

LIMIT BEHAVIOR OF FLUID QUEUES AND NETWORKS

BERNARDO D'AURIA AND GENNADY SAMORODNITSKY

ABSTRACT. A superposition of a large number of infinite source Poisson inputs or that of a large number of ON-OFF inputs with heavy tails can look like either a Fractional Brownian Motion or a stable Lévy Motion, depending on the magnification at which we are looking at the input process (Mikosch et al. (2002)). In this paper we investigate what happens to a queue driven by such inputs. Under such conditions, we study the output of a single fluid server and the behavior of a fluid queuing network. For network we obtain random field limits describing the activity at different stations. In general, both kinds of stations arise in the same network: the stations of the first kind experience loads driven by a Fractional Brownian Motion, while the stations of the second kind experience loads driven by a stable Lévy motion.

1. INTRODUCTION

After the discovery that the internet data traffic has unusual statistical characteristics such as self-similarity, long range dependence (LRD) and heavy tails, many new models have been developed to explain the origin of these new features, interaction between them, and to take these features into account.

Two of the most popular models view the data traffic as the superposition of many contributions due to independent sources of data. They are the so called infinite source Poisson model (otherwise known as the $M/G/\infty$ input flow model) and the superposition of ON/OFF sources.

In the $M/G/\infty$ input flow model, the transmissions (sources) start according to a homogeneous Poisson process. The durations of the transmissions are mutually independent and identically distributed, and independent of the Poisson process. If the transmission times are heavy tailed the resulting input process has slowly decaying correlations and, hence, is viewed as having long range dependence (see e.g. Samorodnitsky (2002)).

In the superposition of ON/OFF sources model, the number of sources is constant and each source switches between ON-periods (transmission) and OFF-periods (silence). For each source,

2000 *Mathematics Subject Classification.* Primary 60K25; secondary 60F05, 60F17, 60G18.

Key words and phrases. queue, queuing network, output process, heavy tailed distribution, long range dependence, fractional Brownian Motion, stable Lévy process, weak convergence.

D'Auria's research was supported by the University of Salerno and the Cornell University during his visit to Cornell University. Samorodnitsky's research was partially supported by NSF grant DMS-0303493 and NSA grant MSPF-02G-183 at Cornell University.

the ON-periods and OFF-periods are assumed to be two sequences of i.i.d. random variables, mutually independent, and different sources are independent as well. Similarly to the previous model, when the transmission times and/or silence times are heavy tailed the resulting input process has slowly decaying correlations and is also viewed as having long range dependence (see Heath et al. (1998)). Slow decay of correlation often comes together with a certain type of scaling, hence the observed self-similarity of the traffic.

While in the literature there is, largely, a consensus on the self-similar nature of the aggregate traffic, different authors report divergent conclusions about the marginal distributions for the cumulative traffic. Indeed, these have been at times reported as light tailed, heavy tailed or intermediate tailed; some of the more recent references are Smith et al. (2001), Campos et al. (2002), Downey (2000) and Gong et al. (2001). One explanation of this phenomenon can be found in the recent papers of Mikosch et al. (2002) and Gaigalas and Kaj (2003) which studied the limit of a sequence of properly scaled input processes both for the $M/G/\infty$ and ON-OFF models. It turns out that the limit depends on relation between the rate at which the transmissions are initiated (called *connection rate* in Mikosch et al. (2002)) and the time scale at which the system is considered. If the connection rate is relatively high, the deviations of the input process from its average look like a Fractional Brownian motion, while if the connection rate is relatively low, these deviations look like a stable motion.

In these papers the attention was focused on the input flow to a single server. We, on the other hand, are interested in the output flows. The knowledge of the output process properties is very useful as the output flow from one station is usually (a part of) the input process for a subsequent queue. Following this line of reasoning it is possible to get insights into the behavior of a whole queuing network.

In this paper we show that the deviations of the output flow from its average in a single queuing system behave similarly to those in the input flow and, hence, satisfy limit results of the kind given in Mikosch et al. (2002). We extend then these results to fluid queuing networks. The results rely on the fact that the stability of the involved queuing systems assures tightness of the buffer content processes. As a consequence, we will see how the marginal distributions, together with the correlation structure, are propagated across the network.

This fact has direct practical applications. For example, in the situations where our limiting results are applicable, empirical finding of sufficiently different marginal behavior in two separate points of a network may give information of low coupling (reducibility) of the network. For

instance, if we find that the properly normalized cumulative input flow at one station has approximately a stable distribution and at another station it has approximately a Gaussian distribution, then the output of the former station does not reach the latter. Otherwise, according to the Theorems 7.1 and 7.2 below, the second distribution would have a heavy tailed stable component as well.

From the modeling point of view, the knowledge of the nature of the limit flows legitimates the study of queuing networks in simplified assumption of Brownian Motion - or stable Lévy Motion - based flows. Note that our flows are, really, random fields. Having a random field description of the limiting flows in a queuing network gives one a state-time description of the loads experienced by the network and, hence, allows one a better understanding of what may happen in such a network. In particular, it is possible to simulate scenarios of behavior of the entire network or its part.

This paper is organized as follows. Section 2 formally defines the models we are studying in the case of a single queuing system. Section 3 contains the results on the tightness of the buffer content process and Section 4 gives limit results for the output process of a single fluid queue. Sections 5, 6 and 7 extend these results to queuing networks.

2. TWO MODELS OF A FLUID QUEUE

In this section we describe two model for a single fluid queue. The queue has an infinite buffer and drains it at constant service rate, say C . The instantaneous input and output flows are denoted respectively by $i(t)$ and $o(t)$ while the cumulative input and output processes are

$$I(t) = \int_0^t i(s) ds \quad \text{and} \quad O(t) = \int_0^t o(s) ds.$$

The buffer content, that is also the workload in the system, is denoted by $W(t)$. Unless noted otherwise we assume that $W(0) = 0$. The workload is related to the cumulative input and output processes through the relation

$$(2.1) \quad O_T(t) = I_T(t) - W_T(t),$$

that expresses the conservation of the flow.

We now introduce the two different models for the instantaneous input process $i(t)$.

2.1. $M/G/\infty$ input flow. This model is also known as the infinite source Poisson input model. Here the input flow $i(t)$ is given by the sum of the contributions of various sources arriving according to a Poisson point process and being active for a random duration.

If $\{t_i\}_{i \in \mathbb{N}}$ are the arrival points of a Poisson point process with intensity λ , and $\{X_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. finite mean random variables independent of the Poisson process, then

$$(2.2) \quad i(t) = \sum_{i=-\infty}^{\infty} 1\{t_i \leq t < t_i + X_i\}.$$

In order to study the limit behavior of the output process we suppose to have a sequence of fluid queues, indexed by T , and we suppose that the intensity of the arrival Poisson process will depend on T , i.e. $\lambda = \lambda(T)$. In particular $\lambda(T)$ is assumed to be an unbounded function of T . The subscripts as in i_T , I_T , etc, will show that a process corresponds to a particular system.

We will show that under our limiting procedure, the normalized difference between the input and output processes converges to 0 (the zero process). To show that this is true, we may allow the random variables $\{X_i\}_{i \in \mathbb{N}}$ to have an arbitrary distribution with a finite mean $\mu = E[X_i]$. However, in order to use the known results on convergence of the input process, in the sequel we suppose that these random variables are appropriately heavy tailed. Specifically, denoting by $F(x)$ their common distribution function, we make the following assumption on its tail $\bar{F}(x) \doteq 1 - F(x)$

$$(2.3) \quad \bar{F}(x) = x^{-\alpha} L(x), \quad x > 0, 1 < \alpha < 2,$$

where $L(x)$ is a slowly varying at infinity function (Feller (1971)).

The behavior of the normalized version of $I_T(Tt)$ as $T \rightarrow \infty$ was studied in Mikosch et al. (2002). In the next section we prove that the finite dimension distributions of $O_T(Tt) - I_T(Tt)$ are tight for $T \geq 0$.

2.2. Superposition of ON-OFF sources. The superposition of ON-OFF sources input model views the traffic as the sum of independent contributions of M flows, each one being a sequence of ON and OFF periods. Specifically, suppose that $\{X_{i,j}^{\text{on}}\}_{i=1}^{\infty}$, $\{X_{i,j}^{\text{off}}\}_{i=1}^{\infty}$ are M independent pairs of independent sequences of i.i.d. random variables, $1 \leq j \leq M$. A generic random variable X^{on} is distributed according to some distribution $F^{\text{on}}(x)$, while a generic random variable X^{off} has a distribution denoted by $F^{\text{off}}(x)$. In the following $\mu^{\text{on}} = E[X^{\text{on}}]$, $\mu^{\text{off}} = E[X^{\text{off}}]$ are assumed to be finite. Denote $\mu = \mu^{\text{on}} + \mu^{\text{off}}$ and $p^{\text{on}} = \frac{\mu^{\text{on}}}{\mu}$.

Construct independent sequences of i.i.d. random variables $\{\tilde{X}_j^{\text{on}}\}_{j=1}^M$ and $\{\tilde{X}_j^{\text{off}}\}_{j=1}^M$ having, respectively, distributions $\tilde{F}^{\text{on}}(x) = \frac{1}{\mu^{\text{on}}} \int_0^x \bar{F}^{\text{on}}(s) ds$, $\tilde{F}^{\text{off}}(x) = \frac{1}{\mu^{\text{off}}} \int_0^x \bar{F}^{\text{off}}(s) ds$, independent of the previously constructed random objects. Following the construction used in Mikosch et al. (2002), we define the random variables

$$T_{0,j} = B_j(\tilde{X}_j^{\text{on}} + X_j^{\text{off}}) + (1 - B_j)\tilde{X}_j^{\text{off}}$$

with (B_j) being a sequence of i.i.d. Bernoulli random variables, once again independent of everything else and with $P\{B = 1\} = p^{\text{on}}$. Define, further, for $n \geq 1$, $T_{n,j} = T_{0,j} + \sum_{i=1}^n (\{X_{i,j}^{\text{on}} + X_{i,j}^{\text{off}}\})$, then

$$(2.4) \quad i(t) = \sum_{j=1}^M \left(B_j 1\{t < \tilde{X}_j^{\text{on}}\} + \sum_{n=1}^{\infty} 1\{T_{n,j} \leq t < T_{n,j} + X_{n+1,j}^{\text{on}}\} \right).$$

As with the previous model, we will restrict our attention to the case

$$(2.5) \quad \bar{F}^{\text{on}}(x) = x^{-\alpha^{\text{on}}} L^{\text{on}}(x), \quad x > 0, 1 < \alpha^{\text{on}} < 2,$$

and

$$(2.6) \quad \bar{F}^{\text{off}}(x) = x^{-\alpha^{\text{off}}} L^{\text{off}}(x), \quad x > 0, 1 < \alpha^{\text{off}} < 2,$$

with $L^{\text{on}}(x)$, $L^{\text{off}}(x)$ slowly varying at infinity functions and $\alpha = \alpha^{\text{on}} < \alpha^{\text{off}}$.

Once again, we will suppose to have a sequence of queuing systems indexed by T whose input process is a superposition of sources. The number of sources $M(T)$ is an unbounded function of T .

As in the $M/G/\infty$ case, we will prove that the finite dimensional distributions of $O_T(Tt) - I_T(Tt)$ are tight for $T \geq 0$.

3. TIGHTNESS OF THE BUFFER CONTENT

In this section we show that the finite dimension distributions of $O_T(T \cdot) - I_T(T \cdot)$ are tight for $T \geq 0$. That is done in the next subsections separately for the two models.

3.1. $M/G/\infty$ input flow. In the statement of the lemma below the notation $X \stackrel{st}{\leq} Y$ means that $P(X \leq x) \geq P(Y \leq x)$ for all $x \in \mathbb{R}$.

Lemma 3.1. *Suppose we have a series of $M/G/\infty$ -input fluid queuing systems indexed by $T \geq 0$ with service rate $C(T) = \gamma\lambda(T)$ and the arrival rate of the $M/G/\infty$ input process, $\lambda(T) \neq 0$, such that $\gamma > \mu$. Then, there exists a random variable H such that*

$$(3.1) \quad W_T(t) \stackrel{st}{\leq} H \quad \forall T, t \geq 0.$$

Proof. The buffer process $W_T(t)$ satisfies the Reich formula

$$(3.2) \quad W_T(t) = \sup_{0 \leq s \leq t} \{I_T(t) - I_T(s) - C(T)(t - s)\}.$$

Evidently $W_T(t) \stackrel{st}{\leq} W_T(\infty) \doteq W_T$, since I_T has stationary increments. Here W_T has the stationary distribution; a simple regenerative argument shows existence of this distribution thanks to

the fact that $\frac{\mu\lambda(T)}{C(T)} < 1$. So in the sequel we consider only the stationary distributions, and we prove that they are uniformly stochastically bounded in T .

Let us consider the stationary buffer distribution W_0 . As the queuing system is stable, it is a well defined random variable.

Consider for each $T \geq 0$ another queuing system, this time not a fluid one. At each time a new session arrives, it adds to the buffer an amount of workload equivalent to its duration. Let $(H_T(t), t \geq 0)$ be the buffer content process associated with this new queuing system. Evidently we have that

$$(3.3) \quad H_T(t) \geq W_T(t) \quad \forall T, t \geq 0$$

and the relation is valid also for the stationary distribution that exists thanks to the fact that the traffic intensity for the new system is less than 1. Hence in the particular case of $T = 0$ we have $W_0 \stackrel{st}{\leq} H_0$.

To complete the proof, let $\eta = \frac{\lambda(T)}{\lambda(0)} \geq 1$. Observe that $(H_T(\eta^{-1}t), t \geq 0)$ is the buffer content of the system where the Poisson point process of arrivals has intensity $\eta^{-1}\lambda(T) = \lambda(0)$ and service rate $\eta^{-1}C(T) = \eta^{-1}\gamma\lambda(T) = \gamma\lambda(0) = C(0)$. Hence $H_T(\eta^{-1}\cdot) \stackrel{fidi}{=} H_0(\cdot)$ and so $H_T \stackrel{d}{=} H_0$, which proves (3.1) with $H = H_0$. \square

3.2. Superposition of ON-OFF sources.

Lemma 3.2. *Let Q_M be a stationary queuing system whose input process is the superposition of M independent and identically distributed stationary ON-OFF processes. Let $C_M = \gamma M$ be the service rate of the queuing system Q_M , with $\gamma > \mu$. Then stationary workload $(W_M, M \geq 1)$ is tight.*

Proof. We transform the fluid system Q_M into a $G/G/1/\infty$ queuing system in the following way.

At each instance of activation of an ON period, the amount of work equal to the length of the ON period is added to the buffer content. As in the proof of Lemma 3.1, the buffer content of the new system cannot be smaller than that of the fluid system. That is, if $(W_M(t), t \geq 0)$ is the buffer content process of Q_M and $(W'_M(t), t \geq 0)$ is the buffer content process of the new system, then

$$W_M(t) \leq W'_M(t) \quad t \geq 0.$$

It is clear that the new system has a unique stationary distribution of the buffer content. Let us call it W'_M . Then

$$W'_M \stackrel{st}{\geq} W_M \quad M \geq 1.$$

Changing time by replacing the buffer process $W'_M(\cdot)$ by $W'_M(\frac{1}{C_M}\cdot)$, preserves the stationary distribution W'_M . In the time-changed system, the arrival process is the superposition of M i.i.d. ON-OFF processes with the time dilated by C_M , and service rate 1. Hence, this is a $G/G/1/\infty$ queue, with the arrival rate $\frac{1}{\gamma(\mu^{\text{on}}+\mu^{\text{off}})}$, and mean service time μ^{on} . We will use the following Proposition 0.2. VI. of Daley and Vere-Jones (1988).

Proposition Let \tilde{N} be a simple point process on \mathbb{R} with finite intensity λ , and let \tilde{N}_M denote the point process obtained by superimposing M independent replicates of \tilde{N} and dilating the time scale by a factor M . Then as $M \rightarrow \infty$, \tilde{N}_M converges weakly to a homogeneous Poisson process with intensity λ .

This implies, in particular, that we are in the framework of the continuity theorem, e.g. Theorem 8.5.1 of Brandt and Lisek (1990), with the limiting system being a stable $M/G/\infty$ queue. If W'_∞ is the steady state buffer content of the latter, we have $W'_M \Rightarrow W'_\infty$, hence $(W'_M, M \geq 1)$ is tight and then so is $(W_M, M \geq 1)$. \square

4. STABLE MOTION OR FRACTIONAL BROWNIAN MOTION AT THE OUTPUT FROM A FLUID QUEUE

As a corollary of the results in the previous section, we can extend the convergence results in Mikosch et al. (2002) to the output processes.

In this section it is essential that the random variables that define the input processes are appropriately heavy tailed. Moreover, as stated in Mikosch et al. (2002), the limit process will depend on the rate of convergence to infinity of the quantities $\lambda(T)$ and $M(T)$ as $T \rightarrow \infty$. These is the reason the *Growth Conditions* are introduced.

In the following $b(t)$ is the quantile function

$$(4.1) \quad b(t) = (1/\bar{F})^{\leftarrow}(t) \quad \text{or} \quad b(t) = (1/\bar{F}^{\text{on}})^{\leftarrow}(t),$$

depending on the model, where for a non-decreasing function F ,

$$(F)^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

4.1. $M/G/\infty$ input flow.

Definition 4.1.

$$\text{Slow Growth Condition (SGC):} \quad \lim_{T \rightarrow \infty} \frac{b(\lambda(T)T)}{T} = 0,$$

$$\text{Fast Growth Condition (FGC):} \quad \lim_{T \rightarrow \infty} \frac{b(\lambda(T)T)}{T} = \infty.$$

Using the previous lemma it is easy to prove the following theorem. Our notation for stable distributions follows that in Samorodnitsky and Taqqu (1994).

Theorem 4.2. *Suppose we have a series of $M/G/\infty$ input fluid queuing systems indexed by T , service rate $C(T) = \gamma\lambda(T)$ with $\gamma > \mu$ and the arrival rate $\lambda(T)$, and $\lambda(T) \uparrow \infty$. Consider the output flow $O_T(\cdot)$. Depending on which growth condition is verified, the following holds:*

$$(4.2) \quad SGC : \frac{O_T(T\cdot) - \mu T \lambda(T)(\cdot)}{b(\lambda(T)T)} \xrightarrow{fidi} X_{\alpha,1,1}(\cdot)$$

$$(4.3) \quad FGC : \frac{O_T(T\cdot) - \mu T \lambda(T)(\cdot)}{[\lambda(T)T^3 \bar{F}(T)]^{\frac{1}{2}}} \xrightarrow{d} \sigma B_H(\cdot), \quad H = \frac{3 - \alpha}{2}.$$

Here $X_{\alpha,1,1}(\cdot)$ is an α -stable Lévy Process with positive jumps, $\sigma^2 = \frac{1}{3-\alpha} \left[\frac{\alpha}{2-\alpha} + \frac{2}{\mu} \right]$, and $B_H(\cdot)$ is the standard Fractional Brownian Motion.

Proof. Using the equation (2.1) and subtracting the mean values, we obtain

$$(4.4) \quad \frac{O_T(T\cdot) - \mu T \lambda(T)(\cdot)}{g(T)} = \frac{I_T(T\cdot) - \mu T \lambda(T)(\cdot)}{g(T)} + \frac{W_T(T\cdot)}{g(T)},$$

where $g(T)$ is, depending on the growth condition, equal to

$$(4.5) \quad SGC : \quad g(T) = b(\lambda(T)T)$$

$$(4.6) \quad FGC : \quad g(T) = [\lambda(T)T^3 \bar{F}(T)]^{\frac{1}{2}}.$$

In any case $\lim_{T \rightarrow \infty} g(T) = \infty$, hence for every $t \geq 0$

$$(4.7) \quad \frac{W_T(Tt)}{g(T)} \xrightarrow[T \rightarrow \infty]{\Rightarrow} 0 \quad \text{in probability}$$

and, together with the results of Theorem 1 and Theorem 3 in Mikosch et al. (2002), respectively for the SGC and for the FGC, this completes the proof for the fidi convergence. Tightness required for weak convergence in $\mathbb{C}[0, \infty)$ for the FGC case is established in Lemma 4.5 below. \square

4.2. Superposition of ON-OFF sources.

Definition 4.3.

$$\text{Slow Growth Condition (SGC): } \lim_{T \rightarrow \infty} \frac{b(M(T)T)}{T} = 0,$$

$$\text{Fast Growth Condition (FGC): } \lim_{T \rightarrow \infty} \frac{b(M(T)T)}{T} = \infty.$$

Using the same argument as in the previous theorem and Lemma 4.5 below establishes the following result.

Theorem 4.4. *Suppose we have a series of fluid queuing systems indexed by T , fed by the superposition of $M(T)$ i.i.d. ON-OFF processes, service rate $C(T) = \gamma M(T)$, with $\gamma > \mu$ and $M(T) \uparrow \infty$. Consider the output flow $O_T(\cdot)$. Depending on which growth condition is verified, the following holds:*

$$(4.8) \quad SGC : \frac{O_T(T \cdot) - \mu^{-1} \mu^{on} T M(T)(\cdot)}{b(M(T)T)} \xrightarrow{fidi} cX_{\alpha, \sigma, 1}(\cdot)$$

$$(4.9) \quad FGC : \frac{O_T(T \cdot) - \mu^{-1} \mu^{on} T M(T)(\cdot)}{[M(T)T^3 \bar{F}^{on}(T)]^{\frac{1}{2}}} \xrightarrow{d} \sigma_0 B_H(\cdot), \quad H = \frac{3 - \alpha}{2}.$$

Here $\sigma = C_\alpha^{-\frac{1}{\alpha}}$, $C_\alpha = \frac{1 - \alpha}{\Gamma(2 - \alpha) \cos(\pi\alpha/2)}$, $c = \frac{\mu^{off}}{\mu^{1+1/\alpha}}$, $X_{\alpha, \sigma, 1}(\cdot)$ is an α -stable Lévy Process, $\sigma_0^2 = \frac{2(\mu^{off})^2 \Gamma(2 - \alpha) / (\alpha - 1)}{\mu^3 \Gamma(4 - \alpha)}$, and $B_H(\cdot)$ is a standard Fractional Brownian Motion.

4.3. Tightness. In order to prove weak convergence in $\mathbb{C}[0, \infty)$ in Theorems 4.2 and 4.4, under FGC condition, we need to prove tightness of the probability measures induced by $\frac{O_T(T \cdot) - E[O_T(T \cdot)]}{g(T)}$ (with the function $g(T)$ appropriate to the $M/G/\infty$ and ON-OFF cases).

Lemma 4.5. *If for some function $g(T)$ the probability measures induced by $\frac{I_T(T \cdot) - E[I_T(T \cdot)]}{g(T)}$ are tight in $(\mathbb{C}[0, \infty), J_1)$, then so are the probability measures induced by $\frac{O_T(T \cdot) - E[O_T(T \cdot)]}{g(T)}$.*

Proof. We need to check tightness in $\mathbb{C}[0, 1]$ as intervals of other lengths can be treated similarly. Since weak convergence in $\mathbb{C}[0, \infty)$ is metrizable, instead of dealing with a family of measures indexed by a continuous parameter T , it is enough to work with an increasing sequence $\{T_n\}_n$.

By Theorem 8.2 of Billingsley (1968) we need to check that for every $\epsilon > 0$ and any $\eta > 0$ there is $\delta > 0$ and $n_0 \geq 0$ such that

$$(4.10) \quad P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t - s < \delta}} \left| \frac{W_{T_n}(T_n t) - W_{T_n}(T_n s)}{g(T_n)} \right| > \epsilon \right) \leq \eta,$$

for all $n \geq n_0$.

We prove (4.10) by dividing the probability in two parts

$$\begin{aligned} P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t - s < \delta}} \left| \frac{W_{T_n}(T_n t) - W_{T_n}(T_n s)}{g(T_n)} \right| > \epsilon \right) &\leq P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t - s < \delta}} \frac{W_{T_n}(T_n t) - W_{T_n}(T_n s)}{g(T_n)} > \epsilon \right) \\ &+ P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t - s < \delta}} \frac{W_{T_n}(T_n s) - W_{T_n}(T_n t)}{g(T_n)} > \epsilon \right), \end{aligned}$$

and show that we can make both probabilities in the right hand side arbitrarily small. Fix $\epsilon > 0$ and consider the event

$$\Omega_1(\delta) \doteq \left\{ \omega : \sup_{\substack{0 \leq s < t \leq 1 \\ t-s < \delta}} \frac{W_{T_n}(T_n t) - W_{T_n}(T_n s)}{g(T_n)} > \epsilon \right\}.$$

For any $\omega \in \Omega_1(\delta)$, choose $0 \leq s_0 < t_0 \leq 1$, $t_0 - s_0 < \delta$ such that

$$W_{T_n}(T_n t_0) - W_{T_n}(T_n s_0) > \epsilon g(T_n).$$

Let

$$u_0 = \sup \{r \in [s_0, t_0] : W_{T_n}(T_n r) = 0\},$$

with the convention that $\sup\{\emptyset\} = s_0$. Then

$$\begin{aligned} \epsilon g(T_n) &< W_{T_n}(T_n t_0) - W_{T_n}(T_n u_0) = I_{T_n}(T_n t_0) - I_{T_n}(T_n u_0) - C(T_n)(T_n(t_0 - u_0)) \\ &\leq I_{T_n}(T_n t_0) - I_{T_n}(T_n u_0) - E[I_{T_n}(1)](t_0 - u_0), \end{aligned}$$

as in the period (u_0, t_0) the buffer is always non-empty and, hence, the system constantly drains flow at the rate $C(T_n) > E[I_{T_n}(1)]$.

Since $t_0 - u_0 < \delta$, we conclude that

$$(4.11) \quad \Omega_1(\delta) \subseteq \left\{ \omega : \sup_{\substack{0 \leq s < t \leq 1 \\ t-s < \delta}} \frac{I_{T_n}(T_n t) - I_{T_n}(T_n s) - E[I_{T_n}(1)](t-s)}{g(T_n)} > \epsilon \right\} \doteq \Omega_0(\delta).$$

Since Mikosch et al. (2002) established tightness for $\frac{I_T(\cdot) - E[I_T(\cdot)]}{g(T)}$, we can use Theorem 8.2 of Billingsley (1968) to conclude that there is a $\delta_1 > 0$ and $n_1 > 0$ such that

$$P(\Omega_0(\delta_1)) \leq \frac{\eta}{2} \quad \text{for all } n \geq n_1.$$

Therefore, for each $\epsilon > 0$ and $\eta > 0$, there are $\delta_1 > 0$ and $n_1 \geq 0$ such that

$$(4.12) \quad P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t-s < \delta_1}} \frac{W_{T_n}(T_n t) - W_{T_n}(T_n s)}{g(T_n)} > \epsilon \right) \leq \frac{\eta}{2},$$

for all $n \geq n_1$.

It remains to prove that, for every $\epsilon > 0$ and $\eta > 0$, there is $\delta_2 > 0$ and $n_2 > 0$ such that

$$(4.13) \quad P \left(\sup_{\substack{0 \leq s < t \leq 1 \\ t-s < \delta_2}} \frac{W_{T_n}(T_n s) - W_{T_n}(T_n t)}{g(T_n)} > \epsilon \right) \leq \frac{\eta}{2},$$

since then (4.10) will follow by choosing $n_0 = \max(n_1, n_2)$ and $\delta = \min(\delta_1, \delta_2)$. The argument in Theorem 8.3 in Billingsley (1968) shows that it is enough to prove that for any $\epsilon > 0$ and $\eta > 0$

there is $0 < \delta \leq 1$ and $n_0 \geq 0$ such that

$$\frac{1}{\delta} P \left(\sup_{s < t \leq (s+\delta) \wedge 1} \frac{W_{T_n}(T_n s) - W_{T_n}(T_n t)}{g(T_n)} > \epsilon \right) \leq \eta$$

for all $n \geq n_0$ and any $0 \leq s \leq 1$. Let us fix any $\delta \in (0, 1)$. For $0 \leq s \leq 1$ we have

$$\frac{1}{\delta} P \left(\sup_{s < t \leq (s+\delta) \wedge 1} \frac{W_{T_n}(T_n s) - W_{T_n}(T_n t)}{g(T_n)} > \epsilon \right) \leq \frac{1}{\delta} P (W_{T_n}(T_n s) > \epsilon g(T_n)) \leq \frac{\eta}{2}$$

for n large enough, since $\frac{W_{T_n}(T_n s)}{g(T_n)} \Rightarrow 0$ in probability.

Therefore, (4.13) follows, and so we are done. \square

5. THE FLUID NETWORK MODEL

In the following parts of the paper we use the results of the previous sections to describe the limiting behavior of a network of fluid queues. We start with defining our fluid network.

The network consists of N fluid queuing systems, say $\{Q_i\}_i$, in the sequel referred to as stations, each one with constant service rate C_i . The output from station Q_i is routed to the other stations in a deterministic way according to the routing rates $\{p_{ij}\}_j$, i.e. if $o_i(t)$ is the instantaneous outgoing flow from the station Q_i , the station Q_j receives instantaneous input of $p_{ij}o_i(t)$.

The rates satisfy the obvious conditions

$$(5.1) \quad p_{ij} \geq 0, \quad 1 \leq i, j \leq N \quad \text{and} \quad \sum_{j=1}^N p_{ij} \leq 1, \quad 1 \leq i \leq N.$$

Hence the matrix $P = \{p_{ij}\}$, called *routing matrix*, is a substochastic matrix.

The portion of output flow from the station Q_i to the outside of the network is given by $p_{i0} = 1 - \sum_{j=1}^N p_{ij}$. We remark that the case $p_{ii} > 0$ is not excluded and it refers to the presence of a loop around station Q_i .

Similarly to the previous sections, we use the notation

$$I_i(t) = \int_0^t i_i(s) ds \quad \text{and} \quad O_i(t) = \int_0^t o_i(s) ds,$$

to denote the cumulative input and output processes at station Q_i , while the buffer content is denoted by $W_i(t)$.

Let $\tilde{i}_i(t)$ and $\tilde{I}_i(t)$ be respectively the amount of instantaneous and cumulative input to the queuing station Q_i from the outside of the network.

The cumulative input to the queuing station Q_i can be expressed for each $t > 0$ as

$$I_i(t) = \tilde{I}_i(t) + \sum_{j=1}^N p_{ji} O_j(t)$$

and, using the fact that

$$O_j(t) = I_j(t) - W_j(t), \quad 0 < j \leq N,$$

we obtain the following system of equations

$$(5.2) \quad I_i(t) - \sum_{j=1}^N p_{ji} I_j(t) = \tilde{I}_i(t) - \sum_{j=1}^N p_{ji} W_j(t), \quad 1 \leq i \leq N.$$

Defining the row vectors $\vec{I}(t) = (I_1(t), \dots, I_N(t))$, $\tilde{\vec{I}}(t) = (\tilde{I}_1(t), \dots, \tilde{I}_N(t))$ and $\vec{W}(t) = (W_1(t), \dots, W_N(t))$, this system can be expressed in vector form as

$$\vec{I}(t)(I - P) = \tilde{\vec{I}}(t) - \vec{W}(t)P.$$

Suppose that for every $i = 1, \dots, N$, $p_{i0}^{(k)} > 0$ for some $k \geq 1$ (i.e. no flow is destined to stay in the system forever). Then the matrix $(I - P)$ is invertible (Feller (1957) XV.8). Denoting $H = (I - P)^{-1}$, the solution of the system of equations (5.2) can be written in the form

$$(5.3) \quad \vec{I}(t) = \tilde{\vec{I}}(t)H - \vec{W}(t)PH.$$

5.1. $M/G/\infty$ input processes. In this case, each external input process \tilde{I}_i , results from a $M/G/\infty$ input process.

We denote by λ_i the intensity rates of the Poisson point processes, and by $F^{(i)}$ and μ_i the distribution functions and the means of the activity periods, respectively. For the limit results we derive in the sequel we assume regular variation

$$(5.4) \quad \bar{F}^{(i)}(x) = x^{-\alpha_i} L_i(x), \quad x > 0, 1 < \alpha_i < 2,$$

where $L_i(x)$ is a slowly varying at infinity function. Note that we are not assuming that $\alpha_i, i = 1, \dots, N$ are all equal. As in the case of a single station, the assumption of the regular variation is not required for the buffer content tightness established below.

The traffic intensities at different stations ρ_i are given by

$$(5.5) \quad \rho_i = \frac{1}{C_i} \sum_{j=1}^N h_{ji} E[\tilde{I}_j(1)] = \frac{1}{C_i} \sum_{j=1}^N h_{ji} \lambda_j \mu_j,$$

where $H = (h_{ij})$ is as above.

Suppose we have a sequence of such queuing networks, indexed by T , in which some components of the arrival intensity vector $\vec{\lambda}(T)$ increase to infinity. In this case we characterize the limit behavior of the normalized flows circulating in the network.

As in the previous sections, we keep constant the values of the traffic intensities by varying the service rates of the queuing stations, i.e.

$$(5.6) \quad C_i(T) = \frac{1}{\rho_i} \sum_{j=1}^N h_{ji} \lambda_j(T) \mu_j.$$

5.2. Superposition of ON-OFF sources. In this case, the external input processes $(\tilde{I}_i, 1 \leq i \leq N)$, result from a superposition of i.i.d. stationary M_i ON-OFF sources.

We denote by μ_i^{on} and μ_i^{off} the average ON and OFF times for the external input at the station i , and use the notation $\mu_i = \mu_i^{\text{on}} + \mu_i^{\text{off}}$ and $p_i^{\text{on}} = \frac{\mu_i^{\text{on}}}{\mu_i}$. Furthermore, the limit results we will prove (but not the buffer content tightness) require the regular variation assumptions on the ON times and OFF times distribution functions

$$(5.7) \quad \bar{F}_i^{\text{on}}(x) = x^{-\alpha_i^{\text{on}}} L_i^{\text{on}}(x), \quad x > 0, 1 < \alpha_i^{\text{on}} < 2,$$

and

$$(5.8) \quad \bar{F}_i^{\text{off}}(x) = x^{-\alpha_i^{\text{off}}} L_i^{\text{off}}(x), \quad x > 0, 1 < \alpha_i^{\text{off}} < 2,$$

with $L_i^{\text{on}}(x)$, $L_i^{\text{off}}(x)$ slowly varying at infinity functions and $\alpha_i = \alpha_i^{\text{on}} < \alpha_i^{\text{off}}$. Once again, the indices $\alpha_i, i = 1 \dots, N$ do not have to be all equal.

In the same way as before, the traffic intensities are given by

$$(5.9) \quad \rho_i = \frac{1}{C_i} \sum_{j=1}^N h_{ji} E[\tilde{I}_j(1)] = \frac{1}{C_i} \sum_{j=1}^N h_{ji} M_j p_j^{\text{on}},$$

where $H = (h_{ij})$ is as above.

Once again, we consider a limiting procedure in which we have a sequence of such queuing networks, indexed by T , with some components of the source-number vector $\vec{M}(T)$ increasing to ∞ . To keep the traffic intensities constant, the service rate are assumed to be given by

$$(5.10) \quad C_i(T) = \frac{1}{\rho_i} \sum_{j=1}^N h_{ji} M_j(T) p_j^{\text{on}}.$$

6. TIGHTNESS OF THE BUFFER CONTENT FOR THE STATIONS IN A NETWORK

In this section we establish the workload tightness for a queuing network in a manner similar to that used for a single queue in Section 3.

6.1. $M/G/\infty$ input processes.

Theorem 6.1. *Suppose we have a series of fluid queuing networks, indexed by T , consisting of N stations $\{Q_i\}_i$ and with routing matrix P such that, for each $1 \leq i \leq N$, $p_{i0}^{(k)} > 0$ for some $k \geq 1$. The external input to each station Q_i is a $M/G/\infty$ process with intensity $\lambda_i(T)$. The service rates satisfy the relation (5.6), and so the traffic intensities $\rho_i < 1$ are kept constant.*

Then the stationary workload $(W_{i,T}, T \geq 0)$ at each station $1 \leq i \leq N$ is tight.

Proof. For every station i and $t \geq 0$, $O_i(t) \leq I_i(t)$. Therefore

$$(6.1) \quad I_i(t) \leq \tilde{I}_i(t) + \sum_{j=1}^N p_{ji} I_j(t), \quad 1 \leq i \leq N, t \geq 0.$$

Since the matrix $H = (I - P)^{-1} = \sum_{n=0}^{\infty} P^n$ has nonnegative entries, we conclude that

$$(6.2) \quad I_i(t) \leq \sum_{j=1}^N h_{ji} \tilde{I}_j(t), \quad 1 \leq i \leq N, t \geq 0.$$

Notice that if every $\tilde{I}_i(\cdot)$ is an $M/G/\infty$ input process, then so is $\sum_{j=1}^N h_{ji} \tilde{I}_j(\cdot)$. Hence our statement follows from Lemma 3.1. \square

6.2. Superposition of ON-OFF sources.

Theorem 6.2. *Suppose we have a series of fluid queuing networks, indexed by T , consisting of N stations $\{Q_i\}_i$ and with routing matrix P such that, for each $1 \leq i \leq N$, $p_{i0}^{(k)} > 0$ for some $k \geq 1$. The external input to each station Q_i is a superposition of $M_i(T)$ stationary ON-OFF processes. The service rates satisfy the relation (5.10), and so the traffic intensities $\rho_i < 1$ are kept constant.*

Then the stationary workload $(W_{i,T}, T \geq 0)$ at each station $1 \leq i \leq N$ is tight.

Proof. Using the same argument as for the $M/G/\infty$ case we have

$$(6.3) \quad I_i(t) \leq \sum_{j=1}^N h_{ji} \tilde{I}_j(t) \quad 1 \leq i \leq N, t \geq 0.$$

Letting $(\tilde{W}_i(t), t \geq 0)$ be the buffer content process of a queuing system fed by the process in the right side of (6.3) and with service rate C_i , we have that

$$(6.4) \quad \tilde{W}_i(t) = \sup_{0 \leq s \leq t} \left\{ \sum_{j=1}^N h_{ji} \left(\tilde{I}_j(t) - \tilde{I}_j(s) \right) - C_i(t - s) \right\}$$

$$(6.5) \quad \leq \sum_{j=1}^N h_{ji} \sup_{0 \leq s \leq t} \left\{ \tilde{I}_j(t) - \tilde{I}_j(s) - \tilde{\gamma}_i E \left[\tilde{I}_j(1) \right] (t - s) \right\} = \sum_{j=1}^N h_{ji} \tilde{W}_{ji}(t),$$

where $\tilde{\gamma}_i = \frac{C_i}{\sum_{j=1}^N h_{ji} E[\tilde{I}_j(1)]} > 1$ and where $(\tilde{W}_{ji}(t), t \geq 0)$ is the buffer occupancy process of a fluid queue fed by the input process \tilde{I}_j , with service rate $\tilde{\gamma}_i E[\tilde{I}_j(1)]$. Hence the tightness follows from Lemma 3.2. \square

7. RANDOM FIELD LIMITS FOR THE FLOWS IN FLUID NETWORKS

In order to describe the random field limit for a fluid queuing network described in the previous sections, we need to introduce some additional notation. First of all, let

$$(7.1) \quad f_i(T) = \begin{cases} \lambda_i(T), & \text{for the } M/G/\infty \text{ model} \\ M_i(T), & \text{for the ON-OFF model} \end{cases}$$

be the rate function appropriate to the input model. The quantile functions are denoted by

$$(7.2) \quad b_i(t) = \left(1/\bar{F}^{(i)}\right)^{\leftarrow}(t) \quad \text{or} \quad b_i(t) = \left(1/\bar{F}_i^{\text{on}}\right)^{\leftarrow}(t),$$

depending on the model. Let $\mathcal{N} = \{1, \dots, N\}$,

$$\mathcal{N}_s = \left\{ i \in \mathcal{N} : \lim_{T \rightarrow \infty} \frac{b_i(f_i(T)T)}{T} = 0 \right\}, \quad \text{and} \quad \mathcal{N}_f = \mathcal{N} \setminus \mathcal{N}_s.$$

As usual, we assume that $b_i(f_i(T)T)/T \rightarrow \infty$ for $i \in \mathcal{N}_f$. The sets \mathcal{N}_s and \mathcal{N}_f are the sets of stations the external input to which satisfies, correspondingly, slow growth and fast growth conditions. Let

$$(7.3) \quad d_i(T) = \begin{cases} b_i(\lambda_i(T)T) & \text{if } i \in \mathcal{N}_s \\ (\lambda_i(T)T^3 \bar{F}^{(i)}(T))^{\frac{1}{2}} & \text{if } i \in \mathcal{N}_f \end{cases}$$

for the $M/G/\infty$ model, and

$$(7.4) \quad d_i(T) = \begin{cases} b_i(M_i(T)T) & \text{if } i \in \mathcal{N}_s \\ (M_i(T)T^3 \bar{F}_i^{\text{on}}(T))^{\frac{1}{2}} & \text{if } i \in \mathcal{N}_f \end{cases}$$

for the superposition of ON-OFF models.

For $k = 1, \dots, N$ let

$$R_k = \left\{ i \in \mathcal{N} : \text{there exist } n = 0, 1, 2, \dots \text{ s.t. } p_{ik}^{(n)} > 0 \right\}$$

with the usual convention $p_{ik}^{(0)} = 1$ if $i = k$ and 0 otherwise. We let

$$R_k^s = R_k \cap \mathcal{N}_s \quad \text{and} \quad R_k^f = R_k \cap \mathcal{N}_f$$

to be the sets of the slow (fast) growth external inputs that reach the station k . Let, furthermore,

$$R_{k,+}^s = \left\{ i \in R_k^s : \lim_{T \rightarrow \infty} \frac{d_i(T)}{d_j(T)} > 0 \quad \text{for all } j \in R_k^s \right\}$$

and

$$R_{k,+}^f = \left\{ i \in R_k^f : \lim_{T \rightarrow \infty} \frac{d_i(T)}{d_j(T)} > 0 \quad \text{for all } j \in R_k^f \right\}$$

be the subsets of R_k^s and R_k^f with the fastest connection rates, $k = 1, \dots, N$. We assume that the limits in the definition of these sets exist. Define

$$(7.5) \quad g_k^{(\cdot)}(T) = \sum_{i \in R_{k,+}^{(\cdot)}} d_i(T).$$

Observe that $g_k^s(T)$ and $g_k^f(T)$ are the contributions to the normalization in the limit theorem from the external inputs satisfying slow growth and fast growth conditions respectively. We assume that for every $k = 1, \dots, N$

$$g_k^s(T) + g_k^f(T) \rightarrow \infty \quad \text{as } T \rightarrow \infty.$$

That is, every node in the network is reached by at least one external input whose intensity grows to infinity.

Finally, denote for $k = 1, \dots, N$

$$(7.6) \quad r_k^{(\cdot)}(i) = \lim_{T \rightarrow \infty} \frac{d_i(T)}{g_k^s(T) + g_k^f(T)} \quad \text{for } i \in R_{k,+}^{(\cdot)}$$

Once again, all the limits are assumed to exist.

The following two theorems describing the limiting behavior of the flow in the network follow in the same way as Theorems 6.1 and 6.2 for a single queue.

The first theorem deals with a network fed by $M/G/\infty$ external inputs.

Theorem 7.1. *Suppose we have a series of fluid queuing networks, indexed by T . The external input to each station Q_i is an $M/G/\infty$ process with intensity $\lambda_i(T)$. Let $I_{k,T}(\cdot)$ denote the cumulative input process to station k in the model corresponding to scale T . Under the above conditions*

$$(7.7) \quad \left(\frac{I_{k,T}(T \cdot) - \sum_{i=1}^N h_{ik} \mu_i T \lambda_i(T)(\cdot)}{g_k^s(T) + g_k^f(T)}, k = 1, \dots, N \right) \\ \xrightarrow{fidi} \left(\sum_{i \in R_{k,+}^s} h_{ik} r_k^s(i) X_{\alpha_i, 1, 1}(\cdot) + \sum_{i \in R_{k,+}^f} h_{ik} r_k^f(i) \sigma_i B_{H_i}(\cdot), k = 1, \dots, N \right)$$

where $\sigma_i^2 = \frac{1}{3-\alpha_i} \left[\frac{\alpha_i}{2-\alpha_i} + \frac{2}{\mu_i} \right]$, $H_i = \frac{3-\alpha_i}{2}$, and the stable Lévy processes and Fractional Brownian motions in the right hand side of (7.7) are independent.

Furthermore, let

$$\mathcal{B} = \left\{ k = 1, \dots, N : \lim_{T \rightarrow \infty} \frac{g_k^s(T)}{g_k^s(T) + g_k^f(T)} = 0 \right\}.$$

Then

$$(7.8) \quad \left(\frac{I_{k,T}(T \cdot) - \sum_{i=1}^N h_{ik} \mu_i^T \lambda_i(T)(\cdot)}{g_k^f(T)}, k \in \mathcal{B} \right) \xrightarrow{d} \left(\sum_{i \in R_{k,+}^f} h_{ik} r_k^f(i) \sigma_i B_{H_i}(\cdot), k \in \mathcal{B} \right).$$

The next theorem establishes a similar result for a network fed ON-OFF external inputs.

Theorem 7.2. *Suppose we have a series of fluid queuing networks, indexed by T . The external input to each station Q_i is a superposition of $M_i(T)$ i.i.d. stationary ON-OFF processes. Let $I_{k,T}(\cdot)$ denote the cumulative input process to station k in the model corresponding to scale T . Under the above conditions*

$$(7.9) \quad \left(\frac{I_{k,T}(T \cdot) - \sum_{i=1}^N h_{ik} \mu_i^{-1} \mu_i^{on} T M_i(T)(\cdot)}{g_k^s(T) + g_k^f(T)}, k = 1, \dots, N \right) \xrightarrow{fidi} \left(\sum_{i \in R_{k,+}^s} h_{ik} r_k^s(i) c_i X_{\alpha_i, \sigma_i, 1}(\cdot) + \sum_{i \in R_{k,+}^f} h_{ik} r_k^f(i) \sigma_{0,i} B_{H_i}(\cdot), k = 1, \dots, N \right).$$

Here $\sigma_i = C_{\alpha_i}^{-\frac{1}{\alpha_i}}$, C_{α} is as in Theorem 4.4, $c_i = \frac{\mu_i^{off}}{\mu_i^{1+1/\alpha_i}}$ and $H_i = \frac{3-\alpha_i}{2}$. Furthermore, the stable Lévy processes and Fractional Brownian motions in the right hand side of (7.9) are independent.

Finally, let \mathcal{B} be as in Theorem 7.1. Then

$$(7.10) \quad \left(\frac{I_{k,T}(T \cdot) - \sum_{i=1}^N h_{ik} \mu_i^{-1} \mu_i^{on} T M_i(T)(\cdot)}{g_k^f(T)}, k \in \mathcal{B} \right) \xrightarrow{d} \left(\sum_{i \in R_{k,+}^f} h_{ik} r_k^f(i) \sigma_{0,i} B_{H_i}(\cdot), k \in \mathcal{B} \right).$$

Notice that we have obtained in Theorems 7.1 and 7.2 random field descriptions of the limiting behavior of the network. The limiting random fields are indexed both by time and by station. The conclusions are even stronger for the part of the network subject to only Gaussian limiting fluctuations (denoted by \mathcal{B} in both theorems). Here we have full weak convergence to the limiting random field in the space of continuous functions with the values in the space whose dimension is the cardinality of \mathcal{B} .

This should allow one to try to predict certain patterns of behavior of such a network. In particular, simulation of the limiting network is possible. Additionally, usage of continuous mapping theorem in the ‘‘Gaussian’’ part of the network \mathcal{B} should allow one to get other qualitative results. We intend to pursue such studies in the future.

REFERENCES

P. BILLINGSLEY (1968): *Convergence of Probability Measures*. Wiley, New York.

- F. BRANDT and B. LISEK (1990): *Stationary Stochastic Models*. Wiley, New York.
- F. CAMPOS, J. MARRON, G. SAMORODNITSKY and F. SMITH (2002): Variable heavy tailed durations in internet traffic, Part I: Understanding heavy tails. In *Proceedings of the Tenth IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (IEEE/ACM MASCOTS 2002)*.
- D. DALEY and D. VERE-JONES (1988): *An Introduction to the Theory of Point Processes*. Springer Series in Statistics. Springer-Verlag, New York.
- A. DOWNEY (2000): The structural cause of file size distributions. Wellesley College Tech. Report CSD-TR25-2000. Available at <http://rocky.wellesley.edu/downey/filesize>.
- W. FELLER (1957): *An Introduction to Probability Theory and its Applications*, volume 1. Wiley, New York, 3rd edition.
- W. FELLER (1971): *An Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 2nd edition.
- R. GAIGALAS and I. KAJ (2003): Convergence of scaled renewal processes and a packet arrival model. *Bernoulli* 9:671–703.
- W. GONG, Y. LIU, V. MISRA and D. TOWSLEY (2001): On the tails of web file size distributions. In *Proceedings of 39-th Allerton Conference on Communication, Control, and Computing*.
- D. HEATH, S. RESNICK and G. SAMORODNITSKY (1998): Heavy tails and long range dependence in on/off processes and associated fluid models. *Mathematics of Operations Research* 23:145–165.
- T. MIKOSCH, S. RESNICK, H. ROOTZEN and A. STEGEMAN (2002): Is network traffic approximated by stable Lévy motion or Fractional Brownian motion? *Annals of Applied Probability* 12:23–68.
- G. SAMORODNITSKY (2002): *Long Range Dependence, Heavy Tails and Rare Events*. MaPhySto, Centre for Mathematical Physics and Stochastics, Aarhus. Lecture Notes.
- G. SAMORODNITSKY and M. TAQQU (1994): *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York.
- F. D. SMITH, F. HERNANDEZ, K. JEFFAY and D. OTT (2001): What TCP/IP Protocol Headers Can Tell Us About the Web. In *Proceedings of ACM SIGMETRICS 2001/Performance 2001*. Cambridge MA, pp. 245–256.

DIPARTIMENTO DI INGEGNERIA DELL' INFORMAZIONE E MATEMATICA APPLICATA, UNIVERSITY OF SALERNO,
VIA PONTE DON MELILLO 84084, FISCIANO (SA) ITALY

E-mail address: bdauria@unisa.it, dauria@diima.unisa.it

SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, CORNELL UNIVERSITY, ITHACA, NY
14853

E-mail address: gennady@orie.cornell.edu