i) = \lambda\, t\, E(p_i)$, where λ is the arrival rate of the original Poisson process.

In order to compute the IDC for a batch Poisson process we need $\mathrm{var}\big(p(t)\big)$. Using the properties of conditional probability, we have

$$\mathrm{var}\big(p(t)\big) = E\big(p^2(t)\big) - E^2\big(p(t)\big) = E\left[E\big(p^2(t) \mid N(t)\big) \right] - E^2\left[E\big(p(t) \mid N(t)\big) \right] .$$

We expand the first of the last two terms using the definition of variance

$$\mathrm{var}\big(p(t)\big) = E\left[\mathrm{var}\big(p(t) \mid N(t)\big) \right] + E\left[E^2\big(p(t) \mid N(t)\big) \right] - E^2\left[E\big(p(t) \mid N(t)\big) \right] ,$$

and then condense the two rightmost terms immediately above using the definition of variance again:

$$\mathrm{var}\big(p(t)\big) = E\left[\mathrm{var}\big(p(t) \mid N(t)\big) \right] + \mathrm{var}\left[E\big(p(t) \mid N(t)\big) \right] .$$

Because $N(t)$ is independent of p_i, we can finally derive

$$\mathrm{var}\big(p(t)\big) = E\big(N(t)\mathrm{var}(p_i)\big) + \mathrm{var}\big(N(t)E(p_i)\big) \;=\; \lambda t\big(\mathrm{var}(p_i) + E^2(p_i)\big) .$$

From definition (5), the IDC is then

$$I_t = \frac{\mathrm{var}(p_i)}{E(p_i)} + E(p_i) .$$

Notice that the IDC is constant since a batch Poisson process is a case of a regenerative process with independent increments.

When the distribution of batch arrivals is geometric, i.e., $\Pr(N{=}\,n) = (1-p)p^{n-1}$, with $0 < p < 1$, the IDC of the batch Poisson process becomes $I_t = \dfrac{1+p}{1-p}$.

It should be pointed out that, when a renewal approximation is used for bursty non-renewal processes, a fitting based on the value of the moments may not be the best. As we will see later when we estimate the indices of dispersion for packet arrival processes, the IDI asymptote may be two orders of magnitude larger than the squared coefficient of variation, that is, the value of J_1, or between the values for large t and small t of the IDC. If the main objective of a renewal approximation is to capture the variability of a point process, the fitting should be done in such a way that the resulting constant index of dispersion of the model intersects the estimated index of dispersion of the point process at an intermediate point. For instance, if a batch-Poisson model is used with geometrically distributed batch sizes, the parameter λ could be set on the basis of the estimated mean and the parameter p set to a value that gives $(1+p)/(1-p)$ an appropriate position on the estimated IDC of the point process.

E. IDC for Markov-Modulated Poisson Processes

The Markov-modulated Poisson process (MMPP) is a model that has received much attention in recent years. It is a powerful, analytically treatable model that can represent aggregate traffic generated by the superposition of several point processes. The MMPP process is a doubly stochastic Poisson process whose arrival rate varies according to the state of an n-state irreducible continuous-time Markov chain, and that, unlike renewal models, can represent correlations between interarrival times. When the Markov chain is in state i, the arrival process is Poisson with rate λ_i. In the following we will only consider a 2-state MMPP.

The 2-state MMPP process is fully specified by four parameters: two transition rates of the Markov chain and two arrival rates (one for each state). It is described by the infinitesimal generator matrix $\mathbf{Q}$ of the embedded Markov chain and by a diagonal matrix $\mathbf{\Lambda}$ whose elements are the Poisson rates of the two states:

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{bmatrix} \quad \text{and} \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} .$$

The first moment of the time between arrivals of an MMPP can be expressed as

$$\mu_1 = \mathbf{p}\left(\mathbf{\Lambda} - \mathbf{Q}\right)^{-2} \mathbf{\Lambda e} ,$$

where $\mathbf{p}$ is the vector $\left(\dfrac{\lambda_1\sigma_2}{(\lambda_1\sigma_2 + \lambda_2\sigma_1)}, \dfrac{\lambda_2\sigma_1}{(\lambda_1\sigma_2 + \lambda_2\sigma_1)}\right)$ and $\mathbf{e}$ the vector (1, 1). The second product moment is

$$\mu_2 = 2\mathbf{p}\left(\mathbf{\Lambda} - \mathbf{Q}\right)^{-3} \mathbf{\Lambda e} .$$

For derivations of the above results see [11] or [6]. After expanding and simplifying, we obtain for the two moments

$$\mu_1 = \frac{\sigma_1 + \sigma_2}{\lambda_1\sigma_2 + \lambda_2\sigma_1} \quad \text{and} \quad \mu_2 = \frac{2(\lambda_1\sigma_1 + \lambda_2\sigma_2 + \sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2)}{(\lambda_1\sigma_2 + \lambda_2\sigma_1)(\lambda_1\sigma_2 + \lambda_2\sigma_1 + \lambda_1\lambda_2)} ,$$

from which we can easily derive the squared coefficient of variation for intervals: $C_J^2 = \mu_2/\mu_1^2 - 1$.

Heffes and Lucantoni [9] derive a formula for the IDC of a 2-state MMPP process:

$$I_t = 1 + \frac{2\sigma_1\sigma_2(\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^2(\lambda_1\sigma_2 + \lambda_2\sigma_1)} - \frac{2\sigma_1\sigma_2(\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^3(\lambda_1\sigma_2 + \lambda_2\sigma_1)t}(1 - e^{-(\sigma_1 + \sigma_2)t}) . \tag{7}$$

The asymptote of the IDC is

$$I_\infty = 1 + \frac{2\sigma_1\sigma_2(\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^2(\lambda_1\sigma_2 + \lambda_2\sigma_1)} ,$$

and it is also straightforward to verify that

$$\frac{I_\infty - I_{t_0}}{I_\infty - 1} = \frac{1 - e^{-rt_0}}{rt_0} , \tag{8}$$

where $r = \sigma_1 + \sigma_2$ can be interpreted as the ''rate'' at which the IDC approaches its asymptote. Equation (8) can be used to estimate r for a measured arrival process since the lefthand side can be easily evaluated from a point at t_0 on the IDC and the estimated IDC asymptote; r can then be obtained by solving (8) numerically.

Of the three processes we have analyzed, only the MMPP can be used to represent correlations between subsequent arrivals. A model based on hyperexponential interarrival times is less appropriate than a batch Poisson model to approximate highly variable non-Poisson measured processes since its interarrival-time distribution is close to an exponential distribution when its coefficient of variation is large. In the next sections, we examine measured arrival processes, estimate their indices of dispersion and, finally, outline a procedure to fit the four MMPP parameters to bursty arrival processes.

III. The Traffic Measurements

The data we will present in this and the following sections are based on measurements taken on a large network of workstations at the Sun Microsystems headquarters in Mountain View, California [8]. The measurements were obtained by instrumenting the kernel of a dedicated Sun-4/280 machine to read all the packets of an Ethernet local-area network and to record on the machine's disks the protocol headers along with the time obtained from a hardware clock. Although the clock's resolution was 0.5 μs, the recorded timing information is accurate only to within 100 μs due to limitations in the measurement system. The system was capable of tracing more than 1000 packets/s, the equivalent of a network load of more than 65%, but the actual traffic only occasionally exceeded 400 packets/s. We ran experiments to estimate the fraction of packets that were not recorded because of congestion of the measurement system; no packet loss was detected. For more information on the measurement system, protocol analysis, and general traffic analysis, refer to [8] and to a previous study that examines measurements taken on another network, with characteristics similar to those of the Sun network [7].

Throughout the rest of this paper we will use ''client workstation'' to refer to a single-user high-performance personal computer as opposed to ''server workstation'' or ''file server'', which is a machine that provides file space and file services. Some client workstations are diskless, while others have local disks; almost all use the Network File System protocol (NFS) [14, 18] to access files from a set of file servers. File servers have no users of their own; instead they process requests received through network interfaces and send replies to clients with the requested information, which may take the form of a data file page, a segment of a file path name during path-name translation, acknowledgments, or error messages.

We will focus on message streams generated by individual workstations. In particular, we will consider separately the streams generated by the send queues of the six workstations listed in Table 1. To ensure security of the Sun network and to protect the privacy of individual users, we have substituted the designations ‘‘station 1’’ through ‘‘station 6’’ for the machines’ actual names.

TABLE 1. WORKSTATIONS WHOSE TRAFFIC IS USED IN THE EXAMPLES

WORKSTATION	TYPE	MEMORY SIZE	OS VERSION	LOCAL DISKS
station 1	Sun-3/260 client	16 Mbytes	SunOS 4.0	1 disk
station 2	Sun-4/260 client	32 Mbytes	SunOS 4.0	1 disk
station 3	Sun-3/50 client	4 Mbytes	SunOS 4.1	diskless
station 4	Sun-3/50 client	8 Mbytes	SunOS 4.0	diskless
station 5	Sun-4/280 server	32 Mbytes	SunOS 4.0	2 disks
station 6	Sun-3/50 client	4 Mbytes	SunOS 4.0	diskless

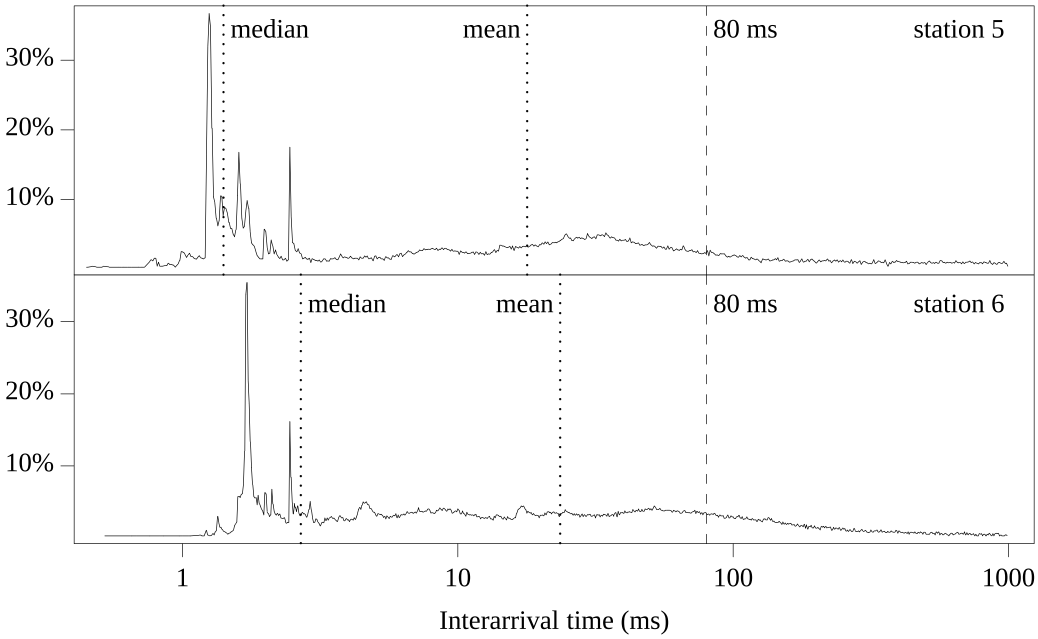
Figure 4 shows histograms of the interarrival times for the send queues of two of the workstations, station 5 and station 6. In order to capture in these histograms interarrival times spanning more than three orders of magnitude, we have made the size of the bins increase exponentially, i.e., the size of bins is constant on a logarithmic scale. On the ordinate, we plot the square root of the bin counts; on the abscissa, the interarrival times on a logarithmic scale. The increasing-bin-size method leads to higher values in the tail regions of the histograms since bin sizes become considerably larger than they would be had we used a constant bin size.

Each histogram can be roughly divided into three areas: an initial series of peaks, a flat middle portion, and a long tail that begins at around 100 ms. (For a more precise identification of these areas and a discussion of the properties of the associated packet arrivals refer to [8, Chapter 4].) The first area, consisting of interarrival times up to 5 ms, shows a strong dependence on protocol time constants. As a result, the two machines produce peaks at about the same points; the peaks of station 5 are only slightly shifted to the left of station 6’s peaks. The relative shift can be explained by the difference in speed between the two CPUs: station 6 is a 15 MHz Motorola 68020 and station 5 is a 16 MHz Sparc, which, in spite of the comparable clock rate, has a much faster architecture.

The highest peak in the station 6 graph at 1.7 ms corresponds to the interarrival times between 1500-byte fragments of 8-Kbytes NFS write transfers. Since 1.2 ms, the time at which station 5’s highest peak is produced, is the minimum permissible time between two 1500-byte Ethernet frames, station 5 is sending these packets back-to-back.

The middle area of the histograms, the area approximately between 5 and 100 ms, is produced by small NFS packets used in request-response dialogues between clients and file servers. After issuing a request, a client must wait until the server replies before sending its next request. Often requests involve operations requiring a server to access a disk device, which causes high variability in the

FIGURE 4. HISTOGRAMS OF INTERARRIVAL TIMES



interarrival times. This variability is indicated by the relatively flat shape of the curve over this interval.

Longer interarrival times, those from around 100 ms to more than 1 s (in Figure 4 we show only interarrivals shorter than 1 s), are produced when pauses in user activity lead to corresponding pauses in packet transmission. Although these long pauses are relevant to the study of user behavior, we will not analyze them in this paper because our focus is on the much faster packet transactions generated by networking protocols.

We will treat arrival processes as point processes. A major aspect of the study of point processes is the analysis of their correlation properties. Since the Ethernet has variable-length packets, there may be, as observed above, correlations between interarrival times and packet lengths. Furthermore, since the baseband transmission medium is shared, i.e., concurrent transmission is serialized, there may be correlations among separate, otherwise independent streams. Finally, shared file servers constitute a possible additional source of correlations.

In order to isolate the effect of packet lengths, in [8] we compared, for the packet streams of several workstations, the autocorrelation coefficients of interarrival times with those of interpacket times. While interarrival times, as defined in Section II, include packet transmission times, interpacket times, defined to be the times between the beginning of a packet's transmission and the end of the transmission of the previous packet, do not. We observed no noticeable difference between the two; since the effects of the variable packet lengths are only present in the interarrival times, we concluded that there is no significant correlation caused by variable packet lengths. This result was confirmed by our study of the cross-correlations between the series of interarrival times and the series packet lengths.

We also studied the cross-correlation between streams generated by two clients on distinct file servers; their only interaction point was the network. In addition, we looked at the cross-correlation between streams generated by workstations that shared a set of file servers. No correlation was observed. Indeed, the only noticeable correlation effects found were between the interarrival times of the stream transmitted by a workstation and the interarrival times of the stream received by the same workstation. Apparently, if any correlations do exist, they are overwhelmed by the large variability in interarrival times. The association detected between the send and receive streams of a workstation is attributable to the use, by many applications, of remote procedure call protocols, which suspend themselves until an answer to the sole outstanding message is received, resulting in interlocks between send and receive operations. In addition, since in systems that use request-response protocols operations are typically initiated by clients (and not by servers), pauses in a user's activity that result in long interarrival times in the send stream are almost invariably associated to pauses in the receive streams.

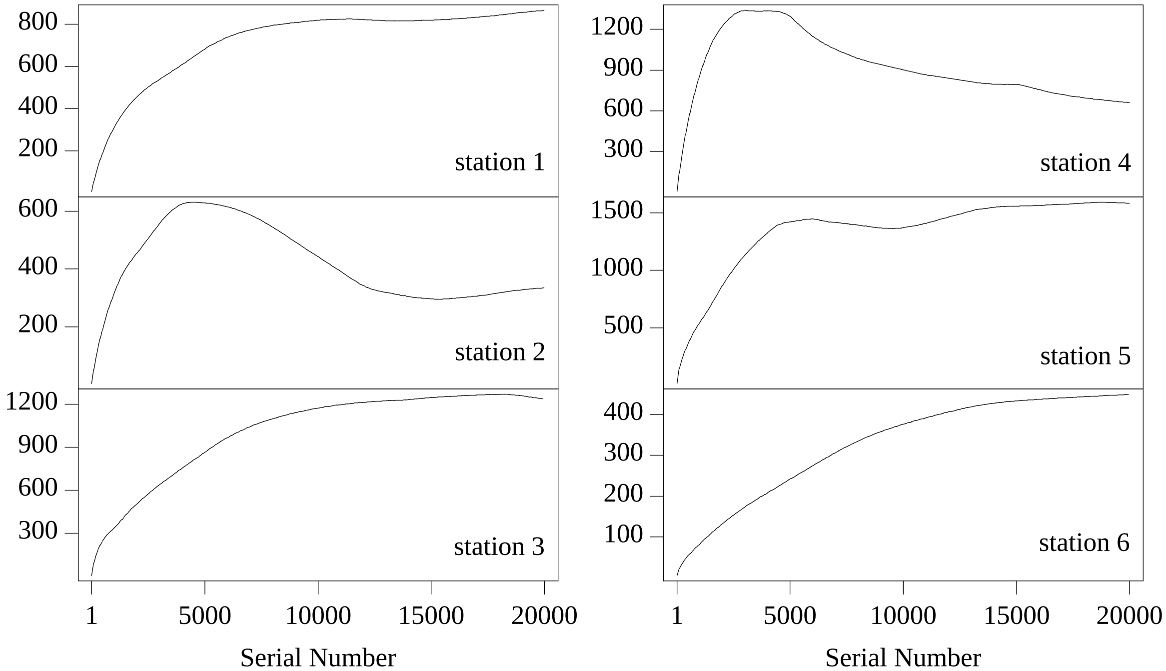
A. Estimated Indices of Dispersion for Intervals

In this section, using index-of-dispersion analyses, we look at the second-order joint probability structure of the packet arrival processes generated by six individual workstations. In Figure 5 we plot the estimated index of dispersion curves for the send queues of these workstations. Each series contains 100 000 interarrival times. The IDI curves are estimated for up to 20% of the length of the original series of data; after this point, with few remaining degrees of freedom, any further estimate would have been inaccurate. For details on how to estimate the indices of dispersion and how to evaluate the precision of the estimates, refer to [3].

The IDI at lag n is the variance of the sum of n successive interarrival times, and the IDI curve indicates the change in the variance as n increases. What appears to be very large variability (the starting values of the IDI for the six curves are, proceeding from station 1 to station 6, 7.2, 3.5, 6.4, 6.8, 13.0, 5.4, while the maxima range from 400 to more than 1500) is caused primarily by nonstationary components in the data. Most remarkable is the effect on the curve of station 4, which increases sharply until approximately lag 3000, then stabilizes for the next 2000 points, and begins decreasing from lag 5000 onwards. From equation (3) we would think that the autocorrelation coefficients of this series of data become negative at around lag 5000, but this is not the case: the coefficients are all positive and only slowly decrease, another sign of nonstationary data.

A look at the smoothed interarrival-time curve of station 4's arrival process, illustrates what is at work. (The smoothing was done by lowpass filtering the data in the frequency domain.) In Figure 6, the two peaks of long interarrival times, one at the beginning and the other at the end of the graph, cause the total variance and the IDI to decrease as more points are averaged together. Nonstationary

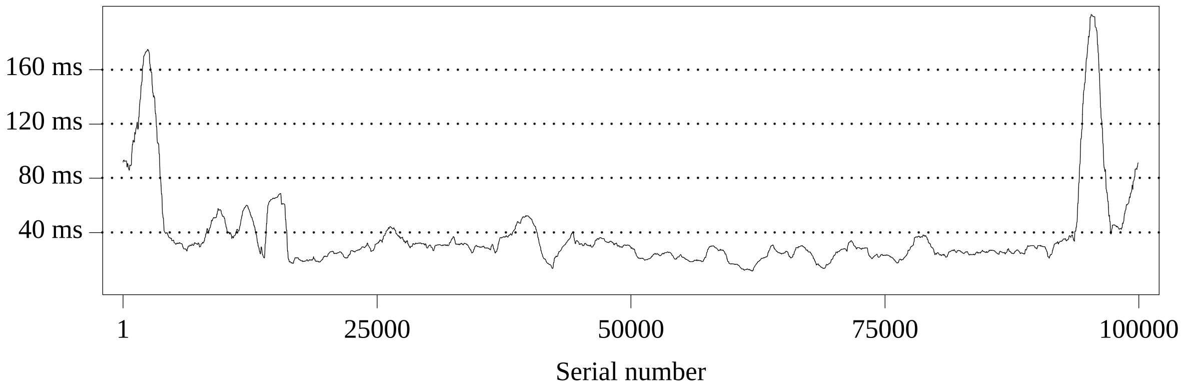
FIGURE 5. ESTIMATED INDEX OF DISPERSION FOR INTERVALS – WORKSTATION SEND QUEUES



data with a minority of large values clustered together generate a gradually decreasing IDI curve such as station 4's.

It could be argued that the apparent nonstationarity of the data in Figure 6, which spans a period of about 1 hour, is a function of the time scale. This particular user might follow the same work patterns and produce roughly the same workload in successive hours; thus, viewed over a period of several hours, the data might appear to be (more) stationary.

FIGURE 6. SMOOTHED INTERARRIVAL TIME CURVE – WORKSTATION STATION 4



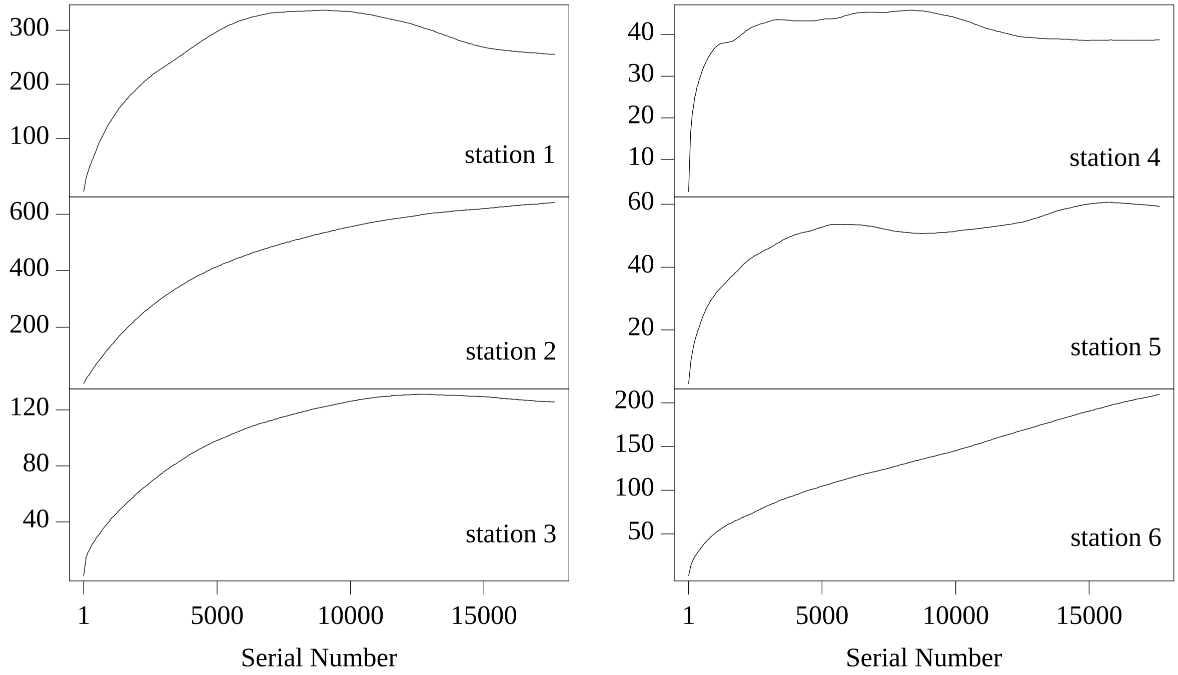
For our purpose of analyzing variability however, the relevant time scale is not one of several hours, or even minutes, but a shorter one, for a workstation will generate a constant packet rate in intervals ranging from milliseconds to seconds. Analyzing the variability of point processes at a micro level, i.e., in terms of the queues to which the processes are fed, implies a time scale that is this short. In contrast, a study of packet arrival processes from the perspective of user behavior would involve a much longer time scale, one defined by the busy/idle intervals of user behavior, which would range from several minutes to hours.

Nonstationary behavior in these types of arrival processes is the norm. It is virtually impossible to isolate a stationary segment of a process long enough for the estimation of many important statistical parameters. It is possible, however, to identify segments of time during which the process has roughly the same characteristics. One can then juxtapose the various segments with the same properties, assemble several series of arrivals, each of which is what we have called a *phase,* and derive the statistical description for each phase. As a very simple illustration of this procedure, let us re-examine the processes shown in Figure 5 by considering subsets of interarrival times shorter than a specified length.

The cut-off time chosen should be short enough to eliminate the two peaks shown in Figure 6 but sufficiently long to capture not only the fast protocol transactions such as file path-name translations, but also the slower disk transactions. In Figure 4, the height of the histograms of interarrival times decreases visibly at or around 80 ms. This decrease indicates that at around 80 ms there is a transition from frequent system-generated events, driven by network protocols and disk-driver software, to infrequent user-generated ones, such as keystrokes. All of the workstations in our traces show similar histograms, an indication that the results described in the reminder of the paper are relatively insensitive to variations of this parameter. Thus, we set the threshold at 80 ms, a value that we have indicated in Figure 4 with a dashed line.

The validity of this approach in reducing the nonstationarity measured by the index of dispersion for intervals can be sustained with the following argument. The probability density functions of the interarrival times for packet arrival processes typically have a very large mass at the beginning and very long and low tails, which results in the median being smaller than the mean value. (In Figure 4 we show the mean and median of the histograms.) The index of dispersion for intervals, as we have seen in equation (1), depends on the autocovariances of the interarrival times. The autocovariance at lag k is defined as $E\left[\big(X_i - E(X)\big)\big(X_{i+k} - E(X)\big) \right]$ where $E(X)$, as usual, indicates the mean interarrival time. We see that, for each lag k, the few long interarrival times, because of their large differences from the mean, account for much of the total covariance. Conversely, the large number of small interarrivals closer to the mean have relatively little effect on the covariance. Hence, removing the largest interarrival times from a time series that is skewed towards small values reduces the index of dispersion for intervals and, if the large values are clustered as in Figure 6, may prevent the IDI's of packet arrival processes, which normally have positive correlation coefficients, from not being monotonic.

FIGURE 7. ESTIMATED INDEX OF DISPERSION FOR INTERVALS – INTERARRIVALS SMALLER THAN 80 MS

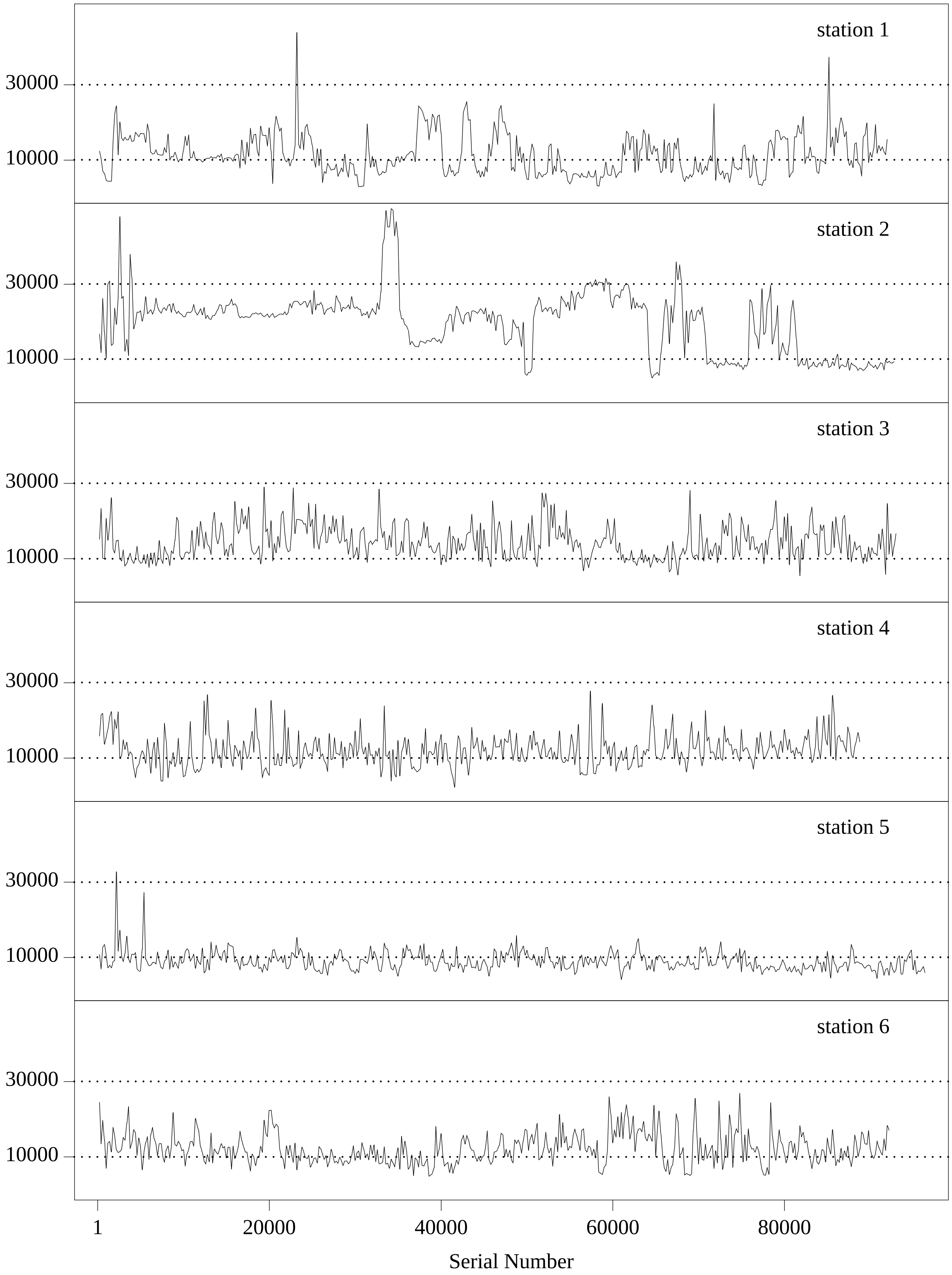


In Figure 7 we re-evaluate the indices of dispersion for the subset of interarrivals that are smaller than 80 ms. We find that, in three cases (station 3, station 4, and station 5), the asymptotic value is reduced by a more than, or nearly, one order of magnitude. The series of data are 4% to 11% shorter than their counterparts in Figure 5 because the interarrivals greater than 80 ms have been dropped.

Analysis of the smoothed interarrival times of these truncated series, which are shown in Figure 8, helps us understand the dynamics of the IDI behavior. Each of these curves was obtained by averaging together 200 successive interarrival times. We notice that the two peaks in station 1's interarrival times are responsible for the decreasing slope of that machine's IDI. The almost linearly increasing slope of station 6's IDI can be understood by looking at two segments of the machine's interarrivals: one between 20 000 and 40 000 and a second between 60 000 and 80 000. The first interval contains interarrival times smaller than the mean; the second, interarrival times larger than the mean. Since the size of each of these segments happens to be equal to the length of the interval over which we estimate the IDI, upon a moment's reflection one will realize that the variance of the sum of consecutive interarrivals will increase linearly.

The range of station 2's IDI in Figure 7 is the same as that in Figure 5, but the second curve increases more gradually and is monotonic. The smoothed interarrival times of station 2 in Figure 8 explain why the range of the truncated series is the same as the original one: despite the elimination of longer interarrival times, the series remains highly variable. The user of this particular machine was not active for much of the time represented in Figure 7; during the inactive stretches, programs left running in the machine invoked remote procedure calls but did not transfer data. The remote

FIGURE 8. SMOOTHED CURVES OF TRUNCATED INTERARRIVAL TIME SERIES

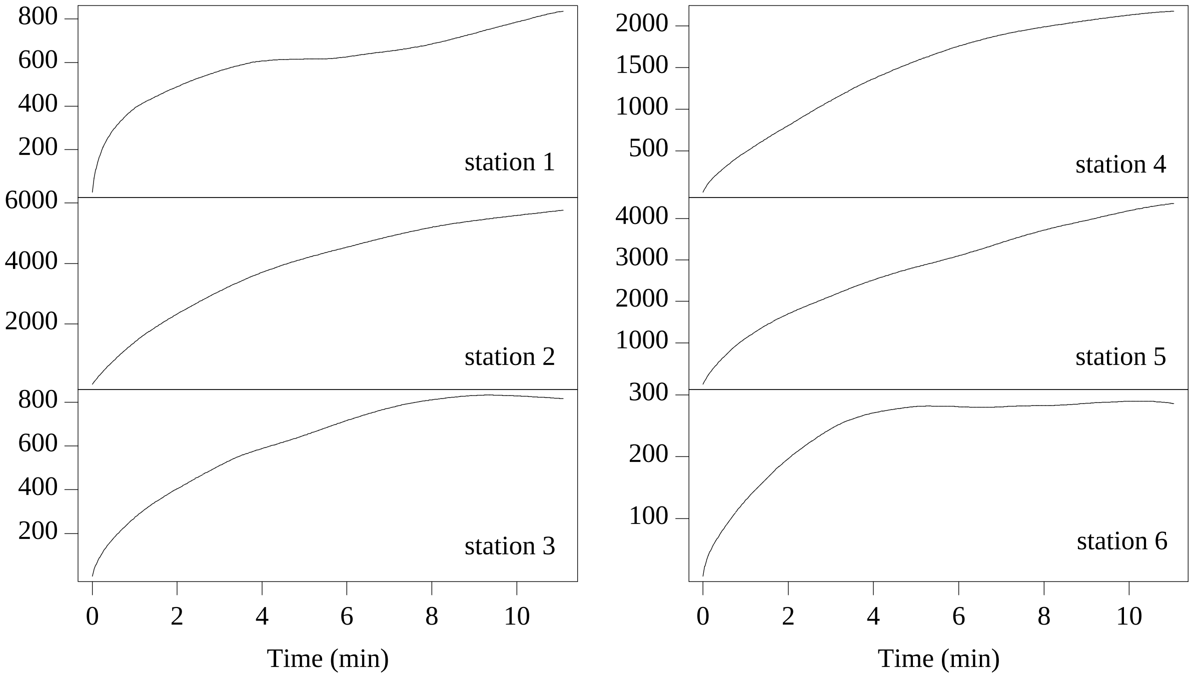


procedure calls produced longer interarrival times, the lack of data transfers did not produce the shortest interarrival times. The nature of station 2's arrival processes reminds us that the study of indices of dispersion is laden with complexities.

B. Estimated Indices of Dispersion for Counts

Figure 9 shows the estimated index of dispersion for counts. The IDC was evaluated only up to one-sixth of the total time length of the traces, or about 11 min. In each of the graphs, packet counts were estimated in slots of size 50 ms.

FIGURE 9. ESTIMATED INDEX OF DISPERSION FOR COUNTS – WORKSTATION SEND QUEUES



A description of a point process in terms of counts is statistically equivalent to a description in terms of intervals [4]. However, they are equivalent only through their complete joint distributions. If we restrict ourselves to first- and second-order properties, the two characterizations are separately informative. For instance, histograms of packet counts, of which we show two examples in Figures 10 and 11, are rather different in shape from the histograms of intervals shown in Figure 4. Analogously, the estimated IDC curves in Figure 9 are substantially different from their IDI counterparts in Figure 5. However, their technical interpretation, since equations (6) and (3) are quite similar, proceeds along the same lines.

Here also we see the effect of nonstationary components. The IDC of station 2 is particularly affected by the grouping of arrivals in some regions of the domain of the packet count process. Station 2 has an almost linearly increasing IDC, which can be attributed to the same sort of nonstationary data structure underlying the linearly growing IDI of station 6. In general, since the domain of values

FIGURE 10. HISTOGRAM OF PACKET COUNTS
WORKSTATION STATION 3

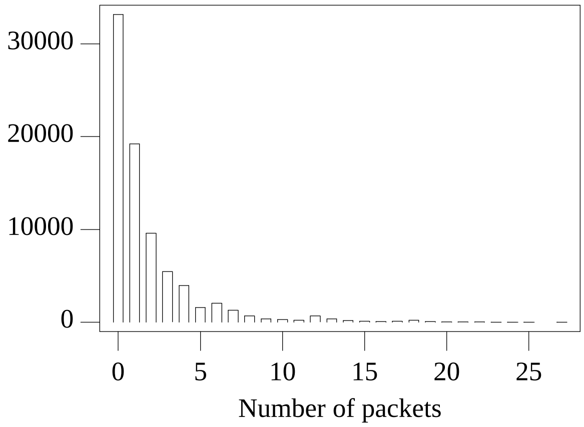
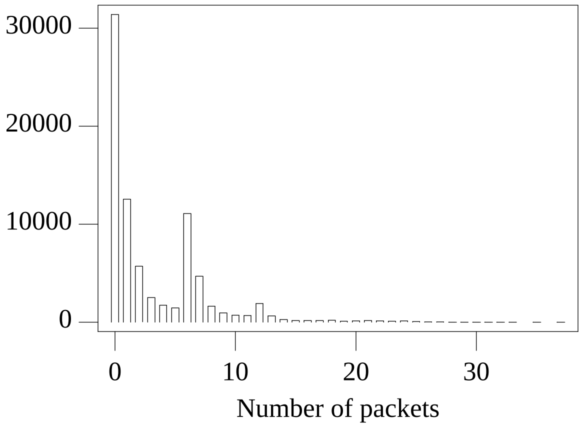


FIGURE 11. HISTOGRAM OF PACKET COUNTS
WORKSTATION STATION 5



of counts is rather limited (especially so when packet counts are estimated over relatively short intervals, as in our case), we can say that IDC curves are more sensitive than IDI curves to the presence of clusters of arrivals. The limited range of the domain of packet counts results in probability density functions of counts with very short tails.

It is important to notice that the asymptotes in Figure 9 are different from those in Figure 5. This is yet another confirmation of the presence of nonstationary data. We have recalculated (but not shown here) the IDC's for interarrivals less than 80 ms. In this case, there is a much better agreement between the data and the limiting result that states that I_∞ and J_∞ are equal (Section II.B). Indeed, if the shapes of the IDI and IDC curves for an arrival process appear monotonic, and the limits of the two indices are the same, one can assume with a considerable degree of confidence that the data are stationary.

In this section we have shown and interpreted the IDI and IDC of several measured packet arrival processes. The probabilistic definitions of these indices makes them suitable for describing the variability of point processes. We next show how this variability representation can be incorporated into analytical modeling.

IV. FITTING A 2-STATE MMPP TO AN ARRIVAL PROCESS

In this section we present a procedure that can be used to fit an MMPP model to packet arrival processes of the type we have been describing as long as the nonstationary data components are somewhat controlled. To provide a concrete example, we will work with the data of workstation station 3, one of the most stationary data sets. Station 3's indices of dispersion appear in Figures 5 and 9. For the mathematical terminology, refer to the MMPP definitions in Section II.E.

We can write the following system of three equations, representing, from top to bottom, the mean interarrival time of an MMPP, the asymptote I_∞, and what we have defined in equation (8) as r, the rate at which the IDC approaches the asymptote:

$$\begin{cases} \dfrac{\sigma_1 + \sigma_2}{\lambda_1 \sigma_2 + \lambda_2 \sigma_1} = a \\ 1 + \dfrac{2\sigma_1 \sigma_2 (\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^2 \, (\lambda_1 \sigma_2 + \lambda_2 \sigma_1)} = b + 1 \\ \sigma_1 + \sigma_2 = c \quad . \end{cases} \tag{9}$$

The quantities a and b represent, respectively, the estimated mean of the interarrival times of a measured point process and the estimated value of the IDC asymptote minus 1, both of which can be obtained with modest effort. An initial value for the parameter c, an estimate of r, can be computed numerically, as indicated in Section II.E, from b and I_{t_0}, the IDC at time t_0. The choice of t_0 is not crucial, as we can repeat part of the procedure we are about to describe until we reach a satisfactory approximation based on some measure of the goodness of fit.

We begin by solving the three equations (9) to obtain the values of λ_1, σ_1, and σ_2 as functions of a, b, c, and the unknown λ_2:

$$\begin{cases} \lambda_1 = \dfrac{2 + abc - 2a\lambda_2}{2a - 2a^2\lambda_2} \\ \sigma_1 = \dfrac{abc^2}{2 + abc - 4a\lambda_2 + 2a^2\lambda_2^2} \\ \sigma_2 = \dfrac{2c(a\lambda_2 - 1)^2}{2 + abc - 4a\lambda_2 + 2a^2\lambda_2^2} \quad . \end{cases} \tag{10}$$

Next, we equate the formula of the squared coefficient of variation for an MMPP, $C_j^2 = \mu_2 / \mu_1^2 - 1$, to the square of the estimated value of the coefficient, which we call d, in order to determine the value of the unknown λ_2. Since μ_1 and μ_2 depend only on the four MMPP parameters, we can substitute the values above to obtain a formula for d in λ_2:

$$d = \frac{2a\lambda_2^2 + (2ac + abc - 2)\lambda_2 - 2c(b+1)}{2a\lambda_2^2 + (2ac - abc - 2)\lambda_2 - 2c} \; . \tag{11}$$

The expression of λ_2 in terms of the quantities a, b, c, and d is simple but rather tedious and we will omit the details here. Note, however, that for the righthand side of equation (11) the limit as λ_2 approaches infinity is 1 and the limit as λ_2 goes to 0 is $b + 1$.

To fit a 2-state MMPP to a measured arrival process we set the four parameters as follows:

1. From the data, estimate a, the mean interarrival time; b, the limiting value of the IDC minus 1; and d, the squared coefficient of variation of the interarrival times.

2. Using b, t_0, and I_{t_0}, the value of the IDC at time t_0, estimate numerically an initial value for the rate c by solving equation (8).

3. From the solutions to equation (11), obtain a value for λ_2, and use it to derive values for λ_1, σ_1, and σ_2 from equations (10). (Note that, in general, there are two solutions for λ_2 from equation (11).)

4. Compute, based on the current values of the parameters, the goodness of the approximation by comparing the estimated IDC with the theoretical one calculated by equation (7). A typical test for the goodness of the fit is one that evaluates the sum of the squared distances between the estimated and the theoretical IDC curves (for some applications it may be worth evaluating the goodness of fit only over a portion of the domain of the IDC). Finally, adjust the value of c as appropriate to improve the fit and repeat steps 3 and 4 of this procedure until the approximation is satisfactory. (A smaller c will make the IDC reach the asymptote more slowly.)

We now apply the procedure outlined above to the arrival process generated by station 3's send queue. We use the index of dispersion for counts computed only for interarrival times smaller than 80 ms. The resulting IDC is related to station 3's IDI shown in Figure 7. The estimated mean interarrival time is 0.01376 s, the value of b is 112, and the squared coefficient of variation 1.794. Using the value 20 s for t_0 and setting the estimated I_{20} at 54, we obtain for c a value of 0.073 $[sec]^{-1}$. The resulting parameters (all of them rates with dimensions $[sec]^{-1}$) are

$$\lambda_1 = 9.61\ , \quad \lambda_2 = 77.37\ , \quad \sigma_1 = 0.0680\ , \quad \sigma_2 = 0.0051\ . \tag{12}$$

There is another symmetric set of solutions, which corresponds to the above with the subscripts 1 and 2 exchanged.

In Figure 12 we plot the righthand side of equation (11), the squared coefficient of variation for intervals of an MMPP with parameters given in (12) above. (In this and in the following two figures, black dots indicate the position of the two sets of solutions on the curves.) Since d must be larger than 1, observe that we have not drawn the segment of the squared coefficient of variation curve for points where the function in Figure 12 is smaller than 1. Figure 13 shows a plot of the first of equations (10); there is a vertical asymptote at $\lambda_2 = 1/a$. Here, we have eliminated from the domain the region in which the equation generates a negative value for the Poisson rate. Finally, in Figure 14 (only on points belonging to the domain of the previous two functions) we show the curves for the transition rates of the Markov chain.

In Figure 15 we plot the estimated IDC for the data (depicted as a continuous curve) as well as the model IDC (depicted as a dotted curve). The fitting is very good in the region of the domain from 0 to 1.5 min, the portion of the IDC that is most likely to affect interactive queues.

FIGURE 12. SQUARED COEFFICIENT OF VARIATION OF X_i's

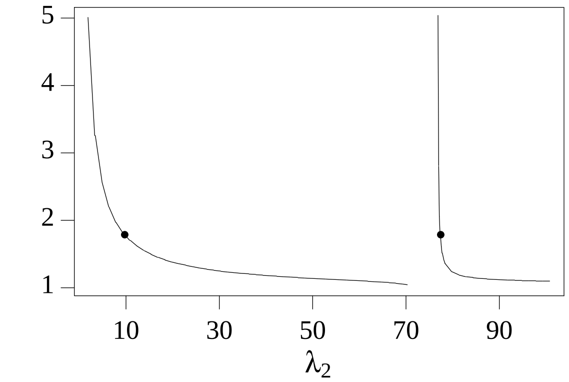


FIGURE 13. POISSON RATE λ_1

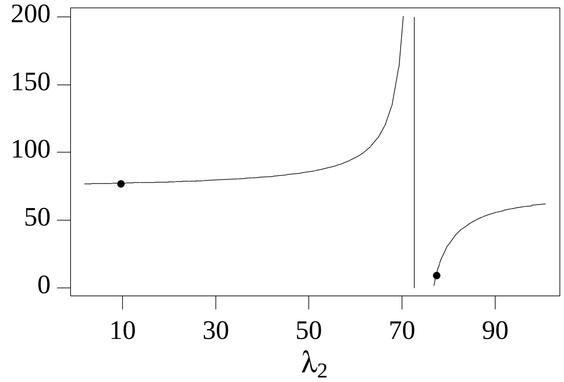


FIGURE 14. MARKOV CHAIN TRANSITION RATES

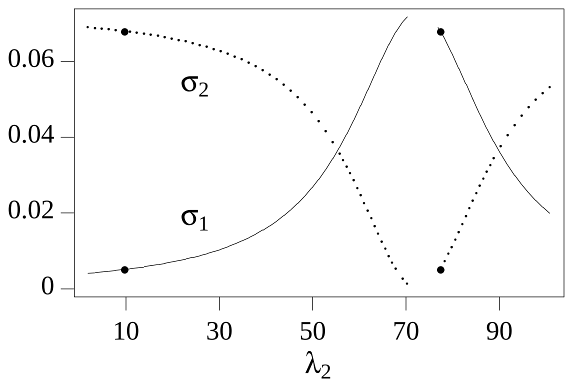
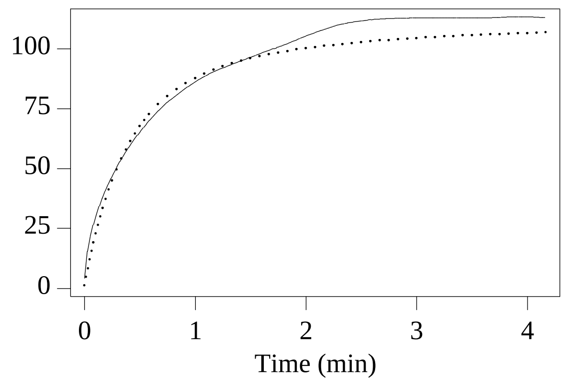


FIGURE 15. IDCS OF DATA AND FITTED MODEL (DOTTED)



V. RELATED WORK

There is a vast quantity of literature that covers the areas of point processes, statistical analysis of point processes, measurements, and fitting. We mention here only some of works most relevant to the analyses presented in this paper. The interested reader will find additional references listed in these studies.

Among related measurements studies, we would like to mention our previous study on Ethernet traffic [7] and a more recent work that provides some statistical characterization of the aggregate network traffic on an Ethernet [21].

The properties of the hyperexponential distribution and the batch Poisson process that we have used are elementary. Most probability textbooks will suffice for background or a fuller explanation of the material presented; see for instance [15]. The main reference for the MMPP process is K. Meier's thesis [11], which includes formal tests of fit, but references [6, 12, 17] are also informative.

The analysis of arrival processes using indices of dispersions is proposed in [5, 9, 20]. These studies, however, deal for the most part with processes measured in the context of telephone communication, i.e., packetized voice data.

Heffes and Lucantoni [9] propose a fitting procedure for the superposition of packetized voice processes to an MMPP process. Their characterization uses the estimated third moment and is more complex than ours. Our approach, however, is based on measurements of both intervals and counts, whereas Heffes and Lucantoni only use packet counts.

VI. SUMMARY AND CONCLUSIONS

In this paper we have used the index of dispersion for intervals and the index of dispersion for counts to characterize arrival processes consisting of packets sent by workstations in an Ethernet local-area network. By evaluating the indices for three analytical models we have illustrated some of their properties. We have suggested that renewal approximations of non-renewal point processes may benefit from index of dispersion analyses. Because indices of dispersion reveal a good deal about the correlation structure of point processes occurring in communication networks, index-of-dispersion analysis should be adopted as one of the basic tools for examining these point processes.

We have introduced a procedure for utilizing indices of dispersion to fit an MMPP model to measured data. We propose testing the quality of the approximation by comparing the IDI or IDC of the data with the model's corresponding indices. Our purpose has not been to propose models for packet arrival processes, or to demonstrate that any one model best approximates workstation traffic, but to demonstrate how to incorporate short- and long-term variability characterizations based on indices of dispersion into arrival models.

Computer communication research is proceeding apace toward future Gigabit networks. These networks will require knowledge of the statistical behavior of data streams if they are to switch data efficiently. Analyzing the variability of arrival processes with indices of dispersion can provide some of that knowledge.

ACKNOWLEDGMENT

This study was made possible by M. Itzkowitz and M. Liu, who arranged for the use of the Sun network and machinery, and provided technical advice. C. Rigatuso and M. Shoemaker of Sun Microsystems also provided valuable assistance. The careful reviews of D. Brillinger and D. Ferrari improved the paper substantially. Finally, the author would like to thank K. Meier-Hellstern for helpful discussions on the MMPP process, and P. Danzig for his comments on the paper.

REFERENCES

1. Brillinger, D. D., Comparative Aspects of the Study of Ordinary Time Series and of Point Processes, in Developments in Statistics, P.R. Krishnaiah, Ed., *Academic Press, vol. 1,* New York, 1978, 33-133.

2. Cox, D. R. and H. D. Miller, *The Theory of Stochastic Processes*, Methuen, London, 1965.

3. Cox, D. R. and P. A. W. Lewis, *The Statistical Analysis of Series of Events*, Chapman and Hall, London, 1966.

4. Cox, D. R. and V. Isham, *Point Processes*, Chapman and Hall, New York, 1980.

5. Fendik, K. W. and W. Whitt, Measurements and Approximations to Describe the Offered Traffic and Predict the Average Workload in a Single-Server Queue, *Proceedings of the IEEE 77*, 1 (January 1989), 171-194.

6. Fisher, W. and K. S. Meier-Hellstern, *The MMPP Cookbook*, INRS-Telecommunications, Verdun, Quebec, and AT&T Bell Laboratories, Holmdel, NJ, in preparation, 1990.

7. Gusella, R., A Measurement Study of Diskless Workstation Traffic on an Ethernet, *IEEE Transactions on Communications*, (to appear) September 1990.

8. Gusella, R., The Characterization of Variability of Packet Arrivals in Networks of High-Performance Workstations, Ph.D. dissertation, Computer Science Division (EECS), University of California, Berkeley, 1990.

9. Heffes, H. and D. M. Lucantoni, A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance, *IEEE Journal on Selected Areas in Communications SAC-4*, 6 (September 1986), 856-868.

10. Kleinrock, L., Computer-Communication Networks: Measurements, Flow Control, and ARPANET Traps, in *Queueing Systems*, vol. 2: Computer Applications , Wiley and Sons, 1976.

11. Meier, K. S., A Statistical Procedure for Fitting Markov-Modulated Poisson Processes, Ph.D. dissertation, University of Delaware, 1984.

12. Neuts, M. F., A Versatile Markovian Point Process, *Journal of Applied Probability 14* (1979), 764-779.

13. Nomura, M., T. Fujii and N. Ohta, Basic Characteristics of Variable Rate Video Coding in ATM Environment, *IEEE Journal on Selected Areas in Communications 7*, 5 (June 1989), 752-760.

14. Nowicki, B., NFS, Network File System Protocol, Internet Network Working Group, RFC 1094, Sun Microsystems, Inc., March 1989.

15. Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

16. Pawlita, P. F., Two Decades of Data Traffic Measurements: a Survey of Published Results, Experiences and Applicability, *Proceedings of the 12th International Telecommunication Conference*, Torino, June 1988.

17. Ramaswami, V., The N/G/1 Queue and its Detailed Analysis, *Advanced Applied Probability 12* (1980), 222-261.

18. Sandberg, R., D. Goldberg, S. Kleinman, D. Walsh and B. Lyon, Design and Implementation of the Sun Network Filesystem, *Usenix Conference Proceedings*, June 1985, 119-130.

19. Shoch, J. F. and J. F. Hupp, Measured Performance of an Ethernet Local Network, *Communications of the ACM 23*, 12 (December 1980), 711-721.

20. Sriram, K. and W. Whitt, Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data, *IEEE Journal on Selected Areas in Communications SAC-4*, 7 (September 1986), 833-846.

21. Wilson, D. V. and W. E. Leland, High Time-Resolution Measurement and Analysis of LAN Traffic: Implications for BISDN Networks, *Proceedings of the 1990 Bellcore Symposium on Performance Modeling*, Livingston, NJ, May 1990.