Positive correlations and buffer occupancy: Lower bounds via supermodular ordering

Sarut Vanichpun, Armand M. Makowski
Department of Electrical and Computer Engineering
Institute for Systems Research
University of Maryland, College Park
Email: sarut@eng.umd.edu, armand@isr.umd.edu

***Abstract*—We use recent advances from the theory of multivariate stochastic orderings to formalize the "folk theorem" to the effect that positive correlations lead to increased buffer occupancy and larger buffer levels at a discrete-time infinite capacity multiplexer queue. Input sequences will be compared in the supermodular (sm) ordering and buffer contents in the increasing convex (icx) ordering, respectively.**

Three popular classes of (discrete-time) traffic models are discussed, namely the Fractional Gaussian Noise traffic model, the on-off source model and the $M|G|\infty$ traffic model. The independent version of an input process in each of these classes of traffic models is a member of the same class. In varying degree of generality, we show that this independent version is smaller than the input sequence itself, and that the corresponding buffer content processes are similarly ordered.

I. Introduction

A. Buffer provisioning

A basic design problem in the engineering of store-and-forward networks is buffer provisioning, namely the determination of buffer sizes at various network nodes. This question is often addressed through the analysis of an appropriate queueing system. The simplest of models operates in discrete time and considers a flow of packets arriving to a finite buffer with a capacity of at most B packets; packets are transmitted out of the buffer in order of arrival over a communication link of constant rate. More precisely, with time organized in contiguous timeslots of identical duration, let Q_t^B denote the number of packets still present in the system at the beginning of timeslot $[t, t+1)$ and let A_t denote the number of new packets arriving into the buffer during that timeslot. If the buffer output link can transmit c packets/slot, then the buffer content evolves according to the recursion

$$Q_{t+1}^B = \min\left(B, [Q_t^B + A_t - c]^+\right) \quad (1)$$

for all $t = 0, 1, \ldots$, for some given intial condition Q_0^B. If the input sequence $\{A_t,\ t = 0, 1, \ldots\}$ is stationary and ergodic, then the system eventually reaches a statistical equilibrium or steady-state regime in that $Q_t^B \Longrightarrow_t Q^B$ for some random variable (rv) Q^B (with $\Longrightarrow_t$ denoting convergence in law as t goes to infinity). [1]

Determining the distribution of Q^B is a natural step towards the evaluation of key design quantities such as the blocking probability and the packet loss rate. This task is a very difficult one; closed-form solutions are available in only a few instances of input sequences $\{A_t,\ t = 0, 1, \ldots\}$, and numerical techniques need to be developed to handle most cases of practical interest.

However, in many situations (e.g., ATM networks), the blocking probability and cell loss rate assume acceptable levels only when B is large. With this in mind, it is reasonable to look instead at the infinite buffer system ($B = \infty$) associated with (1). The evolution of the buffer content sequence $\{Q_t,\ t = 0, 1, \ldots\}$ is now governed by the Lindley recursion

$$Q_{t+1} = [Q_t + A_t - c]^+, \quad t = 0, 1, \ldots \quad (2)$$

for some given initial condition Q_0. It is well known [19] that if the input sequence $\{A_t,\ t = 0, 1, \ldots\}$ is stationary and ergodic with $\mathbf{E}[A_0] < c$, then the system will reach statistical equilibrium, i.e., $Q_t \Longrightarrow_t Q$ for some $\mathbb{R}_+$-valued rv Q.

The relevance of this approach is reinforced by the observation that $Q_t^B \le Q_t$ for all $t = 0, 1, \ldots$ and all $B > 0$ as can be shown recursively through a direct sample path comparison (provided $Q_0^B \le Q_0$). Thus, the upper bounds $\mathbf{P}\left[Q_t^B = B\right] \le \mathbf{P}[Q_t \ge B]$ are valid for all $t = 0, 1, \ldots$, a fact which translates to steady state as

$$\mathbf{P}\left[Q^B = B\right] \le \mathbf{P}[Q \ge B], \quad B > 0 \quad (3)$$

under the appropriate conditions. As argued earlier, we need secure reasonably good approximations to the blocking probability $\mathbf{P}\left[Q^B = B\right]$ only for large B. Hence, as engineering designs tend to be conservative, (3) suggests that this objective can be achieved by evaluating the upper bound $\mathbf{P}[Q \ge B]$ for large B.

B. Dependencies in traffic models

In the past decade this evaluation task has been the subject of intense investigations in the wake of several traffic measurement studies which have concluded to the "failure of Poisson

This material is based upon work supported by the Space and Naval Warfare Systems Center – San Diego under Contract No. N66001-00-C-8063. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Space and Naval Warfare Systems Center – San Diego.

[1] The existence of Q^B is always guaranteed in the stationary and ergodic framework, but additional assumptions are required to have uniqueness and independence with respect to the initial condition.

modeling" [31]. Indeed, starting with the landmark data set collected at BellCore [17], a growing number of measurement studies have by now concluded that network traffic exhibits time dependencies at a much larger number of time scales than had been traditionally observed. This *long-range dependence* has been detected in a wide range of applications and over multiple networking infrastructures, e.g., Ethernet LANs [12], [17], [39], VBR traffic [7], [14], Web traffic [10] and WAN traffic [31].

Roughly speaking, long–range dependence amounts to correlations in the packet stream spanning multiple time scales, which are individually rather small but which decay so slowly as to be non–summable. This is expected to affect performance in a manner drastically different from that predicted by (traditional) summable correlation structures which typically arise in Markovian traffic models and Poisson–like sources. This state of affairs has generated a strong interest in a number of alternative traffic models which capture observed (long–range) dependencies; good surveys are available in [13], [22], [26]. Proposed models include Fractional Brownian Motion [25] and its discrete–time analog, Fractional Gaussian Noise [1], on-off sources with subexponential activity periods [15] (and references therein), and the $M|G|\infty$ traffic model with subexponential session duration [28].

Under these new models the buffer distribution displays much heavier tails than the *exponential* tails typically associated with short–range dependent Markovian models. Thus, from these analyses emerges theoretical support for the recommendation that in networks carrying long-range dependent traffic, buffers should be provisioned more generously than would otherwise be the case with short-range dependent traffic.

C. Positive correlations

This recommendation is often based on asymptotic results of the form

$$\lim_{B\to\infty} \frac{1}{v(B)} \ln \mathbf{P}\left[Q > B\right] = -\gamma \tag{4}$$

with constant $\gamma > 0$ and monotone function $v : (0,\infty) \to (0,\infty)$ increasing at infinity. Of course, γ and v are determined from the statistics of the input sequence $\{A_t,\ t = 0, 1, \ldots\}$ to the buffer dynamics (2) – Typical examples include $v(B) = B$, $v(B) = B^\beta$ ($\beta > 0$) and $v(B) = \ln B$ [26] (and references therein).

Thus, (4) implies tails of the form

$$\mathbf{P}\left[Q > B\right] \sim e^{-v(B)(\gamma+o(1))} \quad (B \to \infty) \tag{5}$$

with more detailed information on the tail of Q rarely available as closed-form expressions are simply not known, or hard to come by due to the inherent computational complexity of these models. However, in most traffic models known to the authors for which (4) has been developed, these asymptotics already suggest the following: Assume the input process $\{A_t,\ t = 0, 1, \ldots\}$ to be *positively correlated*, say associated [Definition 12], and let $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ denote its independent version [Definition 11]. Then, the corresponding buffer content processes $\{Q_t,\ t = 0, 1, \ldots\}$ and $\{\hat{Q}_t,\ t = 0, 1, \ldots\}$ are "ordered" in some suitable stochastic sense (to be defined shortly), with $\hat{Q}$ "smaller" than Q (where $\hat{Q}$ and Q denote the steady state versions whenever appropriate). In other words, positive correlations lead to increased buffer occupancy and larger buffer levels.

This "folk theorem" has been observed by others, e.g., the simulation study in [18] with the help of the TES modeling tool. Moreover, when Large Deviations arguments are used to validate (4) with $v(B) = B$, the constant γ can often be related to the Large Deviations rate functional of the input sequence $\{A_t,\ t = 0, 1, \ldots\}$, and under association, it is then easy to see that

$$\lim_{B\to\infty} \frac{1}{B} \ln \mathbf{P}\left[\hat{Q} > B\right] = -\hat{\gamma} \tag{6}$$

with $\gamma \leq \hat{\gamma}$. Consequently, $\mathbf{P}\left[\hat{Q} > B\right]$ is less than $\mathbf{P}\left[Q > B\right]$ for large values of B.

D. The results

In this paper we consider this "folk theorem" on a more formal basis with the help of recent advances from the theory of multivariate stochastic orderings [21], [34]: We compare input sequences to (2) in the supermodular (sm) ordering [Definition 4] and the buffer contents in the increasing convex (icx) ordering [Definition 3]. The sm ordering is well suited to capture positive dependence in the components of a random vector, while the icx ordering formalizes comparability in terms of variability and size.

In our discussion we consider three popular (discrete-time) traffic models, namely the Fractional Gaussian Noise traffic model [Section VII], the on-off source model [Section VIII] and the $M|G|\infty$ traffic model [Section IX]. For each of these classes of models we obtain the following result: Let $\{A_t, t = 0, 1, \ldots\}$ and $\{\hat{A}_t, t = 0, 1, \ldots\}$ denote the input traffic process and its independent version [Definitions 7 and 11]. Then, in varying degree of generality, we show that

$$\{\hat{A}_t, t = 0, 1, \ldots\} \leq_{sm} \{A_t, t = 0, 1, \ldots\}. \tag{7}$$

with the independent version being a member of the *same* class of traffic models as the input traffic process. Moreover, the corresponding buffer content processes $\{Q_t, t = 0, 1, \ldots\}$ and $\{\hat{Q}_t, t = 0, 1, \ldots\}$ are icx ordered with

$$\hat{Q}_t \leq_{icx} Q_t, \quad t = 0, 1, \ldots. \tag{8}$$

provided $\hat{Q}_0 = Q_0$. [2] In other words, the independent version does act as a lower bound, thereby providing a formalization of the "folk theorem" mentionned above for these classes of traffic models. For the Fractional Gaussian Noise traffic model, we are also able to show the stronger result [Theorem 24] that increasing the Hurst parameter necessarily increases the variability of the buffer levels. For on-off sources, conditions on the distributions of the on- and off-periods are needed to obtain (7) and (8) [Proposition 25].

The passage from (7) to (8) is a simple consequence of properties of the sm ordering [Theorem 19]. The key idea behind the

[2] As briefly explained in Section VI, the steady state comparison $\hat{Q} \leq_{icx} Q$ is easily derived from (8) in a standard manner whenever appropriate [36], [38].

discussion is a form of positive dependence, known as *stochastic increasingness in sequence*. As shown by Meester and Shanthikumar [21], this notion provides a sufficient condition for (7) to hold [Theorem 15]. While these ideas are applied without too much difficulty to the Fractional Gaussian Noise traffic models and to on-off sources, the discussion is more delicate in the case of $M|G|\infty$ traffic models. Many proofs and details are omitted in the interest of brevity; they are available in [38].

The paper is organized as follows: Some basic notation and definitions are collected in Section II, and stochastic orderings are introduced in Section III. Section IV is devoted to multivariate orderings that capture the dependence structure among the components of random vectors. The key notion of stochastic increasingness in sequence is presented in Section V, and its use for the buffer sizing problem is discussed in Section VI. Finally, Sections VII, VIII and IX discuss the results (7)–(8) for the Fractional Gaussian Noise traffic model, the on-off source and the $M|G|\infty$ traffic model, respectively.

II. Notation and Definitions

A word on the notation in use: Equivalence in law or in distribution between rvs (and stochastic processes) is denoted by $=_{st}$.

For any vector μ in $\mathbb{R}^n$ and for any symmetric non-negative $n \times n$ matrix $\mathbf{\Sigma} = (\Sigma_{ij})$, we write $\mathbf{X} =_{st} \mathcal{N}(\mu, \mathbf{\Sigma})$ to indicate that the $\mathbb{R}^n$-valued rv $\mathbf{X}$ is normally distributed with mean vector μ and covariance matrix $\mathbf{\Sigma}$. For α $(0 < \alpha < 1)$, an $\{1, 2, \ldots\}$-valued rv X is said to a geometric rv with parameter α if it is distributed according to the pmf

$$\mathbf{P}[X = k] = \alpha^{k-1}(1-\alpha), \quad k = 1, 2, \ldots, \tag{9}$$

in which case we write $X =_{st} \mathcal{G}(\alpha)$.

For any $\mathbb{N}$-valued rv X, set

$$\mathcal{S}(X) := \{t = 1, 2, \ldots : \mathbf{P}[X \geq t] > 0\} \tag{10}$$

and define the *hazard* function (also known as the *failure rate* function) of the rv X by

$$h_X(t) = \frac{\mathbf{P}[X = t]}{\mathbf{P}[X \geq t]}, \quad t \in \mathcal{S}(X).$$

Definition 1: We say that the $\mathbb{N}$-valued rv X is increasing failure rate (IFR) (resp. decreasing failure rate (DFR)) if the mapping $\mathcal{S}(X) \to \mathbb{R}_+ : t \to h_X(t)$ is increasing (resp. decreasing).

If the $\mathbb{N}$-valued rv X has finite mean, we define its *forward recurrence time* $\hat{X}$ to be the $\mathbb{N}$-valued rv with pmf given by

$$\mathbf{P}\left[\hat{X} = t\right] = \frac{\mathbf{P}[X \geq t]}{\mathbf{E}[X]}, \quad t = 0, 1, \ldots. \tag{11}$$

Note that $\mathbf{P}\left[\hat{X} \geq t\right] = 0$ if and only if $\mathbf{P}[X \geq t] = 0$, and we conclude $\mathcal{S}(\hat{X}) = \mathcal{S}(X)$.

Throughout, increasing (resp. decreasing) is to be understood to mean non-decreasing (resp. non-increasing).

III. Stochastic orderings

In this section, we summarize basic definitions concerning the stochastic orderings of random vectors. Additional information can be found in the monographs by Shaked and Shanthikumar [35], and by Stoyan [36].

Definition 2: Let Φ be a class of Borel measurable functions $\varphi : \mathbb{R}^n \to \mathbb{R}$. We say that the two $\mathbb{R}^n$-valued rvs $\mathbf{X}$ and $\mathbf{Y}$ satisfy the relation $\mathbf{X} \leq_\Phi \mathbf{Y}$ if

$$\mathbf{E}[\varphi(\mathbf{X})] \leq \mathbf{E}[\varphi(\mathbf{Y})] \tag{12}$$

for all functions φ in Φ, whenever the expectations exist.

This generic definition has been specialized in the literature; here are two important examples.

Definition 3: The $\mathbb{R}^n$-valued rvs $\mathbf{X}$ and $\mathbf{Y}$ are said to be ordered according to

- the usual stochastic ordering, written $\mathbf{X} \leq_{st} \mathbf{Y}$, if (12) holds for all increasing functions $\varphi : \mathbb{R}^n \to \mathbb{R}$;
- the increasing convex ordering, written $\mathbf{X} \leq_{icx} \mathbf{Y}$, if (12) holds for all increasing convex functions $\varphi : \mathbb{R}^n \to \mathbb{R}$.

The icx ordering is appropriate for comparing the variability of rvs. However, in the case of random vectors, it is also desirable to compare their "dependence" structures.

IV. Dependence orderings

Several stochastic orderings have been found well suited for comparing the dependence structures of random vectors. Here we rely on the *supermodular* ordering which has recently been used in several queueing and reliability applications [4], [5], [34]. We begin by introducing the class of functions associated with this ordering.

Definition 4: A function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is said to be

- supermodular (sm) if
$$\varphi(\mathbf{x} \vee \mathbf{y}) + \varphi(\mathbf{x} \wedge \mathbf{y}) \geq \varphi(\mathbf{x}) + \varphi(\mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$
where we set $\mathbf{x} \vee \mathbf{y} = (x_1 \vee y_1, \ldots, x_n \vee y_n)$ and $\mathbf{x} \wedge \mathbf{y} = (x_1 \wedge y_1, \ldots, x_n \wedge y_n)$;
- increasing supermodular (ism) if it is increasing in addition to being sm.

We are now ready to define the supermodular orderings.

Definition 5: The $\mathbb{R}^n$-valued rvs $\mathbf{X}$ and $\mathbf{Y}$ are said to be ordered according to

- the supermodular ordering, written $\mathbf{X} \leq_{sm} \mathbf{Y}$, if (12) holds for all supermodular Borel measurable functions $\varphi : \mathbb{R}^n \to \mathbb{R}$;
- the increasing supermodular ordering, written $\mathbf{X} \leq_{ism} \mathbf{Y}$, if (12) holds for all increasing supermodular Borel measurable functions $\varphi : \mathbb{R}^n \to \mathbb{R}$.

It is a simple matter to check that for any $\mathbb{R}^n$-valued rvs $\mathbf{X}$ and $\mathbf{Y}$, the comparison $\mathbf{X} \leq_{sm} \mathbf{Y}$ necessarily implies the stochastic equalities

$$X_i =_{st} Y_i, \quad i = 1, 2, \ldots, n. \tag{13}$$

Additional information on the sm and ism orderings can be found in [4], [5], [21], [24], [34], [37]. In particular, we shall use repeatedly the fact that the sm ordering is closed under convolution.

Lemma 6: *Let $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ be independent $\mathbb{R}^n$-valued rvs. If $\mathbf{X} \leq_{sm} \mathbf{Y}$, then $\mathbf{X} + \mathbf{Z} \leq_{sm} \mathbf{Y} + \mathbf{Z}$.*

Iterating Lemma 6 readily leads to a useful fact contained in Corollary 8, but first, a definition:

Definition 7: For $\mathbb{R}^n$-valued rvs $\mathbf{X}$ and $\hat{\mathbf{X}}$, we say that $\hat{\mathbf{X}} = (\hat{X}_1, \ldots, \hat{X}_n)$ is an independent version of $\mathbf{X} = (X_1, \ldots, X_n)$ if the rvs $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n$ are mutually independent with $\hat{X}_k =_{st} X_k$, $k = 1, \ldots, n$.

Corollary 8: Let $\{\mathbf{X}_k,\ k = 1, 2, \ldots\}$ denote a sequence of mutually independent $\mathbb{R}^n$-valued rvs. For each $k = 1, 2, \ldots$, let $\hat{\mathbf{X}}_k = (\hat{X}_{k1}, \ldots, \hat{X}_{kn})$ denote an independent version of $\mathbf{X}_k$. If $\hat{\mathbf{X}}_k \leq_{sm} \mathbf{X}_k$ for all $k = 1, 2, \ldots$, then for each $N = 1, 2, \ldots$, the rv $\sum_{k=1}^N \hat{\mathbf{X}}_i$ is an independent version of $\sum_{k=1}^N \mathbf{X}_k$ and

$$\sum_{k=1}^N \hat{\mathbf{X}}_k \leq_{sm} \sum_{k=1}^N \mathbf{X}_k. \tag{14}$$

We also note [24, Thm. 3.1, p. 112]

Lemma 9: *Let $\{\mathbf{X}_i,\ i = 1, 2, \ldots\}$ and $\{\mathbf{Y}_i,\ i = 1, 2, \ldots\}$ denote two sequences of $\mathbb{R}^n$-valued rvs such that $\mathbf{X}_n \Longrightarrow_n \mathbf{X}_\infty$ and $\mathbf{Y}_n \Longrightarrow_n \mathbf{Y}_\infty$. If $\mathbf{X}_n \leq_{sm} \mathbf{Y}_n$ for each $n = 1, 2, \ldots$, then $\mathbf{X}_\infty \leq_{sm} \mathbf{Y}_\infty$.*

Finally, we shall find it useful to extend some of these definitions to sequences of rvs.

Definition 10: We say that the two $\mathbb{R}$-valued sequences $\mathbf{X} = \{X_n,\ n = 1, 2, \ldots\}$ and $\mathbf{Y} = \{Y_n,\ n = 1, 2, \ldots\}$ satisfy the relation $\mathbf{X} \leq_{sm} \mathbf{Y}$ if $(X_1, \ldots, X_n) \leq_{sm} (Y_1, \ldots, Y_n)$ for all $n = 1, 2, \ldots$.

Definition 11: For sequences of $\mathbb{R}$-valued rvs $\mathbf{X} = \{X_n,\ n = 1, 2, \ldots\}$ and $\hat{\mathbf{X}} = \{\hat{X}_n,\ n = 1, 2, \ldots\}$, we say that $\hat{\mathbf{X}}$ is an independent version of $\mathbf{X}$ if the rvs $\{\hat{X}_n,\ n = 1, 2, \ldots\}$ are mutually independent with $\hat{X}_n =_{st} X_n$ for all $n = 1, 2, \ldots$.

V. Positive dependence

Positive dependence in a collection of rvs can be captured in several ways. The association of rvs is one of the most useful such characterizations; it was introduced by Esary, Proschan and Walkup [11] and has proved useful in various settings [3], [9], [16].

Definition 12: The $\mathbb{R}$-valued rvs $\{X_1, \ldots, X_n\}$ are said to be associated if, with $\mathbf{X} = (X_1, \ldots, X_n)$, the inequality

$$\mathbf{E}\left[f(\mathbf{X})g(\mathbf{X})\right] \geq \mathbf{E}\left[f(\mathbf{X})\right] \mathbf{E}\left[g(\mathbf{X})\right]$$

holds for all increasing functions $f, g : \mathbb{R}^n \to \mathbb{R}$ for which the expectations exist and are finite.

Here, we focus on a stronger notion of positive dependence:

Definition 13: The $\mathbb{R}$-valued rvs $\{X_1, \ldots, X_n\}$ are said to be sequentially stochastically increasing (SSI) if for each $k = 1, 2, \ldots, n-1$, the family of conditional distributions $\{[X_{k+1} | X_1 = x_1, \ldots, X_k = x_k]\}$ is stochastically increasing in $\mathbf{x} = (x_1, \ldots, x_k)$.

More precisely, this definition states that for each $k = 1, 2, \ldots, n-1$, for $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^k$ with $\mathbf{x} \leq \mathbf{y}$ componentwise, it holds that

$$[X_{k+1} | (X_1, \ldots, X_k) = \mathbf{x}] \leq_{st} [X_{k+1} | (X_1, \ldots, X_k) = \mathbf{y}]$$

where $[X_{k+1} | (X_1, \ldots, X_k) = \mathbf{x}]$ denotes any rv distributed according to the conditional distribution of X_{k+1} given $(X_1, \ldots, X_k) = \mathbf{x}$ (with a similar interpretation for $[X_{k+1} | (X_1, \ldots, X_k) = \mathbf{y}]$).

These definitions can be extended to sequences in a natural way along the lines of Definition 10:

Definition 14: We say that the $\mathbb{R}$-valued sequence $\mathbf{X} = \{X_n,\ n = 1, 2, \ldots\}$ is SSI (resp. associated) if for each $n = 1, 2, \ldots$, the rvs $\{X_1, \ldots, X_n\}$ are SSI (resp. associated).

If the $\mathbb{R}$-valued rvs $\{X_1, \ldots, X_n\}$ are SSI, then they are necessarily associated [3, Thm. 4.7, p. 146] but the converse may not be true. The next result was established in [21], and relates the SSI property of rvs to the supermodular ordering. This fact will prove crucial for subsequent developments in this paper:

Theorem 15: *If the $\mathbb{R}_+$-valued rvs $\{X_1, \ldots, X_n\}$ are SSI, then*

$$(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n) \leq_{sm} (X_1, X_2, \ldots, X_n), \tag{15}$$

where $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n)$ is the independent version of $(X_1, X_2, \ldots, X_n)$.

VI. The buffer sizing problem

We now return to the buffer model with infinite buffer capacity and constant service rate of c packets/slot: If $\{A_t,\ t = 0, 1, \ldots\}$ denotes the input traffic feeding into the system, then the buffer content sequence $\{Q_t,\ t = 0, 1, \ldots\}$ is characterized by the Lindley recursion (2) (reproduced here for easy reference)

$$Q_0 = q; \quad Q_{t+1} = [Q_t + A_t - c]^+, \quad t = 0, 1, \ldots \tag{16}$$

for some fixed initial condition q.

For each $t = 1, 2 \ldots$, it is plain that the buffer content Q_t is a function of the input traffic $A_0, \ldots, A_{t-1}$ (and of the initial condition q). Thus, $Q_t = T_t(A_0, \ldots, A_{t-1}, Q_0)$ for some mapping $T_t : \mathbb{R}^t \times \mathbb{R} \to \mathbb{R}$. This function is readily obtained by iterating the mapping $T : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$T(q, a) := (q + a - c)^+, \quad (q, a) \in \mathbb{R}^2.$$

As we have in mind to apply Theorem 15, we are interested in the supermodularity of the mappings $\{T_t,\ t = 1, 2, \ldots\}$ (with $T_1 = T$). The main result along these lines is contained in

Proposition 16: *For each $t = 1, 2, \ldots$ and each q in $\mathbb{R}$, the mapping $\mathbb{R}^t \to \mathbb{R} : (a_0, \ldots, a_{t-1}) \to T_t(a_0, \ldots, a_{t-1}, q)$ is ism.*

Proposition 16 is readily established by induction (on t) with the help of the following fact due to Bäuerle [4].

Lemma 17: *For any ism function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and any icx function $g : \mathbb{R} \to \mathbb{R}$, the composition $g \circ \varphi : \mathbb{R}^n \to \mathbb{R}$ is ism.*

The following consequence of Lemma 17 is straightforward:

Lemma 18: *If $\mathbf{X} \leq_{sm} \mathbf{Y}$, then $\varphi(\mathbf{X}) \leq_{icx} \varphi(\mathbf{Y})$ for any ism Borel measurable mapping $\varphi : \mathbb{R}^n \to \mathbb{R}$.*

Combining Proposition 16 and Lemma 18, we have a useful result already obtained by Bäuerle [4].

Theorem 19: *Let $\{A_t^1,\ t = 0, 1, \ldots\}$ and $\{A_t^2,\ t = 0, 1, \ldots\}$ be input traffic processes to the discrete-time single server queue (16). If*

$$\{A_t^1,\ t = 0, 1, \ldots\} \leq_{sm} \{A_t^2,\ t = 0, 1, \ldots\},$$

then the corresponding buffer contents $\{Q_t^1,\ t=0,1,\ldots\}$ *and* $\{Q_t^2,\ t=0,1,\ldots\}$ *are ordered in the icx ordering, i.e., for any fixed initial condition* $Q_0^1=Q_0^2=q$*, we have* $Q_t^1 \leq_{icx} Q_t^2$ *for all* $t=1,2,\ldots$*.*

With additional assumptions on the input sequences, the transient results of Theorem 19 can be carried over to steady state:

Theorem 20: *Under the assumptions of Theorem 19, if the input sequence* $\{A_t^i,\ t=0,1,\ldots\}$ *is stationary and ergodic with* $\mathbf{E}\left[A_0^i\right] < c$*,* $i=1,2$*, then* $Q_t^i \Rightarrow_t Q^i$*, independently of the initial condition, and the comparison* $Q^1 \leq_{icx} Q^2$ *holds.*

This steady state result is made possible by the fact that the buffer sequence $\{Q_t^i, t=0,1,\ldots\}$ is stochastically increasing with t if $Q_0^i=0$, $i=1,2$ [19], In other words, $Q_t^i \leq_{st} Q_{t+1}^i$ for all $t=0,1,\ldots$ and the desired result follows by using this monotonicity in the proof of the stability of the icx ordering under weak convergence [36, Prop. 1.3.2, p. 10]. A non-trivial comparison is obtained when the limiting rvs Q^1 and Q^2 have finite first moments. However, in all cases of interest here, this finiteness property can be established under mild moment conditions on the input sequences. In what follows, in the interest of brevity, we shall only discuss the transient results; full details concerning the steady state results are available in [38].

VII. Fractional Gaussian Noise traffic

A detailed treatment of Fractional Gaussian Noise (FGN) can be found in [33]. Its use for traffic modeling is discussed in [25] and in [26] (and references therein).

A. Fractional Gaussian Noise (FGN)

With $0<H<1$, *Fractional Gaussian Noise* with Hurst parameter H is a zero-mean stationary Gaussian sequence $\{N_t^H,\ t=0,1,\ldots\}$ with (auto)covariance function

$$\Gamma_H(k)=\frac{\sigma^2}{2}(|k+1|^{2H}-2|k|^{2H}+|k-1|^{2H}),\quad k=0,1,\ldots \tag{17}$$

for some $\sigma^2>0$. We refer to this sequence by FGN(H).

We consider only the range $0.5 \leq H < 1$, which corresponds to positive correlations as was found appropriate for network traffic modeling. When $0.5<H<1$, the asymptotics [33]

$$\Gamma_H(k) \sim \sigma^2 H(2H-1)k^{2H-2} \quad (k \to \infty)$$

show that FGN(H) exhibits long-range dependence [6]. It is also clear from (17) that FGN(H) is an exactly second-order self-similar sequence, thus a self-similar sequence, since it is a Gaussian sequence.

The FGN(H) traffic model we use as input traffic is the sequence $\{A_t^H,\ t=0,1,\ldots\}$ defined by

$$A_t^H = m + N_t^H, \quad t=0,1,\ldots$$

where $\{N_t^H,\ t=0,1,\ldots\}$ is FGN(H) ($0.5 \leq H < 1$) and the scalar m denotes the average traffic rate. The sequence $\{A_t^H,\ t=0,1,\ldots\}$ is thus also a stationary Gaussian process with $\mathbf{E}\left[A_t^H\right]=m$ ($t=0,1,\ldots$) and with covariance function still given by (17). Therefore, $\{A_t^H, t=0,1,\ldots\}$ is also a self-similar process.

The independent version $\{\hat{A}_t,\ t=0,1,\ldots\}$ of FGN(H) must be a sequence of i.i.d. Gaussian rvs with $\mathbf{E}\left[\hat{A}_t\right]=m$ ($t=0,1,\ldots$). Since $\Gamma_H(0)=\sigma^2$, its covariance function is given by $\hat{\Gamma}(k)=\sigma^2\delta(k)$ where $\delta(k)=1$ when $k=0$ and $\delta(k)=0$ when $k\neq 0$. Equivalently, this independent process is simply FGN(H) with $H=0.5$. Indeed, when $H=0.5$ in (17), we get $\Gamma_H(k)=0$ for all $k=1,2,\ldots$ and $\{N_t^{0.5},\ t=0,1,\ldots\}$ is indeed a sequence of i.i.d. Gaussian rvs, and so is $\{A_t^{0.5},\ t=0,1,\ldots\}$. Since the independent version of FGN(H) does not depend on H, it is appropriate to simply refer to it as $\{\hat{A}_t,\ t=0,1,\ldots\}$ without further reference to H in the notation.

B. Comparisons for FGN traffic models

For each $t=0,1,\ldots$, $[A_{t+1}^H|A_0^H,\ldots,A_t^H]$ is normally distributed. Since a Gaussian rv is stochastically increasing in the mean [35], the SSI property will follow if we can show that the conditional mean $\mathbf{E}\left[A_{t+1}^H|A_0^H=a_0,\ldots,A_t^H=a_t\right]$ is an increasing function in $(a_0,\ldots,a_t)$. Although the covariance function of the underlying sequence is explicitly given, we were unable to obtain usable closed-form expressions for these conditional expectations. This is due to the very complicated structure of the involved matrices (and their inverses). Instead, we turn to the comprehensive characterization of stochastic orderings given for Gaussian rvs by Müller [23, Thm. 3.8].

Theorem 21: *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be* $\mathbb{R}^n$*-valued rvs such that* $\mathbf{X} =_{st} \mathcal{N}(\mu,\mathbf{\Sigma})$ *and* $\mathbf{Y} =_{st} \mathcal{N}(\mu',\mathbf{\Sigma}')$*. Then* $\mathbf{X} \leq_{sm} \mathbf{Y}$ *if and only if* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same marginals and* $\Sigma_{ij} \leq \Sigma'_{ij}$ *for all* $i,j=1,\ldots,n$*.*

From (17), with $0.5 \leq H < 1$, $\Gamma_H(0)=\sigma^2$ and $\Gamma_H(k)\geq 0$ for all $k=1,2,\ldots$. Moreover, $\mathbf{E}\left[\hat{A}_t\right]=\mathbf{E}\left[A_t^H\right]=m$ ($t=0,1,\ldots$). As a direct application of Theorem 21, we conclude that the independent version (FGN(H) with $H=0.5$) is indeed a lower bound process for the FGN(H) with $0.5 \leq H < 1$.

Theorem 22: *Let* $\{A_t^H,\ t=0,1,\ldots\}$ *be a FGN(H) traffic model with parameter* $0.5 \leq H < 1$*. Its independent version* $\{\hat{A}_t,\ t=0,1,\ldots\}$ *coincides with the FGN(*0.5*) traffic model, and satisfies*

$$\{\hat{A}_t,\ t=0,1,\ldots\} \leq_{sm} \{A_t^H,\ t=0,1,\ldots\}.$$

Moreover, the corresponding buffer contents $\{\hat{Q}_t,\ t=0,1,\ldots\}$ *and* $\{Q_t^H,\ t=0,1,\ldots\}$ *are ordered in the icx ordering, i.e., for any fixed initial condition* $\hat{Q}_0=Q_0^H=q$*, we have* $\hat{Q}_t \leq_{icx} Q_t^H$ *for all* $t=0,1,\ldots$*.*

In fact, Theorem 21 makes it possible to compare two FGN traffic models with different Hurst parameters H and H' in $(0.5,1)$ such that $H'<H$. To do this, we use elementary calculus to derive a simple monotonicity result for the covariance function (17).

Lemma 23: *For each* $k=0,1,\ldots$*, the mapping* $H \to \Gamma_H(k)$ *given by (17) is monotone increasing on the interval* $(0.5,1)$*.*

The following strengthening of Theorem 22 is now within reach.

Theorem 24: *With H and H' in the interval $(0.5, 1)$ such that $H' < H$, we have the comparison*

$$\{A_t^{H'},\ t = 0, 1, \ldots\} \leq_{sm} \{A_t^{H},\ t = 0, 1, \ldots\},$$

between the FGN traffic models with parameter H' and H, respectively. Moreover, for the corresponding buffer contents $\{Q_t^{H'},\ t = 0, 1, \ldots\}$ and $\{Q_t^{H},\ t = 0, 1, \ldots\}$, it holds for any fixed initial condition $Q_0^{H'} = Q_0^{H} = q$ that $Q_t^{H'} \leq_{icx} Q_t^{H}$ for all $t = 0, 1, \ldots$.

Proof. The comparison in the sm ordering follows by the combined application of Theorem 21 and Lemma 23 under the condition $H' < H$. The comparison in the icx ordering is now obtained via Theorem 19. ■

From Theorem 22, we can conclude, as expected, that the long-range dependent traffic ($0.5 < H < 1$) requires more buffer space than the short-range dependent traffic ($H = 0.5$). Moreover, when $H' < H$, $\Gamma_{H'}(k) \leq \Gamma_H(k)$, i.e., FGN($H$) is more correlated than FGN(H'), and by Theorem 24, the more correlated the traffic, the larger the buffer space needed to meet the same QoS requirement.

VIII. ON–OFF SOURCES

A. Modeling on-off sources

A discrete-time on-off source with peak rate r is described by a succession of cycles, each such cycle comprising an off-period followed by an on-period. During the on-periods the source is active and produces traffic at the constant rate of r (packet/slot) [3]; the source is silent during the off-periods: For each $n = 0, 1, \ldots$, let B_n and I_n denote the durations (in timeslots) of the on-period and off-period in the $(n+1)^{st}$ cycle, respectively. Thus, if the epochs $\{T_n,\ n = 0, 1, \ldots\}$ denote the beginning of successive cycles, with $T_0 := 0$ we have $T_{n+1} := \sum_{\ell=0}^{n} I_\ell + B_\ell$ ($n = 0, 1, \ldots$). The activity of the source is then described by the $\{0, 1\}$-valued process $\{A_t,\ t = 0, 1, \ldots\}$ given by

$$A_t := \sum_{n=0}^{\infty} \mathbf{1}\left[T_n + I_n \leq t < T_{n+1}\right] \tag{18}$$

for all $t = 0, 1, \ldots$, with the source active (resp. silent) during timeslot $[t, t+1)$ if $A_t = 1$ (resp. $A_t = 0$).

An *independent on-off* source is one for which (i) the $\{1, 2, \ldots\}$-valued rvs $\{I_n,\ n = 1, \ldots\}$ and $\{B_n,\ n = 1, \ldots\}$ are mutually independent rvs which are independent of the pair of rvs I_0 and B_0 associated with the initial cycle; and (ii) the rvs $\{I_n,\ n = 1, \ldots\}$ (resp. $\{B_n,\ n = 1, \ldots\}$) are i.i.d. rvs with generic off-period duration rv I (resp. on-period duration rv B). Throughout the generic rvs B and I are assumed to be independent $\{1, 2, \ldots\}$-valued rvs such that $0 < \mathbf{E}[B], \mathbf{E}[I] < \infty$, and we simply refer to the independent on-off process just defined as the on-off source (B, I).

In general, the activity process (18) is not stationary unless the $\mathbb{N}$-valued rvs I_0 and B_0 are selected appropriately. One possible way is to use the following variation on constructions given in [2], [32]: With

$$p := \frac{\mathbf{E}[B]}{\mathbf{E}[B] + \mathbf{E}[I]}, \tag{19}$$

we introduce the $\{0, 1\}$-valued rv U given by

$$\mathbf{P}[U = 1] = p = 1 - \mathbf{P}[U = 0]. \tag{20}$$

Let $\hat{B}$ and $\hat{I}$ denote two $\{1, 2, \ldots\}$-valued rvs distributed according to the forward recurrence time (11) associated with B and I, respectively. A stationary version of (18), still denoted $\{A_t,\ t = 0, 1, \ldots\}$, is now obtained by selecting (I_0, B_0) so that

$$(I_0, B_0) =_{st} (0, \hat{B})U + (\hat{I}, B)(1 - U) \tag{21}$$

with rvs U, B, $\hat{B}$ and $\hat{I}$ taken to be mutually independent and independent of the rvs $\{B_n, I_n,\ n = 1, \ldots\}$. In that case, we see that

$$\mathbf{P}[A_t = 1] = 1 - \mathbf{P}[A_t = 0] = p, \quad t = 0, 1, \ldots$$

where p is the average rate (19).

Thus, the independent version of the stationary on-off process is simply a sequence $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ of i.i.d. $\{0, 1\}$-valued rvs, with

$$\mathbf{P}\left[\hat{A}_t = 1\right] = 1 - \mathbf{P}\left[\hat{A}_t = 0\right] = p, \quad t = 0, 1, \ldots$$

where p is as above. It is easily seen that $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ is also a stationary on-off process with geometric on-period and off-period, i.e., the corresponding on-period duration rv B (respectively, off-period duration rv I) is geometrically distributed with parameter p (respectively, $1 - p$), i.e.,

$$B =_{st} \mathcal{G}(p) \quad \text{and} \quad I =_{st} \mathcal{G}(1 - p)$$

in the notation (9). In other words, $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ can be interpreted as the discrete-time stationary on-off process $(\mathcal{G}(p), \mathcal{G}(1 - p))$.

B. Lower bounds for on-off sources

In [38], several sets of conditions on B and I were derived for the discrete-time on-off source (B, I) to have the SSI property. One such set of conditions is presented next:

Proposition 25: *The discrete-time stationary on-off source (B, I) satisfies the SSI property if the conditions (i)-(iv) below hold, where*

(i) *The rvs B and I are DFR;*
(ii) *The rvs $\hat{B}$ and $\hat{I}$ are DFR;*
(iii) $\mathbf{E}[B]^{-1} + \mathbf{E}[I]^{-1} \leq 1$*;*
(iv) $\mathbf{P}[B = 1] + \mathbf{P}[I = 1] \leq 1$.

The proof of Proposition 25 relies on calculations that are available in [38]. Upon combining Proposition 25 with Theorem 19, we have

Theorem 26: *Let $\{A_t,\ t = 0, 1, \ldots\}$ be a discrete-time on-off source (B, I) satisfying the conditions of Proposition 25. Its independent version $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ is a sequence of i.i.d.*

[3] For simplicity, we set this rate to be unity, say one packet/slot, i.e., $r = 1$.

$\{0,1\}$-valued rvs with $\mathbf{P}\left[\hat{A}_t = 1\right] = p$ for all $t = 0, 1, \ldots$, and we have the comparison

$$\{\hat{A}_t,\ t = 0, 1, \ldots\} \leq_{sm} \{A_t,\ t = 0, 1, \ldots\}.$$

Moreover, the corresponding buffer contents $\{\hat{Q}_t,\ t = 0, 1, \ldots\}$ and $\{Q_t,\ t = 0, 1, \ldots\}$ are ordered in the icx ordering, i.e., for any fixed initial condition $\hat{Q}_0 = Q_0 = q$, we have $\hat{Q}_t \leq_{icx} Q_t$ for all $t = 0, 1, \ldots$.

To close the discussion for the class of on-off sources, we observe that the steady state comparison $\hat{Q} \leq_{icx} Q$ can be established under weaker conditions than the ones used in Theorem 26. Indeed, fewer conditions are needed to establish a version of the pivotal Proposition 25 for the non-stationary (i.e., non-delayed) version of the on-off source (B, I); details can be found in [38].

IX. $M|G|\infty$ INPUT TRAFFIC

The discrete-time $M|G|\infty$ input traffic is simply the number of busy servers in the infinite server system fed by a discrete-time Poisson process with rate λ (customers per timeslot) and with generic service time S (expressed in timeslots). A more detailed treatment of $M|G|\infty$ input processes can be found in [20], [27], [28], [29], [30]. This process is a versatile class of input traffic since both short-range and long-range dependent traffic can be generated by properly selecting the service distribution.

A. Modeling $M|G|\infty$ traffic sources

Consider a system of infinitely many servers, and suppose B_t customers arrive to the system during timeslot $[t-1, t)$ ($t = 1, 2, \ldots$). Customer i ($i = 1, 2, \ldots, B_t$) is assigned its own server from which it starts receiving service with duration $S_{t,i}$ (number of timeslots) in timeslot $[t, t+1)$. If there are b initial customers present in the system at time $t = 0$, initial customer i ($i = 1, 2, \ldots, b$) will have service time duration $S_{0,i}$ (starting at $t = 0$). Let A_t be the number of busy servers, or equivalently, the number of customers still present at the beginning of the timeslot $[t, t+1)$. The busy server process $\{A_t,\ t = 0, 1, \ldots\}$ defines the $M|G|\infty$ input process.

To define the stationary and ergodic version of the $M|G|\infty$ input process, we need to make some assumptions on the $\mathbb{N}$-valued rvs b, $\{B_t,\ t = 1, 2, \ldots\}$, $\{S_{t,i},\ t = 1, 2, \ldots, i = 1, 2, \ldots\}$ and $\{S_{0,i}, i = 1, 2, \ldots\}$: (i) These rvs are mutually independent; (ii) The rv b is Poisson distributed with mean $\lambda \mathbf{E}[S]$; (iii) The rvs $\{B_t,\ t = 1, 2, \ldots\}$ are i.i.d. Poisson rvs with mean $\lambda > 0$; (iv) The rvs $\{S_{t,i},\ t = 1, 2, \ldots, i = 1, 2, \ldots\}$ are i.i.d. with common pmf G on $\{1, 2, \ldots\}$. Let S be a generic rv distributed according to the pmf G and assume throughout that $\mathbf{E}[S] < \infty$; and (v) The rvs $\{S_{0,i}, i = 1, 2, \ldots\}$ are i.i.d. $\{1, 2, \ldots\}$-valued rvs with common pmf $\hat{G}$ which is the forward recurrence pmf associated with G via (11). Let $\hat{S}$ be a generic $\mathbb{N}$-valued rv distributed according to $\hat{G}$.

Hereafter, by an $M|G|\infty$ input process we mean the stationary and ergodic version, still denoted $\{A_t,\ t = 0, 1, \ldots\}$, which is determined by the conditions above. Also, we shall write $\{\hat{S}_i,\ i = 1, 2, \ldots,\}$ instead of $\{S_{0,i},\ i = 1, 2, \ldots,\}$. Since the $M|G|\infty$ process can be characterized by two parameters, namely λ and S, we refer to it as the $M|G|\infty$ input process (λ, S). The next proposition summarizes needed properties of the stationary $M|G|\infty$ input process $\{A_t,\ t = 0, 1, \ldots\}$ [29].

Proposition 27: *Under assumptions (i)-(v) above, the $M|G|\infty$ input process (λ, S) is a (strictly) stationary and ergodic process $\{A_t,\ t = 0, 1, \ldots\}$ with the following properties: (i) For each $t = 0, 1, \ldots$, the rv A_t is a Poisson rv with parameter $\lambda \mathbf{E}[S]$; and (ii) Its covariance function is given by*

$$\mathrm{cov}(A_t, A_{t+h}) = \lambda \mathbf{E}[S]\, \mathbf{P}\left[\hat{S} > h\right], \quad h = 0, 1, \ldots.$$

for all $t = 0, 1, \ldots$.

We now argue that the independent version of $M|G|\infty$ input process (λ, S) is also an $M|G|\infty$ input process, say (λ^0, S^0), where λ^0 and S^0 are properly selected. Indeed, if we take $S^0 \equiv 1$, then each customer (each session) requires exactly one timeslot of service before leaving the system at the end of that timeslot. Therefore, the number of customers in the system at the beginning of timeslot $[t, t+1)$ is simply the number of customers who arrive in timeslot $[t-1, t)$ independently of arrivals in the past and future timeslots. Let $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ denote the $M|G|\infty$ input process $(\lambda^0, S^0 \equiv 1)$ as specified above. By the first part of the discussion, it is plain that the rvs $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ are mutually independent, in agreement with Claim (ii) of Proposition 27 which would yield in that case

$$\mathrm{cov}(\hat{A}_t, \hat{A}_{t+h}) = \lambda^0 \mathbf{P}\left[\hat{S^0} > h\right] = \lambda^0 \delta(0, h)$$

for all $t, h = 0, 1, \ldots$. By Claim (i) of Proposition 27, for each $t = 0, 1, \ldots$, the rv A_t is Poisson rv with parameter $\lambda \mathbf{E}[S]$ and $\hat{A}_t$ is a Poisson rv with parameter λ^0. Thus, the marginals of the sequence $\{A_t,\ t = 0, 1, \ldots\}$ for the given $M|G|\infty$ input process (λ, S) will coincide with those of its independent version $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ provided $\lambda^0 = \lambda \mathbf{E}[S]$. Combining these observations we conclude that the independent version of the $M|G|\infty$ input process (λ, S) is the $M|G|\infty$ input process with $(\lambda \mathbf{E}[S], 1)$.

B. Lower bounds for $M|G|\infty$ models

We now turn to establishing (7) and (8) for $M|G|\infty$ input processes. Unfortunately, we were not able to show directly that $M|G|\infty$ input processes are SSI [38], although they *are* associated [27], [29]. Instead we took our cue from Corollary 8 to the effect that the sm ordering is stable under independent summation: This property suggests the following very natural approach, whereby we seek an additive decomposition of the $M|G|\infty$ input process into several *independent* components, each with the SSI property. In that case, the independent version of each component will act as a lower bound to the corresponding component in the sm ordering. The sum of the independent versions of the component processes is statistically indistinguishable from $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ and satisfies (7) and the desired result will then follow from Corollary 8.

To carry out this program, fix $t = 0, 1, \ldots$. From the system description in Section IX-A, we can write

$$A_t = A_t^{(0)} + A_t^{(a)} \tag{22}$$

where $A_t^{(0)}$ and $A_t^{(a)}$ are the numbers of busy servers in the system at the beginning of the timeslot $[t, t+1)$ contributed by the initial customers and new arrivals during $[0, t)$, respectively. From the b initial customers, customer i ($i = 1, 2, \ldots, b$) will be in the system at the beginning of timeslot $[t, t+1)$ if and only if $\hat{S}_i > t$, whence

$$A_t^{(0)} = \sum_{i=1}^{b} \mathbf{1}\left[\hat{S}_i > t\right]. \tag{23}$$

Similarly, it is easy to see [27], [29] that

$$A_t^{(a)} = \sum_{r=1}^{t} \sum_{i=1}^{B_r} \mathbf{1}\left[S_{r,i} > t - r\right] = \sum_{r=1}^{\infty} A_t^{(r)} \tag{24}$$

where for each $r = 1, 2, \ldots$, the sequence $\{A_t^{(r)},\ t = 0, 1, \ldots\}$ corresponds to those B_r customers who arrive in timeslot $[r-1, r)$, i.e., for all $t = 0, 1, \ldots$, we have

$$A_t^{(r)} = \mathbf{1}\left[t \geq r\right] \sum_{i=1}^{B_r} \mathbf{1}\left[S_{r,i} > t - r\right]. \tag{25}$$

The sequences $\{A_t^{(0)},\ t = 0, 1, \ldots\}$ and $\{A_t^{(r)}, r = 1, 2 \ldots\}$, $r = 1, 2, \ldots$, are *mutually independent* and display a *very similar structure*. To exploit this fact we shall make use of the following result:

Proposition 28: *Let K be an $\mathbb{N}$-valued rv which is indepedent of the i.i.d. $\{1, 2, \ldots\}$-valued rvs $\{\xi, \xi_k,\ k = 1, 2, \ldots\}$. Then, the sequence $\{\sum_{k=1}^{K} \mathbf{1}\left[\xi_k > t\right],\ t = 0, 1, \ldots\}$ is SSI.*

Proof. Write $X_t = \sum_{k=1}^{K} \mathbf{1}\left[\xi_k > t\right]$ for all $t = 0, 1, \ldots$, and let $Bin(N, p)$ denote the Binomial distribution with paremeters N ($N = 1, 2, \ldots$) and p ($0 \leq p \leq 1$). Elementary arguments [38] show that for each $t = 0, 1, \ldots$,

$$[X_{t+1} | X_0 = x_0, \ldots, X_t = x_t] =_{st} Bin(N, p)$$

for all $(x_0, \ldots, x_t)$ in $\mathbb{N}^{t+1}$ such that

$$\mathbf{P}\left[X_0 = x_0, \ldots, X_t = x_t\right] > 0,$$

with $N = x_t$ and $p = \mathbf{P}\left[\xi > t + 1 | \xi > t\right]$. The result is obtained by recalling that for a given p, the Binomial rv $Bin(N, p)$ is increasing in N in the st ordering [35]. ■

Proposition 28 leads to the next two lemmas.

Lemma 29: *The sequence $\{A_t^{(0)},\ t = 0, 1, \ldots\}$ is SSI.*

Proof. The result is a direct consequence of Proposition 28 once we note that the sequence $\{A_t^{(0)},\ t = 0, 1, \ldots\}$ has the same structure as that introduced in Proposition 28, with $K = b$ and $\xi = \hat{S}$. ■

Lemma 30: *For each $r = 1, 2, \ldots$, the sequence $\{A_t^{(r)},\ t = 0, 1, \ldots\}$ is SSI.*

Proof. Fix $r = 1, 2, \ldots$. Note that $A_t^{(r)} = 0$ whenever $t = 0, \ldots, r-1$, but $A_r^{(r)} =_{st} B_r$. Hence, the SSI property of the sequence $\{A_t^{(r)},\ t = 0, 1, \ldots\}$ is equivalent to that of the sequence $\{A_{t+r}^{(r)},\ t = 0, 1, \ldots\}$. This latter sequence of rvs has the same structure as that introduced in Proposition 28, with $K = B_r$ and $\xi = S$, and the desired result follows. ■

Collecting the various strands of the discussion thus far, we get the following result:

Theorem 31: *Let $\{A_t,\ t = 0, 1, \ldots\}$ be an $M|G|\infty$ input process (λ, S). Its independent version $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ is the $M|G|\infty$ input process $(\lambda \mathbf{E}\left[S\right], 1)$, and we have the comparison*

$$\{\hat{A}_t,\ t = 0, 1, \ldots\} \leq_{sm} \{A_t,\ t = 0, 1, \ldots\}.$$

Moreover, the corresponding buffer contents $\{\hat{Q}_t,\ t = 0, 1, \ldots\}$ and $\{Q_t,\ t = 0, 1, \ldots\}$ are ordered in the icx ordering, i.e., for any fixed initial condition $Q_0 = q$, we have $\hat{Q}_t \leq_{icx} Q_t$ for all $t = 0, 1, \ldots$.

Proof. By Lemma 29 and Theorem 15, we have

$$\{\hat{A}_t^{(0)},\ t = 0, 1, \ldots\} \leq_{sm} \{A_t^{(0)},\ t = 0, 1, \ldots\}$$

where $\{\hat{A}_t^{(0)},\ t = 0, 1, \ldots\}$ is the independent version of $\{A_t^{(0)},\ t = 0, 1, \ldots\}$. On the other hand, by Lemma 30 and Theorem 15, we conclude for each $r = 1, 2, \ldots$ that

$$\{\hat{A}_t^{(r)},\ t = 0, 1, \ldots\} \leq_{sm} \{A_t^{(r)},\ t = 0, 1, \ldots\}$$

where $\{\hat{A}_t^{(r)},\ t = 0, 1, \ldots\}$ denotes the independent version of $\{A_t^{(r)},\ t = 0, 1, \ldots\}$. It is always possible to construct all involved rvs on a single probability triple so that the independent versions are mutually independent. Hence, under the enforced independence assumptions, upon invoking Corollary 8, we obtain

$$\begin{aligned} &\{\hat{A}_t^{(0)} + \sum_{r=1}^{R} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\} \\ \leq_{sm} \quad &\{A_t^{(0)} + \sum_{r=1}^{R} A_t^{(r)},\ t = 0, 1, \ldots\} \end{aligned} \tag{26}$$

for each $R = 1, 2, \ldots$.

Now let R go to infinity in (26), and note the convergence $\{A_t^{(0)} + \sum_{r=1}^{R} A_t^{(r)},\ t = 0, 1, \ldots\} \Longrightarrow_R \{A_t,\ t = 0, 1, \ldots\}$ (via (24)) and $\{\hat{A}_t^{(0)} + \sum_{r=1}^{R} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\} \Longrightarrow_R \{\hat{A}_t^{(0)} + \sum_{r=1}^{\infty} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\}$ (where in fact the limiting rvs exist by pointwise convergence). By Lemma 9, the sm ordering is stable under weak convergence, and we obtain the comparison

$$\{\hat{A}_t^{(0)} + \sum_{r=1}^{\infty} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\} \leq_{sm} \{A_t,\ t = 0, 1, \ldots\}.$$

For each $R = 1, 2, \ldots$, the sequence $\{\hat{A}_t^{(0)} + \sum_{r=1}^{R} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\}$ is an independent version of $\{A_t^{(0)} + \sum_{r=1}^{R} A_t^{(r)},\ t = 0, 1, \ldots\}$. Therefore, in the limit, the sequence $\{\hat{A}_t^{(0)} + \sum_{r=1}^{\infty} \hat{A}_t^{(r)},\ t = 0, 1, \ldots\}$ is statistically indistinguishable of the independent version $\{\hat{A}_t,\ t = 0, 1, \ldots\}$ of $\{A_t,\ t = 0, 1, \ldots\}$ discussed in Section IX-A. In sum, the comparison

$$\{\hat{A}_t,\ t = 0, 1, \ldots\} \leq_{sm} \{A_t,\ t = 0, 1, \ldots\}$$

holds. The second half of the result is now immediate from Theorem 19. ■

REFERENCES

[1] R.G. Addie, M. Zukerman and T. Neame, "Fractal traffic: Measurements, modeling and performance evaluation," in *Proceedings of INFOCOM 95*, Boston (MA), April 1995, pp. 985–992.

[2] R. Agrawal, A.M. Makowski and Ph. Nain, "On a reduced load equivalence for a multi–fluid buffer under heavy tail assumptions," *Queueing Systems – Theory & Applications* **33** (1999), pp. 5–41.

[3] R.E. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, International Series in Decision Processes, Holt, Rinehart and Winston, New York (NY), 1975.

[4] N. Bäuerle, "Monotonicity results for $MR|GI|1$ queues," *Journal of Applied Probability* **34** (1997), pp. 514–524.

[5] N. Bäuerle and T. Rolski, "A monotonicity result for the work-load in Markov-modulated queues," *Journal of Applied Probability* **35** (1998), pp. 741–747.

[6] J. Beran, *Statistics for Long–Memory Processes*, Chapman and Hall, New York (NY), 1994.

[7] J. Beran, R. Sherman, M.S. Taqqu and W. Willinger "Long–range dependence in variable bit–rate video traffic," *IEEE Transactions on Communications* **43** (1995), pp. 1566–1579.

[8] F. Brichet, J. Roberts, A. Simonian, and D. Veitch, "Heavy traffic analysis of a storage model with long range dependent on/off sources," *Queueing Systems – Theory and Applications* **23** (1996), pp. 197–215.

[9] A. Cohen and J.G.B. Kemperman and H.B. Sackrowitz, "Unbiased testing in exponential family regression," *Annals of Statistics* **4** (1994), pp. 1931–1946.

[10] M. Crovella and A. Bestavros, "Self–similarity in World Wide Web traffic: Evidence and possible causes," *Performance Evaluation Review* **24** (1996), pp. 160–169.

[11] J.D. Esary and F. Proschan and D.W. Walkup, "Association of random variables, with applications," *Annals of Mathematical Statistics* **38** (1967), pp. 1466–1474.

[12] H. J. Fowler and W. E. Leland, "Local area network traffic characteristics, with implications for broadband network congestion management," *IEEE Journal on Selected Areas in Communications* **9** (1991), pp. 1139–1149.

[13] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine* **32** (1994), pp. 70–81.

[14] M. Garrett and W. Willinger, "Analysis, modeling and generation of self–similar VBR video traffic," in *Proceedings of the SIGCOMM 94 Conference*, University College London, London (UK), August 1994, pp. 269–280.

[15] P.R. Jelenković, "Asymptotic analysis of queues with subexponential arrivals processes," in *Self–Similar Network Traffic and Performance Evaluation*, Eds. K. Park and W. Willinger, John Wiley & Sons, New York (NY) (2000), pp. 249–268.

[16] K. Joag-Dev and M.D. Perlman and L.D. Pitt, "Association of normal random variables and Slepian's inequality," *Annals of Probability* **11** (1983), pp. 451–455.

[17] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self–similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking* **2** (1994), pp. 1–15.

[18] M. Livny, B. Melamed and A.K. Tsiolis, "The impact of autocorrelation on queuing systems," *Management Science* **39** (1993), pp. 322–339.

[19] R.M. Loynes, "The stability of a queue with non–independent inter–arrival and service times," *Proceedings of the Cambridge Philosophical Society* **58** (1962), pp. 497–520.

[20] A.M. Makowski and M. Parulekar, "Buffer Asymptotics for $M|G|\infty$ input processes." in *Self–Similar Network Traffic and Performance Evaluation*, Eds. K. Park and W. Willinger, John Wiley & Sons, New York (NY) (2000), pp. 215–248.

[21] L.E. Meester and J.G. Shanthikumar, "Regularity of stochastic processes: A theory of directional convexity," *Probability in the Engineering and Informational Sciences* **7** (1993), pp. 343–360.

[22] H. Michiel and K. Laevens, "Teletraffic engineering in a broadband era," *Proceedings of the IEEE* **85** (1997), pp. 2007–2033.

[23] A. Müller, "Stochastic ordering of multivariate normal distributions," *Annals of the Institute of Statistical Mathematics* **53** (2001), pp. 567–575.

[24] A. Müller and M. Scarsini "Some remarks on the supermodular order," *Journal of Multivariate Analysis* **73** (2000), pp. 107–119.

[25] I. Norros, "A storage model with self-similar input," *Queueing Systems* **16** (1994), pp. 387–396.

[26] K. Park and W. Willinger (Eds.), *Self–Similar Network Traffic and Performance Evaluation*, John Wiley & Sons, New York (NY), 2000.

[27] M. Parulekar, *Buffer Engineering for $M|G|\infty$ Input Processes*, Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Maryland, College Park (MD), 1999.

[28] M. Parulekar and A.M. Makowski, "Tail probabilities for a multiplexer with self–similar traffic," in *Proceedings of IEEE INFOCOM '96*, San Francisco (CA), April 1996, pp. 1452–1459.

[29] M. Parulekar and A.M. Makowski, "Tail probabilities for a multiplexer driven by $M|G|\infty$ input processes (I): Preliminary Asymptotics," *Queueing Systems – Theory & Applications* **27** (1997), pp. 271–296.

[30] M. Parulekar and A.M. Makowski, "$M|G|\infty$ input processes: A versatile class of models for traffic network," in *Proceedings of IEEE INFOCOM '97*, Kobe (Japan), April 1997, pp. 1452–1459.

[31] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Transactions on Networking* **3** (1993), pp. 226–244.

[32] S. Resnick and G. Samorodnitsky, "Fluid queues, on/off processes, and teletraffic modeling with highly variable and correlated inputs," in *Self–Similar Network Traffic and Performance Evaluation*, Eds. K. Park and W. Willinger, John Wiley & Sons, New York (NY) (2000), pp. 171–192.

[33] G. Samorodnitsky and M.S. Taqqu, *Stable Non-Guassian Random Processes: Stochastic Models With Infinite Variance*, Chapman and Hall, New York (NY), 1994.

[34] M. Shaked and J.G. Shanthikumar, "Supermodular stochastic orders and positive dependence of random vectors," *Journal of Multivarate Analysis* **61** (1997), pp. 86–101.

[35] M. Shaked and J.G. Shanthikumar, *Stochastic Orders and Their Applications*, Academic Press, New York (NY), 1994.

[36] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, English Translation (D.J. Daley, Editor), J. Wiley & Sons, New York (NY), 1984.

[37] H. Toyoizumi and J.G. Shanthikumar and R.W. Wolff, "Two extremal autocorrelated arrival processes," *Probability in the Engineering and Informational Sciences* **11** (1997), pp. 441–450.

[38] S. Vanichpun, *The Effects of Positive Correlations On Buffer Occupancy: Comparisons and Lower-Bounds Via Supermodular Ordering*, M.S. Thesis, Department of Electrical and Computer Engineering, University of Maryland, College Park (MD), December 2001.

[39] W. Willinger, M.S. Taqqu, W.E. Leland and D.V. Wilson, "Self–similarity in high–speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Statistical Science* **10** (1995), pp. 67–85.