TECHNICAL RESEARCH REPORT

Heavy Traffic Limits Associated with M|GI|Ä Input Processes

by K.P. Tsoukatos, A.M. Makowski

CSHCN T.R. 97-22
(ISR T.R. 97-58)

Heavy traffic limits associated with $M|GI|\infty$ input processes

KONSTANTINOS P. TSOUKATOS †
ktsouk@eng.umd.edu
(301) 405-0110

ARMAND M. MAKOWSKI ‡
armand@eng.umd.edu
(301) 405-6648

Abstract

We study the heavy traffic regime of a discrete–time queue driven by correlated inputs, namely the $M|GI|\infty$ input processes of Cox. We distinguish between $M|GI|\infty$ processes with short– and long–range dependence, identifying for each case the appropriate heavy traffic scaling that results in non–degenerate limits. As expected, the limits we obtain for short–range dependent inputs involve the standard Brownian motion. Of particular interest are the conclusions for the long–range dependent case: The normalized queue length can be expressed as a function not of a fractional Brownian motion, but of an α–stable, $1/\alpha$ self–similar independent increments Lévy process. The resulting buffer asymptotics in heavy traffic display a hyperbolic decay, of power $1-\alpha$. Thus, $M|GI|\infty$ processes already demonstrate that, within long–range dependence, fractional Brownian motion does not necessarily assume the ubiquitous role that standard Brownian motion plays in the short–range dependence setup.

1 Introduction

The apparent presence of long–range dependence and self–similarity in network traffic has been suggested by several traffic measurement studies (e.g., WAN [24], Ether-

†Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742. The work of this author was supported through NSF Grant NSFD CDR-88-03012 and the Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0002.

‡Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742. The work of this author was supported partially through NSF Grant NSFD CDR-88-03012, NASA Grant NAGW77S and the Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0002.

net [17] and VBR video [3]), and points to the need of revisiting various performance analysis and design issues. Indeed, as some recent experimental work [10] already suggests, long–range dependence has a tangible and adverse effect on queueing measures such as buffer overflow probabilities, and its presence thus cannot be overlooked or underestimated.

Roughly speaking, long–range dependence amounts to correlations in the packet stream spanning multiple time scales, which are individually rather small but 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 models. In order to gain some understanding into long–range dependence, and its impact on queueing, various traffic models with non–summable correlation patterns have been proposed in the literature. These models include, among others, fractional Gaussian noise inputs [1], fractional Brownian motion [21] and on/off sources with subexponential inter-renewal periods [7, 14] . In all these cases, the corresponding buffer asymptotics do not display the *exponential* tails typically associated with short–range Markovian models.

In this paper, we consider the class of discrete–time $M|GI|\infty$ input processes. An $M|GI|\infty$ input process is understood as the busy server process of a discrete–time infinite server system fed by a discrete–time Poisson process of rate λ (customers/slot) and with generic service time σ. Such $M|GI|\infty$ processes can account both for short and long–range dependent behaviors, with the correlation patterns controlled through σ [Proposition 2.1]. Furthermore, *asymptotic self–similarity* arises when σ is Pareto–like, i.e., has a *regularly varying* tail of the form (4.1). $M|GI|\infty$ processes have already been used by Paxson and Floyd to successfully model WAN traffic [24]. However, the relevance of $M|GI|\infty$ input processes to network traffic modelling is perhaps best explained through the fact they naturally arise as the aggregate limit of a large number of on–off sources [18].

As discussed in [9, 19, 22, 23], $M|GI|\infty$ processes induce a wide variety of asymptotic behaviors for the buffer probabilities at a multiplexer with constant release rate. In particular, when σ has a regularly varying tail – the $M|GI|\infty$ process is now asymptotically self–similar, the buffer asymptotics are hyperbolic in nature, in stark contrast with the Weibullian tails induced by fractional Gaussian noise (or fractional Brownian motion) [21]. Here, we further explore this discrepancy in the heavy traffic regime [12]. In fact, the motivation for going to heavy traffic is manifold:

First, under short–range dependence heavy traffic analysis has offered useful characterization of queueing networks in terms of functionals on Brownian motion [12], and it is therefore of theoretical interest to extend the analysis to the long–range dependence setup. Such an extension would possibly help answer the question as to whether fractional Brownian motion, the long–range dependent analog of standard Brownian motion, does play a similar central role in the modeling of long–range dependent traffic.

Next, heavy traffic information constitutes a key component of the *light traffic interpolation* technique originally developed by Simon and Reiman in a series of papers [25, 26, 27] in order to estimate performance metrics such as response times in queueing systems driven by Poisson–like inputs. As we contemplate the possible use of this technique with $M|GI|\infty$ input streams, we are naturally led to seek a complete classification of the heavy traffic limits which arise when considering such inputs.

Finally, a heavy traffic analysis of $M|GI|\infty$ processes might help elucidate the noted difference in buffer asymptotics between $M|GI|\infty$ and fractional Gaussian noise inputs. Unless this is due to some fundamental structural property, both models are expected to have a heavy traffic characterization in terms of fractional Brownian motion, in very much the same manner that different short–range dependent models eventually collapse to a *single* description involving Brownian motion [12]. Otherwise, despite asymptotically identical correlation patterns, the differences would carry over and manifest themselves even more clearly in the heavy traffic regime.

The results presented here confirm the latter possibility: Under short–range dependence, the class of $M|GI|\infty$ inputs is found to belong to the domain of attraction of the standard Brownian motion, as expected. More significantly, we show that under long–range dependence, with σ belonging to the domain of attraction of a non–normal stable law, the $M|GI|\infty$ process is *not* attracted to a fractional Brownian motion, but instead to a *non*–Gaussian, α–stable *Lévy motion* which is $1/\alpha$ self–similar. As a consequence, the distribution of the heavy traffic queue length does *not* display a Weibullian, but a Pareto tail, with power $1-\alpha$ [Theorem 5.3]. These results underscore the fundamentally different nature of the long–range dependent $M|GI|\infty$ process (when compared to fractional Gaussian noise), and also point to the fact that fractional Brownian motion does not necessarily play for long–

range dependence the same key role that standard Brownian motion assumes under short–range dependence. Within long–range dependence, there seems to be a choice for distinct modeling possibilities, and it is not at all difficult to find rather simple, potentially useful traffic models that are attracted to non–Gaussian limits.

The basic idea behind the proof of these results is a "convergence together" argument which allows us to identify processes with well–known heavy traffic behavior, under both short– and long–range dependence. This is accomplished chiefly by combining standard results on stable random variables and their domain of attraction [11], with a general functional convergence result for processes with stationary independent increments due to Skorokhod [29]. We point out that, even in the short–range dependent case, convergence to Brownian motion does not appear to follow from standard results for stationary processes [4, Thm. 20.1, p. 174], as it is not obvious that the $M|GI|\infty$ busy server process satisfies the required mixing property. Note however that, as shown in [22], the $M|GI|\infty$ busy server process is strongly positively correlated – it is an *associated* process. Because of this property, it is then possible under short–range dependence to develop an alternative approach [32], similar to that used by Newman and Wright in [20] in establishing the Invariance Principle for associated random variables.

Related work on heavy traffic queueing analysis under long–range dependence appears to have been initiated by Norros [21], where the presence of fractional Brownian motion is postulated. This line of inquiry is further pursued in [30], while in [7] Brichet et al. show how fractional Brownian motion can arise from a Gaussian superposition scheme of infinitely many on/off sources with heavy tailed on/off periods. In view of the fact that $M|GI|\infty$ processes arise from a *different* superposition scheme of infinitely many on/off sources [18], it is not too surprising that these lead to a different heavy traffic limit, involving Lévy motions. Heavy traffic results similar and related to the ones given here have been reported in [15], where only convergence of finite dimensional distributions is announced. The conclusions discussed here were obtained independently, and were summarized in the conference paper [31].

The remainder of the paper is organized as follows: The class of $M|GI|\infty$ input processes, along with the discrete–time queueing setup, are introduced in Section 2. We explain how the queue is driven to heavy traffic in Section 3. The main heavy traffic results are then stated in Section 4, while Section 5 discusses their consequences on the queue length asymptotics. An outline of the proofs is presented in Section 6. The proofs of the main results follow in Section 7, while the arguments behind the "convergence together" are discussed in Section 8. In the appendix Sections 9 and 10, we have summarized several technical facts concerning functions of regular variation.

A few words about the notation used here. All rvs are defined on some probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, with $\mathbf{E}$ denoting the corresponding expectation operator. We use $\Longrightarrow_r$ to denote weak convergence [4], and $\xrightarrow{P}_r$ to denote convergence in probability (with r going to infinity). We write $f(x) \sim g(x)$ $(x \to \infty)$ when $\lim_{x\to\infty} f(x)/g(x) = 1$. Equality in distribution is denoted by $=_{st}$, and inequality in the strong stochastic order is denoted by $\leq_{st}$.

2 A multiplexer driven by $M|GI|\infty$ input processes

We begin by collecting some pertinent facts concerning the busy server process of a discrete–time $M|GI|\infty$ system; the reader is referred to [8, 22] for proofs and additional material on this class of processes.

Consider a system with infinitely many servers: During time slot $[n, n+1)$, $n = 0, 1, \ldots$, β_{n+1} new customers arrive into the system. Customer j, $j = 1, \ldots, \beta_{n+1}$, is presented to its own server and begins service by the start of slot $[n+1, n+2)$; its service time has duration $\sigma_{n+1,j}$ (in number of slots). Let b_n denote the number of busy servers, or equivalently, the number of customers still present in the system at the beginning of time slot $[n, n+1)$. If initially (i.e., at time $n = 0$) there were already b busy servers, we denote by $\widehat{\sigma}_j$ the residual service duration (in time slots) for the j^{th} busy server, $j = 1, \ldots, b$. The busy server process $\{b_n,\ n = 0, 1, \ldots\}$ is what we refer to as the $M|GI|\infty$ input process.

Throughout, the $I\!N$–valued rvs b, $\{\beta_{n+1},\ n = 0, 1, \ldots\}$, $\{\sigma_{n,j},\ n = 1, 2, \ldots;\ j = 1, 2, \ldots\}$ and $\{\widehat{\sigma}_j,\ j = 1, 2, \ldots\}$ satisfy the following assumptions: (i) These rvs are mutually independent; (ii) The rvs $\{\beta_{n+1},\ n = 0, 1, \ldots\}$ are *i.i.d.* Poisson rvs with parameter $\lambda > 0$; (iii) The rvs $\{\sigma_{n,j},\ n = 1, \ldots;\ j = 1, 2, \ldots\}$ are *i.i.d.* with common pmf G on $\{1, 2, \ldots\}$. Let σ be a generic $I\!N$–valued rv distributed according to the pmf G, assume throughout that $\mathbf{E}[\sigma] < \infty$; (iv) The rvs $\{\widehat{\sigma}_j,\ j = 1, 2, \ldots\}$ are *i.i.d.* $I\!N$–valued rvs distributed according to the *equilibrium* pmf $\widehat{G}$ associated with G, i.e., if $\widehat{\sigma}$ denotes a generic $I\!N$–valued rv distributed according to the pmf $\widehat{G}$, then

$$\mathbf{P}[\widehat{\sigma} = n] = \frac{\mathbf{P}[\sigma \geq n]}{\mathbf{E}[\sigma]}, \quad n = 1, 2, \ldots \tag{2.1}$$

The tail of the equilibrium pmf $\widehat{G}$ directly controls the correlations of the sequence $\{b_n,\ n = 0, 1, \ldots\}$.

Proposition 2.1 *If b is taken to be a Poisson rv with parameter $\lambda\mathbf{E}[\sigma]$, then the busy server process $\{b_n,\ n = 0, 1, \ldots\}$ is a (strictly) stationary ergodic process with the following properties:*

1. *For each $n = 0, 1, \ldots$, the rv b_n is a Poisson rv with parameter $\lambda\mathbf{E}[\sigma]$;*
2. *Its covariance function is given by*

$$\mathrm{cov}(b_{n+j}, b_n) = \lambda\mathbf{E}\left[[\sigma - j]^+\right] = \lambda\mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > j], \quad n, j = 0, 1, \ldots$$

3. *Its index of dispersion of counts (IDC) is given by*

$$\mathrm{IDC} \equiv \sum_{j=0}^{\infty} \mathrm{cov}(b_{n+j}, b_n) = \lambda\mathbf{E}[\sigma] \sum_{j=0}^{\infty} \mathbf{P}[\widehat{\sigma} > j] = \frac{\lambda}{2}\mathbf{E}[\sigma(\sigma+1)],$$

and the process is short–range dependent (i.e., IDC finite) if and only if $\mathbf{E}[\sigma^2]$ is finite.

In short, a stationary $M|GI|\infty$ input process is fully characterized by the pair (λ, σ), and displays time dependencies which are determined by the tail of σ. We now offer such a stationary $M|GI|\infty$ input process $\{b_n,\ n = 1, 2, \ldots,\}$ to a multiplexer which we model as a discrete–time single server queue with infinite buffer capacity, operating at a constant rate and in a first–come first–served manner. Let q_n denote the number of cells remaining in the buffer by the end of slot $[n-1, n)$, and let b_{n+1} denote the number of new cells which arrive at the start of time slot $[n, n+1)$. If the multiplexer output link can transmit c cells/slot, then the buffer content sequence $\{q_n,\ n = 0, 1, \ldots\}$ evolves according to the Lindley recursion

$$q_0 = 0; \quad q_{n+1} = [q_n + b_{n+1} - c]^+, \quad n = 0, 1, \ldots \tag{2.2}$$

By Part 1 of Proposition 2.1, the average input rate to the multiplexer is simply $\mathbf{E}[b_n] = \lambda\mathbf{E}[\sigma]$, and it can be shown that the system is *stable* if $\lambda\mathbf{E}[\sigma] < c$, in which case $q_n \Longrightarrow_n q_\infty$ for some $I\!N$–valued rv q_∞.

The output to the Lindley recursion (2.2) admits an equivalent representation, which is useful for establishing heavy traffic limit theorems: We can write

$$q_0 = 0; \quad q_n = s_n - nc - \inf(s_j - jc,\ j = 0, 1, \ldots, n), \quad n = 1, 2, \ldots$$

upon defining the partial sums $\{s_n,\ n = 0, 1, \ldots\}$ by

$$s_0 \equiv 0; \quad s_n \equiv \sum_{j=1}^{n} b_j, \quad n = 1, 2, \ldots \tag{2.3}$$

3 The heavy traffic regime

We seek to understand the behavior of the (stable) queue under the assumption that it is almost fully utilized, i.e., $\lambda \mathbf{E}[\sigma]$, though less than the release rate c, is very close to c. This typically involves obtaining limiting expressions of properly rescaled quantities of interest, as the traffic intensity $\rho \equiv \lambda \mathbf{E}[\sigma]/c$ tends towards its critical value 1. Here, the quantity of interest is the steady–state queue size q_∞. A natural setup to investigate this problem consists of embedding the discrete–time queue with release rate c driven by an $M|GI|\infty$ input process (λ, σ) into a parametric family of like queueing systems, indexed by an integer parameter, say r. More precisely, for each $r = 1, 2, \ldots$ we take the r^{th} system to be a discrete–time queue with release rate c driven by an $M|GI|\infty$ input process $\{b_n^r,\ n = 0, 1, \ldots\}$ characterized by the pair (λ_r, σ). The corresponding queue size sequence $\{q_n^r,\ n = 0, 1, \ldots\}$ also obeys the Lindley recursion (2.2), and can be represented as

$$q_n^r = s_n^r - nc - \inf\left(s_j^r - jc,\ j = 0, 1, \ldots, n\right), \quad n = 1, 2, \ldots \tag{3.1}$$

where $\{s_n^r,\ n = 1, 2, \ldots\}$ is the sequence of partial sums (2.3) associated with $\{b_n^r,\ n = 1, 2, \ldots\}$. We take $\lambda_r \mathbf{E}[\sigma] < c$ for all $r = 1, 2, \ldots$, with

$$\lim_{r\to\infty} \lambda_r = c/\mathbf{E}[\sigma]. \tag{3.2}$$

Thus, each one of these systems is stable with $\lim_{r\to\infty} \rho_r = 1$, thereby capturing the notion that "the system is driven to heavy traffic." We seek a scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ such that the convergence in distribution

$$\frac{q_\infty^r}{\zeta_r} \Longrightarrow_r Q \tag{3.3}$$

takes place to some $\mathbb{R}$–valued rv Q.

Unfortunately, this heavy traffic program cannot be carried out in this form as exact expressions are *unavailable* for the distribution of q_∞^r owing to the correlations present in the $M|GI|\infty$ input process, and we need to resort to the following indirect approach where the buffer content is rescaled in both the *time* and *state space* variables. Thus, for each $r = 1, 2, \ldots$, we define the $\mathbb{R}$–valued continuous–time processes $\{S^r(t),\ t \geq 0\}$ and $\{Q^r(t),\ t \geq 0\}$ by

$$S^r(t) \equiv \frac{1}{\zeta_r}\left(s_{[rt]}^r - \mathbf{E}\left[s_{[rt]}^r\right]\right) \quad \text{and} \quad Q^r(t) \equiv \frac{q_{[rt]}^r}{\zeta_r}, \quad t \geq 0,$$

and we introduce the function $\gamma^r : \mathbb{R}_+ \to \mathbb{R}$ given by

$$\gamma^r(t) \equiv \frac{1}{\zeta_r}\left([rt]c - \mathbf{E}\left[s^r_{[tr]}\right]\right) = \frac{[rt]}{\zeta_r}\left(c - \lambda_r \mathbf{E}\left[\sigma\right]\right), \quad t \geq 0.$$

We note that (3.3) can informally be stated as

$$\lim_{r\to\infty}\lim_{t\to\infty} Q^r(t) = Q \tag{3.4}$$

with limits understood in the sense of weak convergence. The approach to heavy traffic followed here is to interchange the order of these limits, i.e., to evaluate

$$\lim_{t\to\infty}\lim_{r\to\infty} Q^r(t) \tag{3.5}$$

which corresponds to first taking $r \to \infty$, and then letting $t \to \infty$. Assuming that the limits can be taken in that order, we are then left with the task of showing that

$$\lim_{r\to\infty}\lim_{t\to\infty} Q^r(t) = Q = \lim_{t\to\infty}\lim_{r\to\infty} Q^r(t). \tag{3.6}$$

In this paper we concentrate only on establishing the first step (3.5), and it is well known [12, 33] that the theory weak convergence on function spaces provides a natural framework for doing so. To that end, we pause briefly to introduce the needed notation, as well as to highlight several points from the theory of weak convergence of processes; this material is drawn mostly from [4, pp. 150–153] to which the reader is referred for additional information:

For each $T > 0$, let $D[0,T]$ denote the space of mappings $[0,T] \to \mathbb{R}$ which are right–continuous with left limits; the space $D[0,T]$ can be equipped with either the uniform topology or the standard Skorokhod topology [4, p. 111]. As in [4, p. 150], a concept prefixed with U (resp. S) refers to the uniform (resp. Skorokhod) topology. For probability measures defined on the collection of U–Borel (resp. S–Borel) sets on $D[0,T]$, we refer to weak convergence in the sense of the uniform (resp. Skorokhod) topology by U–weak (resp. S–weak) convergence, and we write $\overset{U}{\Longrightarrow}_r$ (resp. $\overset{S}{\Longrightarrow}_r$) (with the understanding that r goes to infinity). For probability measures defined on the collection of U–Borel sets, U–weak convergence implies S–weak convergence but the converse is false. This implication will be used repeatedly in various technical arguments [Sections 6 and 8].

Finally, let $D[0,\infty)$ denote the space of mappings $\mathbb{R}_+ \to \mathbb{R}$ which are right–continuous with left limits. In this paper, we present results on the S–weak convergence of the restrictions to finite intervals of sequences of $\mathbb{R}$–valued processes with sample paths in $D[0,\infty)$. More precisely, consider the sequence of $\mathbb{R}$–valued processes $\{X_r(t),\ t \geq 0\}$, $r = 1, 2, \ldots$, with sample paths in $D[0,\infty)$. Whenever, for *each* $T > 0$, we have the S–weak convergence

$$\{X_r(t),\ 0 \leq t \leq T\} \stackrel{S}{\Longrightarrow}_r \{X(t),\ 0 \leq t \leq T\} \quad \text{in } D[0,T] \tag{3.7}$$

for some $\mathbb{R}$–valued process $\{X(t),\ t \geq 0\}$ with sample paths in $D[0,\infty)$, we simplify the notation by writing

$$\{X_r(t),\ t \geq 0\} \Longrightarrow_r \{X(t),\ t \geq 0\}. \tag{3.8}$$

Now, noting that (3.1) can be rewritten as

$$Q^r(t) = S^r(t) - \gamma^r(t) - \inf_{0 \leq x \leq t} \left(S^r(x) - \gamma^r(x)\right), \quad t \geq 0, \tag{3.9}$$

and recalling the continuous mapping theorem [4, Thm. 5.1, p. 30], we conclude that the first limit in (3.5) requires at the very least identifying a scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ that ensures the convergence

$$\{S^r(t),\ t \geq 0\} \Longrightarrow_r \{S(t),\ t \geq 0\} \tag{3.10}$$

for some non–trivial limiting process $\{S(t),\ t \geq 0\}$.

4 The main heavy traffic results

As will become apparent shortly, the choice of the scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ and the characterization of the limiting process $\{S(t),\ t \geq 0\}$ entering (3.10) both depend on the distribution of the rv σ which controls the correlations in the input cell stream. It is nevertheless easy to see that in order to avoid collecting only a law of large numbers, any candidate scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ should obey the following necessary condition:

Condition (A) *The scaling sequence* $\{\zeta_r,\ r = 1, 2, \ldots\}$ *satisfies*

$$\lim_{r \to \infty} \zeta_r = +\infty \quad \textit{with} \quad \lim_{r \to \infty} \frac{\zeta_r}{r} = 0.$$

The heavy traffic assumption below refines (3.2), and guarantees that, as $r \to \infty$, the family of queueing systems described by (3.9) gradually approaches instability at the appropriate speed:

Assumption (A) *The scaling sequence* $\{\zeta_r,\ r=1,2,\ldots\}$ *satisfies*

$$\lim_{r\to\infty}(\lambda_r \mathbf{E}[\sigma]-c)\,\frac{r}{\zeta_r}=-\gamma \quad \textit{or equivalenty,} \quad \lambda_r\mathbf{E}[\sigma]=c-\frac{\zeta_r}{r}(\gamma+o(1))$$

for some $\gamma>0$.

Condition (A) and Assumption (A) are enforced throughout. It is worth pointing out that the scaling sequence $\{\zeta_r,\ r=1,2,\ldots\}$ is *essentially unique*, i.e., any other scaling sequence $\{\zeta'_r,\ r=1,2,\ldots\}$ yielding a non–degenerate limit in (3.10) must satisfy $\lim_{r\to\infty}\zeta'_r/\zeta_r=C$ for some finite constant $C>0$.

We begin with the case where the $M|GI|\infty$ process is short–range dependent and let $\{B(t),\ t\geq 0\}$ denote a standard Brownian motion.

Theorem 4.1 *(Short-range dependence) If* $\mathbf{E}[\sigma^2]<\infty$, *then with* $\zeta_r=\sqrt{r}$, $r=1,2,\ldots$, *it holds that*

$$\{S^r(t),\ t\geq 0\}\Longrightarrow_r \{\sqrt{\frac{c\mathbf{E}[\sigma^2]}{\mathbf{E}[\sigma]}}\,B(t),\ t\geq 0\}.$$

The remaining results are obtained under the additional assumption that the tail of σ is *regularly varying* of order α $(1<\alpha\leq 2)$, i.e., of the form

$$\mathbf{P}[\sigma>n]=n^{-\alpha}h(n), \quad n=0,1,\ldots \tag{4.1}$$

for some slowly varying function $h:\mathbb{R}_+\to\mathbb{R}_+$ [11], in which case the mean $\mathbf{E}[\sigma]$ is finite. Of particular interest for the forthcoming discussion is the realization that the truncated second moment of σ is $(2-\alpha)$–regularly varying. Writing

$$l_\alpha(x)\equiv\begin{cases}\dfrac{\alpha}{2-\alpha}\,h(x) & \text{if } 1<\alpha<2\\[2ex] 2\displaystyle\sum_{r=1}^{[x]}\frac{h(r)}{r} & \text{if } \alpha=2\end{cases} \tag{4.2}$$

for all $x>0$, we can show via Proposition 9.2 that the function $l_\alpha:\mathbb{R}_+\to\mathbb{R}_+$ is slowly varying and, whenever $\mathbf{E}[\sigma^2]=\infty$,

$$\mathbf{E}\left[\mathbf{1}[\sigma\leq n]\,\sigma^2\right]\sim n^{2-\alpha}l_\alpha(n), \quad (n\to\infty). \tag{4.3}$$

The details of the proof of this asymptotic equivalence are identical to those of (7.10) and (7.11).

The next proposition handles the boundary value $\alpha=2$, which represents a hybrid case between short and long–range dependence.

Theorem 4.2 *Assume $\alpha = 2$ in (4.1) with $\mathbf{E}[\sigma^2] = \infty$. Then, with $\{\zeta_r,\ r = 1,2,\ldots\}$ satisfying*

$$\lim_{r\to\infty} \frac{r}{\zeta_r^2} l_2(\zeta_r) = \lim_{r\to\infty} \frac{r}{\zeta_r^2} \mathbf{E}\left[\mathbf{1}[\sigma \le \zeta_r]\,\sigma^2\right] = K \tag{4.4}$$

for some positive constant K, it holds that

$$\{S^r(t),\ t \ge 0\} \Longrightarrow_r \{\sqrt{\frac{cK}{\mathbf{E}[\sigma]}}\ B(t),\ t \ge 0\}.$$

Finally, we turn to the case of *bona fide* long–range dependence, i.e., $1 < \alpha < 2$. Recall that a Lévy process is an $\mathbb{R}$–valued process with *stationary independent increments.* We let $\{L_\alpha(t),\ t \ge 0\}$ denote a standard α–stable, spectrally positive Lévy motion, i.e., a Lévy process such that $L_\alpha(0) = 0$ a.s. and for $t > 0$ the rv $L_\alpha(t)$ is a stable rv $S_\alpha(t^{1/\alpha}, 1, 0)$ [28] characterized by

$$\mathbf{E}[\exp(i\theta L_\alpha(t))] = \exp\left(-t|\theta|^\alpha\left(1 - i\,\mathrm{sgn}(\theta)\tan(\frac{\pi\alpha}{2})\right)\right), \quad \theta \in \mathbb{R}. \tag{4.5}$$

Theorem 4.3 *(Long–range dependence) If $1 < \alpha < 2$ in (4.1), then with $\{\zeta_r, r = 1,2,\ldots\}$ satisfying*

$$\lim_{r\to\infty} \frac{r}{\zeta_r^\alpha} h(\zeta_r) = \lim_{r\to\infty} r\mathbf{P}[\sigma > \zeta_r] = K \tag{4.6}$$

for some positive constant K, it holds that

$$\{S^r(t),\ t \ge 0\} \Longrightarrow_r \{C(K)\ L_\alpha(t),\ t \ge 0\} \tag{4.7}$$

where

$$C(K) \equiv \left(\frac{cK\Gamma(2-\alpha)}{(\alpha-1)\mathbf{E}[\sigma]}\ \cos(\pi\frac{2-\alpha}{2})\right)^{1/\alpha}. \tag{4.8}$$

We close with a characterization of the scaling sequences encountered in Theorems 4.2 and 4.3; its proof is available in Proposition 10.3 of Section 10.

Proposition 4.4 *The scaling sequence $\{\zeta_r,\ r = 1,2,\ldots\}$ of Theorems 4.2 and 4.3 is $1/\alpha$–regularly varying, $1 < \alpha \le 2$, i.e., of the form $\zeta_r = r^{1/\alpha}\widehat{h}(r)$ for some slowly varying function $\widehat{h} : \mathbb{R}_+ \to \mathbb{R}_+$.*

5 Consequences and comments

Several interesting inferences follow from the heavy traffic results obtained so far.

5.1 Queue size

We start with the heavy traffic behavior of the normalized queue length. Whenever the convergence (3.10) holds, we can immediately conclude from (3.9) and from the continuity of the reflection mapping (via the continuous mapping theorem [4, Thm. 5.1, p. 30]) that

$$\{Q^r(t),\ t \geq 0\} \Longrightarrow_r \{Q(t),\ t \geq 0\} \tag{5.1}$$

with

$$Q(t) \equiv S(t) - \gamma t - \inf_{0 \leq x \leq t} (S(x) - \gamma x), \quad t \geq 0. \tag{5.2}$$

The form of the limit derives from (3.9) and the fact that $\lim_{r \to \infty} \gamma^r(t) = -\gamma t$ under Assumption (A).

This observation can now be used to provide a characterization of the steady–state buffer content in heavy traffic under the assumptions of Theorems 4.1–4.3.

In the short–range dependent case, Theorem 4.1 combines with a classical result on the supremum functional of Brownian motion [12, p.15] to yield the following.

Theorem 5.1 *Under the assumptions of Theorem 4.1, the resulting stationary heavy–traffic buffer content is exponentially distributed with*

$$\lim_{t \to \infty} \mathbf{P}\left[Q(t) > b\right] = \exp\left(-\frac{2\gamma \mathbf{E}\left[\sigma\right]}{c\mathbf{E}\left[\sigma^2\right]}\, b\right), \quad b \geq 0.$$

Theorem 4.2 leads via (5.1)–(5.2) to a similar result.

Theorem 5.2 *Under the assumptions of Theorem 4.2, the resulting stationary heavy–traffic buffer content is exponentially distributed, with*

$$\lim_{t \to \infty} \mathbf{P}\left[Q(t) > b\right] = \exp\left(-\frac{2\gamma \mathbf{E}\left[\sigma\right]}{cK}\, b\right), \quad b \geq 0.$$

Finally, in the stable case we combine Theorem 4.3 with the developments in [5, Theorem 12a] to get the following fact.

Theorem 5.3 *Under the assumptions of Theorem 4.3, the distribution of the resulting stationary heavy–traffic buffer content has a Pareto tail given by*

$$\lim_{t \to \infty} \mathbf{P}\left[Q(t) > b\right] \sim \frac{cK}{\gamma(\alpha - 1)\mathbf{E}\left[\sigma\right]}\, b^{1-\alpha} \quad (b \to \infty). \tag{5.3}$$

5.2 On selecting the heavy traffic scaling

As the appropriate scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ is revealing of the nature of the limiting heavy traffic process $\{S(t),\ t \geq 0\}$, we briefly discuss here its selection.

It is known [16] that convergence of a normalized partial sum process, such as $\{S^r(t),\ t \geq 0\}$, can only be to a self–similar process, and that the corresponding Hurst parameter H may be determined through the regularly varying scaling $\{\zeta_r,\ r = 1, 2, \ldots\}$ by

$$\lim_{r\to\infty} \frac{\zeta_{[rx]}}{\zeta_r} = x^H, \quad x > 0.$$

On the other hand, under (4.1) with $1 < \alpha < 2$, the $M|GI|\infty$ busy server process process already possesses the so-called second order asymptotic self–similarity property, with parameter $(3-\alpha)/2$ [18], i.e., by aggregating the original process $\{b_n, n = 0, 1, \ldots\}$ in blocks of size m and dividing by the block size, we can obtain in the limit as m goes to infinity the same correlation function as that of a fractional Brownian motion. Because of this fact it would be tempting to think that the appropriate scaling ensuring (3.10) might be the one that balances the rate of growth of the partial sums variance, so that convergence occurs to a limiting process with finite variance. By standard calculations, we find the partial sums variance to be

$$\mathrm{var}\left[S^r(t)\right] = \frac{\lambda_r \mathbf{E}\left[\sigma\right]}{\zeta_r^2}\left([rt] + 2\sum_{k=1}^{[rt]}([rt]-k)\mathbf{P}\left[\widehat{\sigma} > k\right]\right), \quad t \geq 0 \tag{5.4}$$

for all $r = 1, 2, \ldots$. As shown in [32], when the tail of σ satisfies (4.1) with $1 < \alpha < 2$, the candidate scaling $\{\zeta_r,\ r = 1, 2, \ldots\}$ given by

$$\zeta_r^2 \equiv r\sum_{k=1}^{r} \mathbf{P}\left[\widehat{\sigma} > k\right], \quad r = 1, 2, \ldots \tag{5.5}$$

indeed results in a finite limiting variance, i.e., $\lim_{r\to\infty} \mathrm{var}\left[S^r(t)\right]$ exists and is finite for all $t \geq 0$. In addition, invoking (9.24) we see that this scaling has the asymptotic form

$$\zeta_r^2 \sim \frac{1}{(2-\alpha)(\alpha-1)\mathbf{E}\left[\sigma\right]}\, r^{3-\alpha}h(r) \quad (r \to \infty).$$

As such a scaling is clearly $(3-\alpha)/2$–regularly varying, it suggests possible convergence to a fractional Brownian motion with Hurst parameter $(3-\alpha)/2$. For a single $M|GI|\infty$ process however this is not true as such convergence to a fractional Brownian motion does *not* take place. In fact, the candidate scaling (5.5), which balances the growth of the variance, is *too strong* and yields convergence to a *degenerate limit* – the identically zero process. Theorem 4.3, in conjuction with Proposition 4.4, clearly shows that the correct scaling does not contain any $r^{(3-\alpha)/2}$ factor; but instead contains the weaker $r^{1/\alpha}$ factor associated with the stable law to which the service rv σ is attracted. As a result, the limiting heavy traffic process turns out to be not a fractional Brownian motion but an α–stable $1/\alpha$–self–similar Lévy motion, the stable analog of standard Brownian motion, which has infinite variance and independent increments. In heavy traffic, the corresponding queue length asymptotics are not Weibullian, but hyperbolic in nature, with power $1-\alpha$. Thus, the $M|GI|\infty$ processes demonstrate that, within long–range dependence, fractional Brownian motion does not assume the ubiquitous role that its short–range dependent counterpart, standard Brownian motion, plays in the short–range dependence setup, and that modeling possibilities attracted to non–Gaussian limits are not so hard to find. Clearly, the extent to which such non–Gaussian processes can serve as useful traffic models deserves some further consideration.

6 Outline of proof and preliminary results

In this section we organize the proof of Theorems 4.1–4.3 into a series of steps which we formalize as Propositions; their proofs are given in Section 8.

Look at the r^{th} queueing system, $r = 1, 2, \ldots$, and fix $n = 0, 1, \ldots$. We note the decomposition $b_n^r = b_n^{(0)r} + b_n^{(a)r}$ where the rvs $b_n^{(0)r}$ and $b_n^{(a)r}$ describe the contributions to the number of customers in the system at the beginning of slot $[n, n+1)$ from those initially present (at $n = 0$) and from the new arrivals, respectively. It is easy to see that

$$b_n^{(0)r} = \sum_{j=1}^{b^r} \mathbf{1}\left[\widehat{\sigma}_j > n\right] \quad \text{and} \quad b_n^{(a)r} = \sum_{k=1}^{n} \sum_{j=1}^{\beta_k^r} \mathbf{1}\left[\sigma_{k,j} > n-k\right].$$

It was shown in [22] that

$$s_n^{(0)r} \equiv \sum_{j=1}^{n} b_j^{r(0)} = \sum_{j=1}^{b^r} \min(n, \widehat{\sigma}_j - 1) \tag{6.1}$$

and

$$s_n^{(a)r} \equiv \sum_{k=1}^{n} b_k^{(a)r} = \sum_{k=1}^{n} \sum_{j=1}^{\beta_k^r} \min(\sigma_{k,j}, n-k+1). \tag{6.2}$$

We introduce the rescaled versions

$$S^{(0)r}(t) \equiv \frac{1}{\zeta_r}\left(s^{(0)r}_{[rt]} - \mathbf{E}\left[s^{(0)r}_{[rt]}\right]\right) \quad \text{and} \quad S^{(a)r}(t) \equiv \frac{1}{\zeta_r}\left(s^{(a)r}_{[rt]} - \mathbf{E}\left[s^{(a)r}_{[rt]}\right]\right), \quad t \geq 0$$

so that

$$S^r(t) = S^{(0)r}(t) + S^{(a)r}(t), \quad t \geq 0. \tag{6.3}$$

Also, for each $T > 0$, the identically zero mapping on $[0,T]$ is the element of $D[0,T]$ denoted by θ_T, i.e., $\theta_T : [0,T] \to \mathbb{R}$ with $\theta_T(t) = 0$, $0 \leq t \leq T$.

We first show that the initial condition plays no role in the heavy traffic limit, as should be expected. This reduction step, as well as others taken in this section, is accomplished under the following sufficient condition.

Condition (B) *The scaling sequence* $\{\zeta_r,\ r = 1,2,\ldots\}$ *satisfies*

$$\lim_{r\to\infty} \frac{1}{\zeta_r} \sum_{j=1}^{r} \mathbf{P}\left[\hat{\sigma} > j\right] = 0.$$

Condition (B) holds under each set of assumptions of Theorems 4.1–4.3; this is shown in Proposition 7.1 of Section 7.

Proposition 6.1 *Under Condition (B), for each $T > 0$ it holds that*

$$\{S^{(0)r}(t),\ 0 \leq t \leq T\} \overset{U}{\Longrightarrow}_r \theta_T \quad \text{in } D[0,T].$$

Thus, in order to get (3.10) it suffices to consider the limiting behavior of the rescaled process $\{S^{(a)r}(t),\ t \geq 0\}$. To that end, for each $r = 1,2,\ldots$, we introduce the sequence $\{w^r_n,\ n = 0,1,\ldots\}$ given by

$$w^r_0 \equiv 0, \quad w^r_n \equiv \sum_{k=1}^{n} \sum_{j=1}^{\beta^r_k} \sigma_{k,j}, \quad n = 1,2,\ldots \tag{6.4}$$

which can be interpreted as the contribution to the workload due to arrivals only. The corresponding rescaled process $\{W^r(t),\ t \geq 0\}$ is given by

$$W^r(t) \equiv \frac{1}{\zeta_r}\left(w^r_{[rt]} - \mathbf{E}\left[w^r_{[rt]}\right]\right), \quad t \geq 0. \tag{6.5}$$

The main idea driving the discussion is that in as much as heavy traffic is concerned, the process $\{W^r(t),\ t \geq 0\}$ acts as a surrogate for $\{S^{(a)r}(t),\ t \geq 0\}$. This is made precise through the following "convergence together" result.

Proposition 6.2 *Under Condition (B), for each $T > 0$ it holds that*

$$\{W^r(t) - S^{(a)r}(t),\ 0 \le t \le T\} \stackrel{U}{\Longrightarrow}_r \theta_T \quad \text{in } D[0,T].$$

Combining Propositions 6.1 and 6.2, we immediately get the following conclusion from (6.3).

Corollary 6.3 *Under Condition (B), for each $T > 0$ it holds that*

$$\{W^r(t) - S^r(t),\ 0 \le t \le T\} \stackrel{U}{\Longrightarrow}_r \theta_T \quad \text{in } D[0,T],$$

so that the process $\{S^r(t),\ 0 \le t \le T\}$ is S–weakly convergent if and only if $\{W^r(t),\ 0 \le t \le T\}$ is S–weakly convergent, and convergence is to the same limit.

Thus, we need only consider the convergence of the process $\{W^r(t),\ t \ge 0\}$, and characterize the limiting process. In fact, a further reduction can be achieved by noting that in heavy traffic we can replace $\{\beta_k^r, k = 1, 2, \ldots\}$ by the limiting i.i.d. sequence $\{\beta_k, k = 1, 2, \ldots\}$, where the generic rv β is a Poisson rv with parameter $c/\mathbf{E}[\sigma]$. More precisely, consider the modified workload process $\{v_n,\ n = 0, 1, \ldots\}$ given by

$$v_0 = 0; \quad v_n = \sum_{k=1}^{n} \sum_{j=1}^{\beta_k} \sigma_{k,j}, \quad n = 1, 2, \ldots \tag{6.6}$$

under the assumption that the rvs $\{\beta_k, k = 1, 2, \ldots\}$ are independent of the service time rvs $\{\sigma_{n,j},\ n, j = 1, 2, \ldots\}$. For each $r = 1, 2, \ldots$, the corresponding rescaled process $\{V^r(t),\ t \ge 0\}$ is defined by

$$V^r(t) \equiv \frac{1}{\zeta_r}\left(v_{[rt]} - \mathbf{E}\left[v_{[rt]}\right]\right), \quad t \ge 0. \tag{6.7}$$

Proposition 6.4 *Under Assumption (A), the process $\{W^r(t),\ 0 \le t \le T\}$ is S–weakly convergent if and only if $\{V^r(t),\ 0 \le t \le T\}$ is S–weakly convergent, and convergence is to the same limit.*

Corollary 6.3 and Proposition 6.4 together lead to the following conclusion:

Corollary 6.5 *Under Assumption (A) and Condition (B), the process $\{S^r(t),\ 0 \le t \le T\}$ is S–weakly convergent if and only if $\{V^r(t),\ 0 \le t \le T\}$ is S–weakly convergent, and convergence is to the same limit.*

7 Proofs of Theorems 4.1–4.3

First, the big picture: Corollary 6.5 and Proposition 7.1 (given below) imply that in proving Theorems 4.1–4.3 we need only investigate the convergence of the modified workload process (6.7). This is a much easier task as we now deal with the (normalized) partial sums process associated with a single sequence of *i.i.d.* rvs, of finite mean but possibly infinite variance, an extensively studied situation where the (functional form of the) classical Central Limit Theorem and its generalization to i.i.d. summands with infinite variance, are expected to yield the requested convergence. In fact, as we shall see shortly, the convergence of the finite dimensional distributions of $\{V^r(t),\ t \geq 0\}$ turns out to be an easy by-product of classical results concerning stable distributions and their domains of attraction [11, pp. 574–581]. Finally, the desired S-weak convergence of the process $\{V^r(t),\ t \geq 0\}$, thus of $\{S^r(t),\ t \geq 0\}$, will be validated through the functional convergence results due to Skorokhod [29]. This approach clearly explains the form of the results obtained in this paper, providing insights as to *when* the process $\{V^r(t),\ t \geq 0\}$ is expected to converge, and to *which* limit.

And now, on with the details: In Section 9 we give a proof that the technical Condition (B) required to establish the "convergence together" argument, indeed holds under the assumptions of Theorems 4.1–4.3.

Proposition 7.1 *Condition (B) holds true for each of the scaling sequences $\{\zeta_r,\ r = 1, 2, \ldots\}$ in Theorems 4.1–4.3.*

Next, we consider the generic compound rv Y given by

$$Y \equiv \sum_{j=1}^{\beta} \sigma_j \tag{7.1}$$

where the rv β is a Poisson rv with parameter $c/\mathbf{E}[\sigma]$ and independent of the i.i.d. rvs $\{\sigma_j,\ j = 1, 2, \ldots\}$ which are distributed according to σ. Fixing $t \geq 0$, we note that

$$V^r(t) =_{st} \frac{1}{\zeta_r} \sum_{k=1}^{[rt]} (Y_k - \mathbf{E}[Y_k]), \quad r = 1, 2, \ldots \tag{7.2}$$

where the i.i.d. rvs $\{Y_k,\ k = 1, 2, \ldots\}$ are distributed according to the generic rv Y.

For easy reference, we restate some useful facts concerning stable distributions and their domains of attraction; the reader is refered to [11, pp. 574–581] for additional material: Let L be a rv with distribution not concentrated at one point, and let $\{X_r,\ r = 1, 2, \ldots\}$ be a sequence of i.i.d. rvs, with generic rv X. We say that X belongs to the domain of attraction of the rv L if there exist normalizing constants $\zeta_r > 0$ and $c_r, r = 1, 2, \ldots$, such that

$$\frac{X_1 + \ldots + X_r - rc_r}{\zeta_r} \Longrightarrow_r L. \tag{7.3}$$

By Theorem 1 of [11, p. 576] only stable rvs possess a domain of attraction. By Theorem 2 in [11, p. 577], in order for X to belong to the domain of attraction of a stable law with exponent α, $0 < \alpha \leq 2$, it is necessary that its truncated second moment be regularly varying with exponent $2 - \alpha$, i.e.,

$$\mathbf{E}\left[\mathbf{1}\left[X \leq r\right] X^2\right] \sim r^{2-\alpha} g(r) \quad (r \to \infty), \tag{7.4}$$

for some slowly varying function $g : \mathbb{R}_+ \to \mathbb{R}_+$. The associated scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ in (7.3) must satisfy

$$\lim_{r\to\infty} \frac{r}{\zeta_r^2} \mathbf{E}\left[\mathbf{1}\left[X \leq \zeta_r\right] X^2\right] = M \tag{7.5}$$

for some constant $M > 0$ [11, p. 579]. Moreover, if $\mathbf{E}\left[X\right]$ is finite, then by Theorem 3(ii) of [11, p. 581] we can take $c_r = \mathbf{E}\left[X\right]$, $r = 1, 2, \ldots$.

We are now ready to discuss Theorems 4.1–4.3 which are all proven in the same manner, although for clarity of presentation, we shall consider each of them separately. As $\mathbf{E}\left[Y\right]$ is finite under the enforced assumptions, we conclude from (7.2) and (7.3) that for each $t > 0$, the convergence question concerning $\{V^r(t),\ r = 1, 2, \ldots\}$ is equivalent to determining whether the rv Y is attracted to a stable law, and to which one. In asserting this equivalence we rely on the fact that the scaling sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ so selected is regularly varying; this turn out to be the case by Proposition 4.4, so that

$$\lim_{r\to\infty} \frac{\zeta_{[rt]}}{\zeta_r} = t^{1/\alpha}, \quad t \geq 0. \tag{7.6}$$

In each case, we show that both the necessary condition (7.4) and the accompanying sufficient condition stated in [11, p. 577] are satisfied. This occurs simply because the generic rv Y inherits the tail behavior of the generic service time σ under each set of assumptions of Theorems 4.1–4.3.

A proof of Theorem 4.1. Since $\mathbf{E}\left[\sigma^2\right] < \infty$, the variance of Y is also finite, and is given by

$$\mathrm{var}\,[Y] = \mathrm{var}\,[\beta]\,\mathbf{E}\,[\sigma]^2 + \mathbf{E}\,[\beta]\,\mathrm{var}\,[\sigma] = \frac{c\mathbf{E}\,[\sigma^2]}{\mathbf{E}\,[\sigma]}. \tag{7.7}$$

Hence, the truncated second moment of Y varies slowly, i.e., (7.4) holds with $\alpha = 2$ and as Y is never degenerate at one point, it follows from Corollary 1 to Theorem 2 in [11, p. 578] that Y is attracted to the normal distribution. Obviously, the scaling $\zeta_r = \sqrt{r}$, $r = 1, 2, \ldots$ satisfies (7.5), with $M = c\mathbf{E}\,[\sigma^2]\,/\mathbf{E}\,[\sigma]$. In fact, by a well-known result of Donsker [4, Thm. 16.1, p. 137], selecting $\zeta_r = \sqrt{r}$, $r = 1, 2, \ldots$ ensures that the process $\{V^r(t),\ t \geq 0\}$ is S-weakly convergent to a Brownian motion, with

$$\{V^r(t),\ t \geq 0\} \overset{S}{\Longrightarrow}_r \{\sqrt{M}\ B(t),\ t \geq 0\}. \tag{7.8}$$

Combining (7.8) with Proposition 7.1 and Corollary 6.5 immediately concludes the proof. ■

Under the assumptions of Theorems 4.2 and 4.3, $\mathbf{E}\,[\sigma^2]$ is infinite, and the compound Poisson rv Y now has infinite variance. Also, if σ satisfies the tail condition (4.1), so does Y with

$$\mathbf{P}\,[Y > r] = \mathbf{P}\left[\sum_{j=1}^{\beta} \sigma_j > r\right] \sim \mathbf{E}\,[\beta]\ r^{-\alpha} h(r) \quad (r \to \infty). \tag{7.9}$$

The asymptotic equality in (7.9) is stated as an exercise in [11, Ex. 31, p. 288], where the reader will find hints for its proof. Next, we check that the truncated second moment of Y is given by

$$\mathbf{E}\left[\mathbf{1}\,[Y \leq r]\,Y^2\right] = 2\sum_{n=1}^{r} n\mathbf{P}\,[Y > n] - r(r+2)\mathbf{P}\,[Y > r] + \sum_{n=0}^{r-1} \mathbf{P}\,[Y > n]$$

for each $r = 1, 2, \ldots$, and using (7.9) in this last expression we find that

$$\mathbf{E}\left[\mathbf{1}\,[Y \leq r]\,Y^2\right] \sim \mathbf{E}\,[\beta]\left(2\sum_{n=1}^{r} n^{1-\alpha} h(n) - r^{2-\alpha} h(r)\right) \quad (r \to \infty) \tag{7.10}$$

because $\mathbf{E}\,[Y]$ is finite and $\mathbf{E}\,[Y^2]$ is infinite. We close these preliminary remarks by noting that the truncated second moments of σ and Y are obviously related to each other by

$$\mathbf{E}\left[\mathbf{1}\,[Y \leq r]\,Y^2\right] \sim \mathbf{E}\,[\beta]\,\mathbf{E}\left[\mathbf{1}\,[\sigma \leq r]\,\sigma^2\right] \quad (r \to \infty). \tag{7.11}$$

A proof of Theorem 4.2. Inserting $\alpha = 2$ in (7.10) and using the definition (4.2) (with $\alpha = 2$), we get

$$\mathbf{E}\left[\mathbf{1}\left[Y \leq r\right] Y^2\right] \sim \mathbf{E}\left[\beta\right] \ (l_2(r) - h(r)) \quad (r \to \infty).$$

By Proposition 9.2(ii), $l_2 : \mathbb{R}_+ \to \mathbb{R}_+$ is slowly varying with

$$\lim_{r \to \infty} \frac{h(r)}{l_2(r)} = 0,$$

so that

$$\mathbf{E}\left[\mathbf{1}\left[Y \leq r\right] Y^2\right] \sim \mathbf{E}\left[\beta\right] \ l_2(r) \quad (r \to \infty). \tag{7.12}$$

This time, by Corollary 1 in [11, p. 578], the slow variation of the truncated second moment is a necessary and sufficient condition for Y to be attracted to the normal distribution, with normalizing coefficients selected according to (7.5) (despite the fact that the variance of Y is now infinite). Since the marginals of the process $\{V^r(t), \ t \geq 0\}$, which has stationary, independent increments, converge to a Gaussian distribution, it follows by [29, Theorem 2.7] without any additional conditions that (7.8) takes place. Because of (7.12), selecting the scaling $\{\zeta_r, \ r = 1, 2, \ldots\}$ according to (7.5), with $M = cK/\mathbf{E}\left[\sigma\right]$, is equivalent to (4.4). Combining (7.8) with Proposition 7.1 and Corollary 6.5 completes the proof. ■

A proof of Theorem 4.3. When $1 < \alpha < 2$ in (4.1) the rvs σ and Y have infinite second moment, and Proposition 9.2(i) implies

$$\lim_{r \to \infty} \frac{1}{r^{2-\alpha} h(r)} \sum_{n=1}^{r} n^{1-\alpha} h(n) = \frac{1}{(2-\alpha)}. \tag{7.13}$$

Using this asymptotic in (7.10) we get

$$\mathbf{E}\left[\mathbf{1}\left[Y \leq r\right] Y^2\right] \sim \frac{\alpha}{2-\alpha} \ \mathbf{E}\left[\beta\right] \ r^{2-\alpha} h(r) \quad (r \to \infty). \tag{7.14}$$

Invoking Corollary 2 of [11, p. 578], we see that (7.14) and the tail condition (7.9) are sufficient to ensure that Y belongs to the domain of attraction of a non–normal stable distribution with exponent $1 < \alpha < 2$. The associated scaling sequence $\{\zeta_r, \ r = 1, 2, \ldots\}$, selected according to (7.5), yields convergence of the marginal distribution of $V^r(1)$, as r goes to infinity, to that of an α–stable rv, i.e.,

$$\lim_{r \to \infty} \mathbf{E}\left[\exp(i\theta V^r(1))\right] = \mathbf{E}\left[\exp\left(i\theta \left(\frac{M\Gamma(3-\alpha)}{\alpha(\alpha-1)} \cos(\pi \frac{2-\alpha}{2})\right)^{1/\alpha} L_\alpha(1)\right)\right], \quad \theta \in \mathbb{R}.$$

The exact value of the constant given above can be easily verified, by recalling the expression (4.5) for the characteristic function of $L_\alpha(1)$ and comparing it with Eq. (3.18) of [11, p. 730] (note the unfortunate error in the $\pm$ sign). Next, appealing to [29, Theorem 2.7] again, we conclude that convergence of the marginals also implies S–weak convergence of the process $\{V^r(t),\ t \geq 0\}$, which has stationary, independent increments, to an α–stable Lévy motion. More precisely, it holds that

$$\{V^r(t),\ t \geq 0\} \stackrel{S}{\Longrightarrow}_r \{\Big(\frac{M\Gamma(3-\alpha)}{\alpha(\alpha-1)}\cos(\pi\frac{2-\alpha}{2})\Big)^{1/\alpha} L_\alpha(t),\ t \geq 0\}. \tag{7.15}$$

Using (7.14) in (7.5) with $M = cK\alpha/(2-\alpha)\mathbf{E}[\sigma]$ we obtain the scaling sequence $\{\zeta_r,\ r = 1,2,\ldots\}$ given in (4.6). Finally, combining (7.15) with Proposition 7.1 and Corollary 6.5 shows that (4.7) holds true. ■

8 Proofs of Propositions 6.1, 6.2 and 6.4

We start by remarking that if the sequence $\{\zeta_r,\ r = 1,2,\ldots\}$ is regularly varying (as stated in Proposition 4.4), then Condition (B) also implies

$$\lim_{r\to\infty} \frac{1}{\zeta_r}\sum_{j=1}^{[rt]} \mathbf{P}[\widehat{\sigma} > j] = 0, \quad t \geq 0. \tag{8.1}$$

All three proofs given in this section follow the same pattern, and are based on the following simple idea: Consider a sequence of $\mathbb{R}$–valued processes $\{X^r(t),\ t \geq 0\}$, $r = 1,2,\ldots$, with sample paths in $D[0,\infty)$. Fix $T > 0$. According to Theorem 4.1 of [4, p. 25], in order to prove the U–weak convergence

$$\{X^r(t),\ 0 \leq t \leq T\} \stackrel{U}{\Longrightarrow}_r \theta_T \quad \text{in } D[0,T], \tag{8.2}$$

it suffices to establish the convergence in probability

$$\sup_{0\leq t\leq T} |X^r(t)| \stackrel{P}{\longrightarrow}_r 0. \tag{8.3}$$

A proof of Proposition 6.1. Fix $r = 1,2,\ldots$, and note from (6.1) that

$$\sup_{0\leq t\leq T} |S^{(0)r}(t)| \leq \frac{1}{\zeta_r}\sum_{j=1}^{b^r} \min(\widehat{\sigma}_j - 1, [rT]) + \frac{\lambda_r \mathbf{E}[\sigma]}{\zeta_r}\,\mathbf{E}[\min(\widehat{\sigma} - 1, [rT])].$$

Hence, for every $\varepsilon > 0$, it is plain that

$$\begin{aligned}
&\mathbf{P}\left[\sup_{0\le t\le T} |S^{(0)r}(t)| > \varepsilon\right] \\
&\le \mathbf{P}\left[\frac{1}{\zeta_r}\sum_{j=1}^{b^r}\min(\widehat{\sigma}_j - 1, [rT]) + \frac{\lambda_r \mathbf{E}[\sigma]}{\zeta_r}\,\mathbf{E}\left[\min(\widehat{\sigma} - 1, [rT])\right] > \varepsilon\right] \\
&\le \frac{2\lambda_r \mathbf{E}[\sigma]}{\varepsilon\zeta_r}\,\mathbf{E}\left[\min(\widehat{\sigma} - 1, [rT])\right] \qquad (8.4)
\end{aligned}$$

where the last step follows by Chebyshev's inequality. It is also the case that

$$\begin{aligned}
\mathbf{E}\left[\min(\widehat{\sigma} - 1, [rT])\right] &= \sum_{n=0}^{[rT]-1} \mathbf{P}\left[\min(\widehat{\sigma} - 1, [rT]) > n\right] \\
&= \sum_{n=0}^{[rT]-1} \mathbf{P}\left[\widehat{\sigma} - 1 > n\right] \\
&= \sum_{n=1}^{[rT]} \mathbf{P}\left[\widehat{\sigma} > n\right].
\end{aligned}$$

Appealing to Condition (B) and (8.1), we get

$$\lim_{r\to\infty} \mathbf{E}\left[\frac{1}{\zeta_r}\min(\widehat{\sigma} - 1, [rT])\right] = 0 \qquad (8.5)$$

and the conclusion

$$\lim_{r\to\infty} \mathbf{P}\left[\sup_{0\le t\le T} |S^{(0)r}(t)| > \varepsilon\right] = 0$$

immediately obtains from (3.2) upon letting r go to infinity in (8.4). ■

A proof of Proposition 6.2. Fix $r = 1, 2, \ldots$. From (6.2) and (6.4) we note that

$$\begin{aligned}
w_n^r - s_n^{(a)r} &= \sum_{k=1}^{n}\sum_{j=1}^{\beta_k^r} (\sigma_{k,j} - (n - k + 1))^+ \\
&= \sum_{k=1}^{n}\sum_{j=1}^{\beta_{j-k+1}^r} (\sigma_{n-k+1,j} - k)^+ \\
&=_{st} \sum_{k=1}^{n}\sum_{j=1}^{\beta_k^r} (\sigma_{k,j} - k)^+, \quad n = 1, 2, \ldots \qquad (8.6)
\end{aligned}$$

where in the last step we used the mutual independence of the families of i.i.d. rvs $\{\beta_k^r,\ k = 1, 2, \ldots\}$ and $\{\sigma_{k,j},\ k, j = 1, 2, \ldots\}$ are mutually independent. It is now straightforward to check that

$$\sup_{0 \leq t \leq T} |W^r(t) - S^{(a)r}(t)| \leq_{st} \frac{1}{\zeta_r} \sum_{k=1}^{[rT]} \sum_{n=1}^{\beta_k^r} (\sigma_{k,n} - k)^+ + \frac{\lambda_r}{\zeta_r} \sum_{k=1}^{[rT]} \mathbf{E}\left[(\sigma - k)^+\right]. \quad (8.7)$$

By Chebyshev's inequality, for every $\varepsilon > 0$ we obtain

$$\begin{aligned} \mathbf{P}\left[\sup_{0 \leq t \leq T} |W^r(t) - S^{(a)r}(t)| > \varepsilon\right] &\leq \frac{2\lambda_r}{\varepsilon \zeta_r} \sum_{k=1}^{[rT]} \mathbf{E}\left[(\sigma - k)^+\right] \\ &= \frac{2\lambda_r \mathbf{E}[\sigma]}{\varepsilon \zeta_r} \sum_{k=1}^{[rT]} \mathbf{P}\left[\widehat{\sigma} > k\right], \end{aligned}$$

and the desired convergence

$$\sup_{0 \leq t \leq T} |W^r(t) - S^{(a)r}(t)| \xrightarrow{P}_r 0 \quad (8.8)$$

follows upon letting r go to infinity in the upper bound (8.8), and making use of (3.2), Condition (B) and (8.1). ■

The proof of Proposition 6.4 requires estimates that derive from various martingales inequalities; we now state them in Lemmas 8.1 and 8.2 for easy reference: Consider a collection of integrable rvs $\{X_i,\ i = 1, \ldots, n\}$ adapted with respect to the filtration $\{\mathcal{F}_i,\ i = 1, \ldots, n\}$, i.e., for each $i = 1, \ldots, n$, the rv X_i is $\mathcal{F}_i$-measurable. We also write

$$S_i = X_1 + \ldots + X_i, \quad i = 1, \ldots, n. \quad (8.9)$$

Lemma 8.1 (Maximal inequality [13]) *Assume $\{(S_i, \mathcal{F}_i),\ i = 1, \ldots, n\}$ to form a martingale. Then, for each $p \geq 1$, it holds that*

$$\mathbf{P}\left[\max_{i=1,\ldots,n} |S_i| > \lambda\right] \leq \lambda^{-p} \mathbf{E}\left[|S_n|^p\right], \quad \lambda > 0.$$

Lemma 8.2 (von Bahr – Esseen inequality [2]) *Assume $\{(X_i, \mathcal{F}_i),\ i = 1, \ldots, n\}$ to form a martingale difference. If $\mathbf{E}[|X_i|^p] < \infty$ for all $i = 1, \ldots, n$, then*

$$\mathbf{E}\left[|S_n|^p\right] \leq 2 \sum_{i=1}^{n} \mathbf{E}\left[|X_i|^p\right], \quad 1 \leq p \leq 2.$$

In what follows Lemmas 8.1 and 8.2 are applied to the case when the rvs $\{X_i,\ i = 1,\ldots,n\}$ are zero–mean i.i.d. rvs.

A proof of Proposition 6.4. Recall that the rvs $\{\beta_k,\ k = 1,2,\ldots\}$ are i.i.d. Poisson rvs with parameter $c/\mathbf{E}[\sigma]$, which are independent of the sequence of i.i.d. service rvs $\{\sigma_{k,j},\ k,j = 1,2,\ldots\}$.

Fix $r = 1,2,\ldots$. On the same probability triple $(\Omega,\mathcal{F},\mathbf{P})$ where the previously mentioned rvs are defined, we introduce a family of i.i.d. $\{0,1\}$–valued rvs $\{U^r_{k,j},\ k,j = 1,2,\ldots\}$, i.e.,

$$\mathbf{P}[U^r = 1] = \frac{\lambda_r \mathbf{E}[\sigma]}{c} = 1 - \mathbf{P}[U^r = 0]$$

where U^r denotes the generic rv for this i.i.d. sequence. The rvs $\{U^r_{k,j},\ k,j = 1,2,\ldots\}$ are assumed independent of the collections of rvs mentioned so far. Next, we define the rvs $\{\widetilde{\beta}^r_k, k = 1,2,\ldots\}$ by

$$\widetilde{\beta}^r_k \equiv \sum_{j=1}^{\beta_k} U^r_{k,j}, \quad k = 1,2,\ldots$$

We also define the workload process $\{\widetilde{w}^r_n,\ n = 0,1,\ldots\}$ corresponding to $\{\widetilde{\beta}^r_k,\ k = 1,2,\ldots\}$ by

$$\widetilde{w}^r_0 \equiv 0, \quad \widetilde{w}^r_n \equiv \sum_{k=1}^{n}\sum_{j=1}^{\widetilde{\beta}^r_k} \sigma_{k,j}, \quad n = 1,2,\ldots$$

and its rescaled version $\{\widetilde{W}^r(t),\ t \geq 0\}$ by

$$\widetilde{W}^r(t) \equiv \frac{1}{\zeta_r}\left(\widetilde{w}^r_{[rt]} - \mathbf{E}\left[\widetilde{w}^r_{[rt]}\right]\right), \quad t \geq 0.$$

Under the enforced independence assumptions, it is easy to check that $\{\widetilde{\beta}^r_k,\ k = 1,2,\ldots\} =_{st} \{\beta^r_k,\ k = 1,2,\ldots\}$, and that

$$\{\widetilde{W}^r(t),\ 0 \leq t \leq T\} =_{st} \{W^r(t),\ 0 \leq t \leq T\}.$$

Moreover, these rvs are *all* defined on the *same* probability triple as the rescaled process $\{V^r(t),\ t \geq 0\}$. Thus, the result will be established if it holds that

$$\{V^r(t) - \widetilde{W}^r(t),\ 0 \leq t \leq T\} \stackrel{U}{\Longrightarrow}_r \theta_T \quad \text{in } D[0,T],$$

or equivalently, if we can show that

$$\sup_{0\le t\le T} |V^r(t) - \widetilde{W}^r(t)| \xrightarrow{P}_r 0. \tag{8.10}$$

To that end, for each $r = 1, 2, \ldots$, we note from the definitions that

$$\widetilde{w}_n^r = \sum_{k=1}^{n}\sum_{j=1}^{\beta_k} U_{k,j}^r \sigma_{k,j}, \quad n = 1, 2, \ldots$$

so that

$$v_n^r - \widetilde{w}_n^r = \sum_{k=1}^{n}\sum_{j=1}^{\beta_k} (1 - U_{k,j}^r)\sigma_{k,j}, \quad n = 1, 2, \ldots \tag{8.11}$$

The rvs $\{Z_k^r,\ k = 1, 2, \ldots\}$ defined by

$$Z_k^r \equiv \sum_{j=1}^{\beta_k} (1 - U_{k,j}^r)\sigma_{k,j}, \quad k = 1, 2, \ldots \tag{8.12}$$

are i.i.d., and we denote by Z^r the corresponding generic rv associated with this collection of rvs. It is plain from (8.11) and (8.12) that

$$\sup_{0\le t\le T} |V^r(t) - \widetilde{W}^r(t)| = \frac{1}{\zeta_r} \sup_{1\le n\le [rT]} \left| \sum_{k=1}^{n} (Z_k^r - \mathbf{E}[Z_k^r]) \right|. \tag{8.13}$$

Fix $\varepsilon > 0$. Invoking the maximal inequality for martingale sequences [Lemma 8.1], we get

$$\mathbf{P}\left[\sup_{0\le t\le T} |V^r(t) - \widetilde{W}^r(t)| > \varepsilon \right] \le \frac{1}{(\varepsilon\zeta_r)^p}\, \mathbf{E}\left[\left| \sum_{k=1}^{[rT]} (Z_k^r - \mathbf{E}[Z_k^r]) \right|^p \right] \tag{8.14}$$

with p selected such that $1 < p < \alpha \le 2$. This selection of p ensures $\mathbf{E}[\sigma^p] < \infty$ both under short–range dependence and under the assumption of regularly varying tail (4.1). The von Bahr – Esseen inequality [Lemma 8.2] for martingale differences can now be applied to the right handside of (8.14) to yield

$$\frac{1}{(\varepsilon\zeta_r)^p}\, \mathbf{E}\left[\left| \sum_{k=1}^{[rT]} (Z_k^r - \mathbf{E}[Z_k^r]) \right|^p \right] \le \frac{2[rT]}{(\varepsilon\zeta_r)^p} \mathbf{E}\left[|Z^r - \mathbf{E}[Z^r]|^p \right]. \tag{8.15}$$

By the convexity of $x \to x^p$ $(p > 1)$ on $\mathbb{R}_+$, we find

$$\mathbf{E}\left[|Z^r - \mathbf{E}[Z^r]|^p \right] \le 2^{p-1}(\mathbf{E}[|Z^r|^p] + \mathbf{E}[|Z^r|]^p) \le 2^p \mathbf{E}[|Z^r|^p] \tag{8.16}$$

with the last step validated by Jensen's inequality. Next, using the definition of Z^r, we obtain by the same convexity argument that

$$\mathbf{E}\left[|Z^r|^p|\beta\right] \leq \beta^{p-1}\,\mathbf{E}\left[\sum_{j=1}^{\beta}(1-U_j^r)^p\sigma_j^p\;|\beta\right] = \beta^p\mathbf{E}\left[\sigma^p\right]\mathbf{E}\left[(1-U^r)^p\right] \quad a.s. \quad (8.17)$$

under the enforced independence assumptions (and with an obvious notation). Injecting the bounds (8.16) and (8.17) into (8.15), we conclude from (8.14) that

$$\mathbf{P}\left[\sup_{0\leq t\leq T}|V^r(t)-\widetilde{W}^r(t)|>\varepsilon\right] \leq \frac{2^{p+1}}{\varepsilon^p\zeta_r^{p-1}}\,\frac{\lceil rT\rceil}{r}\mathbf{E}\left[\sigma^p\right]\mathbf{E}\left[\beta^p\right]\frac{r}{\zeta_r}\mathbf{E}\left[(1-U^r)^p\right] \quad (8.18)$$

As the heavy traffic Assumption (A) implies

$$\lim_{r\to\infty}\frac{r}{\zeta_r}\mathbf{E}\left[(1-U^r)^p\right] = \frac{\gamma}{c}, \quad (8.19)$$

the desired conclusion (8.10) now follows by letting r go to infinity in (8.18), using (8.19) and noting that $\lim_{r\to\infty} 1/\zeta_r^{p-1} = 0$ for $p > 1$. ■

9 A proof of Proposition 7.1

In the proof of Proposition 7.1 and elsewhere, we make use of the following fact.

Lemma 9.1 *For any slowly varying function $u : \mathbb{R}_+ \to \mathbb{R}_+$, it holds that*

$$\lim_{x\to\infty} x^\rho u(x) = \infty, \quad \rho > 0, \quad (9.1)$$

while

$$\lim_{x\to\infty} x^\rho u(x) = 0, \quad \rho < 0. \quad (9.2)$$

Proof. By the Representation Theorem for slowly varying functions [6, Theorem 1.3.1, p. 12], we can write

$$u(x) \sim c\exp\left(\int_A^x \frac{\varepsilon(t)}{t}\,dt\right) \quad (x\to\infty) \quad (9.3)$$

with constants $A > 0$ and $c > 0$, and Borel mapping $\varepsilon : \mathbb{R}_+ \to \mathbb{R}$ such that $\lim_{t\to\infty}\varepsilon(t) = 0$. Thus,

$$x^\rho u(x) \sim cA^{-\rho}\exp\left(\int_A^x \frac{\varepsilon(t)+\rho}{t}\,dt\right) \quad (x\to\infty). \quad (9.4)$$

For every $\delta > 0$ there exists $t_\delta > A$ such that $|\varepsilon(t)| < \delta$ for $t > t_\delta$, whence

$$\frac{-\delta+\rho}{t} \leq \frac{\varepsilon(t)+\rho}{t} \leq \frac{\delta+\rho}{t}, \quad t > t_\delta \tag{9.5}$$

so that

$$K + (-\delta+\rho)\ln\left(\frac{x}{t_\delta}\right) \leq \int_A^x \frac{\varepsilon(t)+\rho}{t}\,dt \leq K + (\delta+\rho)\ln\left(\frac{x}{t_\delta}\right), \quad x > t_\delta \tag{9.6}$$

with

$$K \equiv \int_A^{t_\delta} \frac{\varepsilon(t)+\rho}{t}\,dt.$$

The conclusion (9.1) (resp. (9.2)) immediately follows from these inequalities when selecting $\delta > 0$ such that $\delta < \rho$ (resp. $\delta < -\rho$), such a selection is always possible when $\rho > 0$ (resp. $\rho < 0$). ■

The limit (9.1) is useful in the proof of the following discrete analogue to the direct half of Karamata's Theorem [6, p. 26].

Proposition 9.2 *Let $u : \mathbb{R}_+ \to \mathbb{R}_+$ be a slowly varying function. Then the following statements hold:*

(i) For any $p > -1$, we have the asymptotics

$$\sum_{n=1}^{r} n^p u(n) \sim \frac{r^{p+1}}{p+1}\,u(r) \quad (r \to \infty). \tag{9.7}$$

(ii) The mapping $\widehat{u} : \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$\widehat{u}(x) \equiv \sum_{n=1}^{[x]} \frac{u(n)}{n}, \quad x \geq 0 \tag{9.8}$$

is a slowly varying function which satisfies

$$\lim_{x\to\infty} \frac{u(x)}{\widehat{u}(x)} = 0. \tag{9.9}$$

Proof. (i) Pick δ in $(0, 1+p)$. By Potter's bound [6, p. 25], for every $A > 0$ there exists $r_{\delta,A}$ such that

$$\frac{u(n)}{u(r)} < A\left(\frac{n}{r}\right)^{-\delta}, \quad r_{\delta,A} \leq n \leq r. \tag{9.10}$$

Because $p+1>0$, we readily see from (9.1) that

$$\begin{aligned}\frac{1}{r^{p+1}u(r)}\sum_{n=1}^{r} n^p u(n) &\sim \sum_{n=r_{\delta,A}}^{r} \frac{1}{r}\left(\frac{n}{r}\right)^p \frac{u(n)}{u(r)} \quad (r\to\infty)\\ &= \int_0^1 U_r(x)\,dx \qquad (9.11)\end{aligned}$$

where for each $r=1,2,\ldots$, we have defined the functions $U_r:\mathbb{R}_+\to\mathbb{R}$ and $T_r:\mathbb{R}_+\to\mathbb{R}$ by

$$U_r(x)\equiv \mathbf{1}\left[x\geq \frac{r_{\delta,A}}{r}\right] T_r(x)^p\, \frac{u(rT_r(x))}{u(r)}, \quad x\geq 0 \qquad (9.12)$$

and

$$T_r(x)\equiv \frac{[rx]+1}{r}, \quad x\geq 0. \qquad (9.13)$$

For every x in $(0,1]$ and $r=1,2,\ldots$ we note that $x\leq T_r(x)\leq 2$, and the Uniform Convergence Theorem for slowly varying functions [6, Theorem 1.2.1, p. 6] thus implies

$$\lim_{r\to\infty}\frac{u(rT_r(x))}{u(r)}=1, \quad 0<x\leq 1. \qquad (9.14)$$

Combining (9.12), (9.13) and (9.14) we readily conclude to the pointwise convergence

$$\lim_{r\to\infty} U_r(x)=x^p, \quad 0\leq x\leq 1. \qquad (9.15)$$

Simple bounding arguments based on (9.10) and the classical asymptotics

$$\lim_{r\to\infty}\sum_{n=r_{\delta,A}}^{r}\frac{1}{r}\left(\frac{n}{r}\right)^{p-\delta}=\frac{1}{p-\delta+1}<\infty$$

can be used to validate the use of the dominated convergence theorem, so that

$$\lim_{r\to\infty}\int_0^1 U_r(x)dx=\int_0^1 x^p dx=\frac{1}{p+1}.$$

The conclusion (9.7) follows readily from (9.11).

(ii) Pick ε in $(0,1)$. For each $x>0$ we have

$$\begin{aligned}\hat{u}(x) &\geq \sum_{n=[\varepsilon x]+1}^{[x]}\frac{u(n)}{n}\\ &= u(x)\sum_{n=[\varepsilon x]+1}^{[x]}\frac{1}{n}\frac{u(x\frac{n}{x})}{u(x)}\\ &\sim u(x)\sum_{n=[\varepsilon x]+1}^{[x]}\frac{1}{n} \quad (x\to\infty), \qquad (9.16)\end{aligned}$$

where the asymptotic equality follows by the Uniform Convergence Theorem [6, Theorem 1.2.1, p. 6]. Noting that

$$\lim_{x\to\infty} \sum_{n=[\varepsilon x]+1}^{[x]} \frac{1}{n} = -\ln \varepsilon \tag{9.17}$$

where ε can be chosen arbitrarily small, we obtain from (9.16) that

$$\lim_{x\to\infty} \frac{\widehat{u}(x)}{u(x)} = \infty \tag{9.18}$$

or equivalently, (9.9).

To prove that $\widehat{u}$ is slowly varying, pick $y > 1$ and note for every $x > 0$ that

$$\widehat{u}(yx) = \widehat{u}(x) + u(x) \sum_{n=[x]+1}^{[yx]} \frac{1}{n}\frac{u(n)}{u(x)}.$$

Applying the Uniform Convergence Theorem [6, Theorem 1.2.1, p. 6] once more we get

$$\sum_{n=[x]+1}^{[yx]} \frac{1}{n}\frac{u(n)}{u(x)} = \sum_{n=[x]+1}^{[yx]} \frac{1}{n}\frac{u(x\frac{n}{x})}{u(x)} \sim u(x)\ln y \quad (x\to\infty),$$

whence

$$\frac{\widehat{u}(yx)}{\widehat{u}(x)} \sim 1 + \frac{u(x)}{\widehat{u}(x)} \ln y \sim 1 \quad (x \to \infty)$$

upon using (9.9). The case $y < 1$ is handled in a similar way, and the slow variation of $\widehat{u}$ follows. ■

A proof of Proposition 7.1 From (2.1) it always holds that

$$\mathbf{E}\left[\widehat{\sigma}\right] = \sum_{n=0}^{\infty} \mathbf{P}\left[\widehat{\sigma} > n\right] = \frac{\mathbf{E}\left[\sigma^2\right]}{2\mathbf{E}\left[\sigma\right]} + \frac{1}{2}. \tag{9.19}$$

We consider each of the scalings $\{\zeta_r,\ r = 1, 2, \ldots\}$ associated with Theorems 4.1 – 4.3, separately:

[Theorem 4.1] Under short–range dependence, we have $\mathbf{E}\left[\sigma^2\right] < \infty$, and it is immediate from (9.19) that Condition (B) holds for the choice $\zeta_r = \sqrt{r}$, $r = 1, 2, \ldots$ (in fact for any choice such that $\lim_{r\to\infty} \zeta_r = \infty$). ■

We next turn to Theorems 4.2 and 4.3. Upon substituting (4.1) (with $1 < \alpha \leq 2$) into (2.1), we readily get from Proposition 9.2(i) that

$$\mathbf{P}[\widehat{\sigma} > n] = \frac{1}{\mathbf{E}[\sigma]} \sum_{j=n}^{\infty} \mathbf{P}[\sigma > j] \sim \frac{1}{(\alpha - 1)\mathbf{E}[\sigma]} n^{1-\alpha} h(n) \quad (n \to \infty), \tag{9.20}$$

whence

$$\sum_{n=1}^{r} \mathbf{P}[\widehat{\sigma} > n] \sim \frac{1}{(\alpha - 1)\mathbf{E}[\sigma]} \sum_{n=1}^{r} n^{1-\alpha} h(n) \quad (r \to \infty) \tag{9.21}$$

provided $\mathbf{E}[\widehat{\sigma}]$ is infinite.

[Theorem 4.2] When $\alpha = 2$ in (4.1), the condition $\mathbf{E}[\sigma^2] = \infty$ implies that $\mathbf{E}[\widehat{\sigma}]$ is infinite by (9.19). Thus,

$$\sum_{n=1}^{r} \mathbf{P}[\widehat{\sigma} > n] \sim \frac{1}{\mathbf{E}[\sigma]} \sum_{n=1}^{r} \frac{h(n)}{n} \quad (r \to \infty) \tag{9.22}$$

which, from Proposition 9.2(ii) is seen to be slowly varying. By Proposition 4.4, the scaling $\{\zeta_r,\ r = 1, 2, \ldots\}$ is 1/2–regularly varying, so that

$$\frac{1}{\zeta_r} \sum_{n=1}^{r} \mathbf{P}[\widehat{\sigma} > n] \sim \frac{1}{\mathbf{E}[\sigma]} r^{-\frac{1}{2}} \frac{1}{\widehat{h}(r)} \sum_{n=1}^{r} \frac{h(n)}{n} \quad (r \to \infty) \tag{9.23}$$

for some slowly varying function $\widehat{h} : \mathbb{R}_+ \to \mathbb{R}_+$. The ratio of slowly varying functions being itself slowly varying, we readily conclude from Lemma 9.1 and (9.23) that Condition (B) holds. ■

[Theorem 4.3] On the range $1 < \alpha < 2$, $\mathbf{E}[\sigma^2]$ is infinite, and so is $\mathbf{E}[\widehat{\sigma}]$ by virtue of (9.19). Proposition 9.2(i) applied to the right handside of the asymptotic equivalence (9.21) yields

$$\sum_{n=1}^{r} \mathbf{P}[\widehat{\sigma} > n] \sim \frac{1}{(2-\alpha)(\alpha - 1)\mathbf{E}[\sigma]} r^{2-\alpha} h(r) \quad (r \to \infty). \tag{9.24}$$

By Proposition 4.4 the scaling $\{\zeta_r,\ r = 1, 2, \ldots\}$ is $1/\alpha$–regularly varying, so that

$$\frac{1}{\zeta_r} \sum_{n=1}^{r} \mathbf{P}[\widehat{\sigma} > n] \sim \frac{1}{(2-\alpha)(\alpha - 1)\mathbf{E}[\sigma]} \frac{h(r)}{\widehat{h}(r)} \frac{r^{2-\alpha}}{r^{\frac{1}{\alpha}}} \quad (r \to \infty) \tag{9.25}$$

for some slowly varying function $\widehat{h} : \mathbb{R}_+ \to \mathbb{R}_+$. The ratio of slowly varying functions is itself slowly varying, and Condition (B) is now a direct consequence of Lemma 9.1 once we note that $2 - \alpha - \alpha^{-1} < 0$ for $\alpha > 0$. ■

10 Appendix – Asymptotic invertibility of regularly varying functions

The proof of Proposition 4.4 relies on the following general fact.

Proposition 10.1 *Consider a Borel measurable function $\beta : \mathbb{R}_+ \to \mathbb{R}$ such that $\lim_{x\to\infty} \beta(x) = 0$. With $A > 0$, the equation*

$$y = \exp\left(\int_A^x \frac{1-\beta(t)}{t}\, dt\right), \quad x \geq A, y > 0 \tag{10.1}$$

has a unique solution $x \equiv x(y)$ for all y large enough. Moreover,

$$\lim_{y\to\infty} \frac{x(\gamma y)}{x(y)} = \gamma, \quad \gamma > 0, \tag{10.2}$$

or equivalently the mapping $y \to x(y)$ is regularly varying of order 1, i.e., $x(y) \sim yv(y)$ ($y \to \infty$) for some slowly varying function $v : \mathbb{R}_+ \to \mathbb{R}_+$.

Proof. Set

$$B(x) \equiv \int_A^x \frac{1-\beta(t)}{t}\, dt, \quad x \geq A \tag{10.3}$$

and pick ε in $(0,1)$. Since $\lim_{x\to\infty} \beta(x) = 0$, there exists $A^\star = A^\star(\varepsilon) > A$ such that

$$\frac{1-\varepsilon}{t} < \frac{1-\beta(t)}{t} < \frac{1+\varepsilon}{t}, \quad t \geq A^\star. \tag{10.4}$$

It is straightforward to see that $\lim_{x\to\infty} B(x) = \infty$ and $B(A) = 0$, and by continuity, the range of $x \to B(x)$ contains the semi–infinite interval $[1, \infty)$. We also conclude from (10.4) that $x \to B(x)$ is strictly monotone increasing on the interval $[A^\star, \infty)$, and the existence and uniqueness of a solution to (10.1) follows whenever $y \geq y^\star$ with $y^\star \equiv \exp(B(A^\star))$. The solution mapping $y \to x(y)$ is strictly increasing on $[y^\star, \infty)$.

We now turn to proving (10.2). There is nothing to prove when $\gamma = 1$. With $\gamma > 1$, whenever $y \geq y^\star$, we get

$$\frac{\gamma y}{y} = \exp(B(x(\gamma y)) - B(x(y))) = \exp\left(\int_{x(y)}^{x(\gamma y)} \frac{1-\beta(t)}{t}\, dt\right), \tag{10.5}$$

and the use of the inequalities (10.4) yields

$$\left[\frac{x(\gamma y)}{x(y)}\right]^{1-\varepsilon} \leq \gamma \leq \left[\frac{x(\gamma y)}{x(y)}\right]^{1+\varepsilon}, \tag{10.6}$$

or equivalently,

$$\gamma^{\frac{1}{1+\varepsilon}} \leq \frac{x(\gamma y)}{x(y)} \leq \gamma^{\frac{1}{1-\varepsilon}}. \tag{10.7}$$

Letting y go to infinity in (10.7), we get

$$\gamma^{\frac{1}{1+\varepsilon}} \leq \liminf_{y\to\infty} \frac{x(\gamma y)}{x(y)} \leq \limsup_{y\to\infty} \frac{x(\gamma y)}{x(y)} \leq \gamma^{\frac{1}{1-\varepsilon}}, \tag{10.8}$$

and (10.2) is obtained as we note that ε is arbitrary in $(0,1)$. The case $\gamma < 1$ is handled in a similar way; details are omitted in the interest of brevity. ■

Lemma 10.2 *Consider slowly varying functions $u, w : \mathbb{R}_+ \to \mathbb{R}_+$, such that $u(x) \sim w(x)$ $(x \to \infty)$, and let $\alpha > 1$. For any sequences $\{\zeta_r,\ r = 1,2,\ldots\}$ and $\{\eta_r,\ r = 1,2,\ldots\}$ with $\lim_{r\to\infty} \zeta_r = \lim_{r\to\infty} \zeta_r = \infty$ such that*

$$\lim_{r\to\infty} r\zeta_r^{-\alpha} u(\zeta_r) = \lim_{r\to\infty} r\eta_r^{-\alpha} w(\eta_r) = K \tag{10.9}$$

for some finite constant $K > 0$, it holds that $\zeta_r \sim \eta_r$ $(r \to \infty)$.

Proof. We first look at the special case when $u = w$, in which case condition (10.9) implies

$$\lim_{r\to\infty} \frac{\zeta_r^{-\alpha} u(\zeta_r)}{\eta_r^{-\alpha} u(\eta_r)} = 1. \tag{10.10}$$

We refer to the proof of Lemma 9.1, where we introduced the asymptotically equivalent representation (9.3) of the slowly varying function u. Substituting (9.3) in (10.10), we see that

$$\lim_{r\to\infty} \exp\left(-\int_A^{\zeta_r} \frac{\alpha - \varepsilon(t)}{t}\, dt + \int_A^{\eta_r} \frac{\alpha - \varepsilon(t)}{t}\, dt\right) = 1,$$

or, equivalently,

$$\lim_{r\to\infty} \left| \int_{\zeta_r}^{\eta_r} \frac{\alpha - \varepsilon(t)}{t}\, dt \right| = 0. \tag{10.11}$$

Pick δ in $(0, \alpha)$. Because $\lim_{r\to\infty} \zeta_r = \lim_{r\to\infty} \eta_r = \infty$, there exists r_δ such that for $r > r_\delta$ we have $|\varepsilon(t)| < \delta$ whenever $t > \min(\zeta_r, \eta_r)$. Thus,

$$(\alpha - \delta) \left| \ln \frac{\eta_r}{\zeta_r} \right| < \left| \int_{\zeta_r}^{\eta_r} \frac{\alpha - \varepsilon(t)}{t}\, dt \right|, \quad r > r_\delta,$$

and combining this last inequality with (10.11) we obtain the desired conclusion $\lim_{r\to\infty} \eta_r/\zeta_r = 1$.

In general, when u and w are not necessarily equal, we note the easy relation

$$\frac{\zeta_r^{-\alpha}u(\zeta_r)}{\eta_r^{-\alpha}u(\eta_r)} = \frac{r\zeta_r^{-\alpha}u(\zeta_r)}{r\eta_r^{-\alpha}w(\eta_r)} \cdot \frac{w(\eta_r)}{u(\eta_r)}, \quad r = 1, 2, \ldots \tag{10.12}$$

Condition (10.9) and the asymptotic equivalence of u and w together imply that the relation (10.10) still holds, and the conclusion $\zeta_r \sim \eta_r$ $(r \to \infty)$ follows from the first part of the proof. ■

Proposition 10.3 *Consider a slowly varying function $u : \mathbb{R}_+ \to \mathbb{R}_+$, and let $\alpha > 1$. For any sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ with $\lim_{r\to\infty} \zeta_r = \infty$ such that (10.9) holds, we have*

$$\zeta_r \sim r^{\frac{1}{\alpha}} w(r) \quad (r \to \infty) \tag{10.13}$$

for some slowly varying function $w : \mathbb{R}_+ \to \mathbb{R}_+$.

Proof. We go back to the proof of Lemma 9.1, where we introduced the asymptotically equivalent representation (9.3) of the slowly varying function u. In view of Lemma 10.2, it suffices to consider a sequence $\{\zeta_r,\ r = 1, 2, \ldots\}$ determined by the relations

$$r\zeta_r^{-\alpha} \cdot c \exp\left(\int_A^{\zeta_r} \frac{\varepsilon(t)}{t}\, dt\right) = K, \quad r \geq r_\star \tag{10.14}$$

with $r^\star$ large enough, with constants $A > 0$ and $c > 0$, and Borel mapping $\varepsilon : \mathbb{R}_+ \to \mathbb{R}$ such that $\lim_{t\to\infty} \varepsilon(t) = 0$. By simple manipulations, we can write (10.14) in the equivalent form

$$Br^{\frac{1}{\alpha}} = \exp\left(\int_A^{\zeta_r} \frac{1 - \beta(t)}{t}\, dt\right), \quad r \geq r_\star \tag{10.15}$$

with

$$B \equiv \left(\frac{c}{KA^\alpha}\right)^{\frac{1}{\alpha}} \quad \text{and} \quad \beta(t) \equiv \frac{1}{\alpha}\varepsilon(t), \quad t \geq 0. \tag{10.16}$$

Hence, by Proposition 10.1, for large enough r we see that ζ_r is the unique solution $x(y)$ of the equation (10.1) with $y = Br^{\frac{1}{\alpha}}$. By the second part of Proposition 10.1, we have

$$\zeta_r = x(Br^{\frac{1}{\alpha}}) \sim Br^{\frac{1}{\alpha}} v(Br^{\frac{1}{\alpha}}) \quad (r \to \infty), \tag{10.17}$$

and the desired conclusion is now immediate once we note that the mapping w : $x \to Bv(Bx^{\frac{1}{\alpha}})$ is slowly varying whenever v is. ■

References

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

[2] B. von Bahr and C. G. Esseen, "Inequalities for the r^{th} absolute moment of a sum of random variables, $1 \leq r \leq 2$," *Annals of Mathematical Statistics*, **36** (1965), pp. 299–303.

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

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York (NY), 1968.

[5] N. H. Bingham, "Fluctuation theory in continuous time," *Advances in Applied Probability*, **7** (1975), pp. 705–766.

[6] N. H. Bingham, C. M. Goldie and J. T. Teugels, *Regular Variation. Encyclopedia of Mathematics and its Applications*, Cambridge Univerity Press, Cambridge (UK), 1987.

[7] 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.

[8] D. R. Cox, "Long–range dependence: A review," In H. A. David and H. T. David, editors, *Statistics: An Appraisal*, The Iowa State University Press, Ames (IA), 1984, pp. 55-74.

[9] N. G. Duffield, "On the relevance of long–tailed durations for the statistical multiplexing of large aggregations," in *Proceedings of the 34th Annual Allerton Conference on Communications, Control and Computing*, October 1996.

[10] A. Erramilli, O. Narayan, and W. Willinger, " Experimental queueing analysis with long–range dependent packet traffic," *IEEE/ACM Transactions on Networking* **4** (1996), pp. 209–223.

[11] W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume II, Second Edition, John Wiley & Sons, New York (NY), 1972.

[12] J. M. Harrison, *Brownian Motion and Stochastic Flow systems.* John Wiley & Sons, New York (NY), 1985.

[13] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, New York (NY), 1980.

[14] P. R. Jelenković and A. A. Lazar, "Multiplexing on–off sources with subexponential on periods: part I," *Proceedings of IEEE Infocom 97*, Kobe (Japan), April 1997.

[15] T. Konstantopoulos and S.–J. Lin, "Fractional Brownian motions and Lévy motions as limits of stochastic traffic models," in *Proceedings of the 34th Annual Allerton Conference on Communication, Control and Computation*, pp. 913–922, October 1996.

[16] J. Lamperti, "Semi–stable stochastic processes," *Transactions of the American Mathematical Society* **104** (1962), pp. 62–78.

[17] W. Leland, M. S. 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] N. Likhanov, B. Tsybakov, and N. D. Georganas, "Analysis of an ATM buffer with self–similar (fractal) input traffic," in *Proceedings of IEEE Infocom 95*, pp. 985–992, Boston (MA), April 1995.

[19] Z. Liu, Ph. Nain, D. Towsley and Z.–L. Zhang, "Asymptotic behavior of a multiplexer fed by a long–range dependent process," Preprint, February 1997.

[20] C. M. Newman and A. L. Wright, "An invariance principle for certain dependent sequences," *The Annals of Probability* **9** (1981), pp. 671–675.

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

[22] M. Parulekar and A. M. Makowski, "Tail probabilities for $M|GI|\infty$ processes (I): Preliminary asymptotics," submitted to *Queueing Systems – Theory and Applications*, 1996.

[23] M. Parulekar and A. M. Makowski, "$M|GI|\infty$ input processes : A versatile class of models for network traffic," in *Proceedings of IEEE Infocom 97*, Kobe (Japan), April 1997.

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

[25] M. I. Reiman and B. Simon, "An interpolation approximation for queueing systems with Poisson input," *Operations Research* **36** (1988), pp. 454–469.

[26] M. I. Reiman and B. Simon, "Light traffic limits of sojourn time distributions in Markovian queuing networks," *Stochastic Models* **4** (1988), pp. 191–233.

[27] M. I. Reiman and B. Simon, "Open queueing systems in light traffic," *Mathematics of Operations Research* **14** (1989), pp. 26-59.

[28] G. Samorodnitsky and M. S. Taqqu, *Stable Non–Gaussian Random Processes*, Chapman and Hall, London (UK), 1994.

[29] A. V. Skorokhod, "Limit theorems for stochastic processes with independent increments," *Theory of Probability and its Applications*, **2** (1957), pp. 138–171.

[30] V. Solo, "On queueing theory for broadband communication network traffic with long range correlation,"in *Proceedings of the 34th Conference on Decision and Control*, pp. 853–858, New Orleans (LA), December 1995.

[31] K. P. Tsoukatos and A. M. Makowski, "Heavy traffic analysis for a multiplexer driven by $M|GI|\infty$ input processes," in *Proceedings of ITC 15*, Washington (DC), June 1997.

[32] K. P. Tsoukatos, Ph.D. Thesis, Electrical Engineering Department, University of Maryland, College Park (MD), in preparation.

[33] W. Whitt, "Some useful functions for functional limit theorems," *Mathematics of Operations Research* **5** (1980), pp. 67–85.