Long-Range Dependence as a Consequence of Balanced Queueing

Ken Duffy[a], Christopher King[a,b] and David Malone[a].
(a) Communications Network Research Institute,
Dublin Institute of Technology, Rathmines,
Dublin 6W, Ireland.
(b) Department of Mathematics, Northeastern University,
Boston, MA 02115.

October 2002

Abstract

Consider a queue with infinite waiting space at which the mean arrival-rate is equal to the mean service-rate. We call this system a balanced queue. Using a mathematical model, this paper proves that the departure process from such a system can be Long-Range Dependent (LRD) when the input process is Short-Range Dependent (SRD). Furthermore, the departure process does not fall within the class of LRD processes normally considered, and its power-spectrum is investigated using non-standard techniques.

Simulations and experiments are described which demonstrate that the induced LRD is not a peculiarity of the model chosen, and can occur in real networks. In particular, evidence is presented to show that TCP may itself cause balanced queues. This finding is consistent with the observation that one of the primary goals of TCP is to obtain maximum network throughput without causing over-load, thereby leading to balanced queues throughout a network.

1 Introduction

Long-range dependence (LRD) in network traffic has been observed and documented over the last 10 years (for example, see [1, 2, 3, 4, 5]). Two important

[1]Work supported by Science Foundation Ireland under the National Development Plan.

effects produced by LRD traffic are: (i) less benefits arise from economies of scale [6]; (ii) the queue-length distribution at a buffered device admitting such traffic has a thicker tail (for example, see [7, 8, 9, 10, 11]). Models using SRD traffic-sources have been used to dimension networks. If traffic is LRD, it is possible these models will result in under-provisioning of network resources.

The Hurst parameter, introduced originally in hydrology [12], is a common measure of long-range dependence. For a stochastic process, a Hurst parameter of greater than 0.5 indicates that its power-spectrum diverges at the origin, and signals the presence of LRD. We will take power-spectrum divergence at the origin as our working definition of long-range dependence, and we will measure it using the wavelet-based LRD estimator of Abry et. al. [13, 14].

Estimation of the Hurst parameter for network traffic-traces has led to values greater than 0.5. The range of values determined by these estimations does not vary greatly, even though the traffic analyzed comes from different sources over a ten year period. It seems likely that the nature of network traffic has changed during this period. For example, it is known that http traffic is the most common on today's Internet (see [15]), but the World Wide Web did not exist when the famous BellCore traces, [16], were recorded.

Consistent Hurst parameter estimation across such a large range of traffic suggests a common cause. We will search for a possible explanation by examining the behavior of queues within real networks. As a model for this behavior, we consider the departure process from a single device with infinite buffer and constant service-rate, under the assumption that the mean arrival-rate of traffic is equal to the device's service-rate. We call this system a *balanced queue*. We prove that the departure process from a balanced queue has LRD, and we compute the strength of this LRD. As a by-product, we show that the Hurst parameter fails to fully describe the nature of LRD. We then describe simulations which show how TCP can create these balanced queues in networks, and consequently how TCP can produce LRD-type behavior. Estimations of the strength of this LRD are consistent with the mathematical results from the balanced queue model.

The first mathematical result concerns the presence of LRD in the departure process. With a SRD ON/OFF arrival process, Theorem 1 generalizes a result for simple random walks; it proves that the ON times of the departure process from a balanced queue are distributed such that $P(\mathrm{ON} > x) \sim x^{-0.5}$. This indicates that Pareto $x^{-\alpha}$ burst-time distributions with $0 < \alpha < 1$ are important in the study of network traffic. Such distributions are more troublesome that those with $1 < \alpha < 2$, as they have infinite mean.

The second mathematical result concerns the strength of LRD processes.

The power-spectrum for an ON/OFF process with ON times distributed as $x^{-\alpha}$ for $0 < \alpha < 1$ is not defined. We introduce a regularized version of the power-spectrum and determine its divergence at the origin for such processes; Theorem 2 relates this divergence to the Hurst parameter. For $1 < \alpha < 2$, the relationship $H = (3-\alpha)/2$ is deduced in Heath et. al. [17]; for $0 < \alpha < 1$, Theorem 2 proves that $H = (1+\alpha)/2$. Hence, the Hurst parameter does not differentiate these processes.

The rest of this paper is organized as follows: in Section 2, balanced queues are considered and Theorem 1 stated; in Section 3, the power-spectrum regularization necessary to deal with the departures from a balanced queue is introduced and Theorem 2 is stated; simulation results verifying the theorems are presented in Section 4; experimental results using TCP and UDP traffic are presented in Section 5; concluding remarks appear in Section 6; all proofs are presented in Appendix A.

2 Mathematical Model for a Balanced Queue

The balanced queue is analyzed using a continuous-time fluid model. In this model, a single server is fed by an ON/OFF arrival process that carries traffic at constant rate, r, during ON periods and at rate 0 during OFF periods. The durations of ON periods, $\{A_n\}$, are assumed to be drawn independently from a distribution with finite mean, $\mathbf{E}[A]$, and variance. The durations of the OFF periods are i.i.d. with exponential distribution of rate λ.

The server has an infinite buffer and serves at constant rate, s, whenever its buffer is non-empty. It is assumed that $r > s$, so that the server is idle only when both the buffer is empty and the arrival-rate is zero. This ensures that the departure process forms an alternating sequence of rate-s ON periods and rate-0 OFF periods. Furthermore, as the OFF periods of the arrival process are exponentially distributed, the ON and OFF periods of the departure process are independent and the OFF periods are i.i.d. with rate λ. The goal is to analyse the distribution of the ON periods of this departure process.

This system has been analyzed in detail by other authors. A thorough review can be found in Boxma and Dumas [18]. We use results from [18] in our analysis. They define a parameter, ρ, called the traffic intensity,

$$\rho := \frac{r\lambda\mathbf{E}[A]}{s(1+\lambda\mathbf{E}[A])},$$

which is the ratio of the arrival rate of traffic to the maximum possible service rate. The condition for stability is $\rho \leq 1$. We define the *balanced queue* by the condition $\rho = 1$, which is the borderline case for stability.

Theorem 1 is a result regarding the output process from a balanced queue. It is assumed that the arrival process starts at time $t = 0$ with an ON period, hence the departure process is not stationary. Indeed, in some cases the departure process cannot converge to a stationary distribution.

Theorem 1 *Let $\{B_1, B_2, \ldots\}$ be the sequence of ON period durations of the departure process from the model described above. Assuming $\rho = 1$, the durations $\{B_n\}$ are i.i.d. and their distribution, B, satisfies*

$$\mathbf{P}(B > x) \sim x^{-1/2} \qquad as \quad x \to \infty.$$

Theorem 1 shows that the output from the balanced queue fed with a SRD source has heavy-tailed ON periods with exponent 1/2.

3 The regularized power-spectrum

The power-spectrum, $S(\theta)$, of a wide-sense stationary process, $\{X(t)\}$, is the Fourier transform of its autocorrelation function $R_{XX}(t) = \mathbf{E}[X(s+t)X(s)]$; that is

$$S(\theta) := \int_{-\infty}^{\infty} e^{i\theta t} R_{XX}(t) dt. \tag{1}$$

Unless $\{X(t)\}$ has mean zero, the expression (1) has an impulse singularity at $\theta = 0$. For this reason it is often preferable to work instead with the Fourier transform of the auto-covariance function $C_{XX}(t) = \mathbf{E}[X(s+t)X(s)] - (\mathbf{E}[X])^2$, which agrees with (1) for $\theta \neq 0$, and is usually continuous at $\theta = 0$.

However, if the process $\{X(t)\}$ has LRD, then both $R_{XX}(t)$ and $C_{XX}(t)$ are infinite, and (1) cannot be used to compute the power spectrum. In particular this is the case for the ON/OFF process that describes the output from the balanced queue in Theorem 1. We resolve this difficulty by defining a regularized Fourier transform, $\tilde{X}_\epsilon(\theta)$, of $X(t)$. We use the expectation of $|\tilde{X}_\epsilon(\theta)|^2$ as an approximation to the power-spectrum (this is a standard methodology, for example, see Section 7.1 of [19]). Assuming that $X(t)$ is uniformly bounded, the following integral exists for any $\epsilon > 0$:

$$\tilde{X}_\epsilon(\theta) = \int_0^{\infty} X(t) e^{(i\theta - \epsilon)t} dt. \tag{2}$$

If $X(t)$ is Wide-Sense Stationary (WSS), then the usual power-spectrum, $S(\theta)$ defined in Equation (1), is recovered from (2) by taking the limit

$$S(\theta) = \lim_{\epsilon \to 0} \epsilon \, \mathbf{E}\Big[|\tilde{X}_\epsilon(\theta)|^2\Big]. \tag{3}$$

(Technically, the convergence requires that $X(t)$ be replaced by $X(t) - \mathbf{E}[X]$, in which case the right side of (3) converges to the Fourier transform of the auto-covariance function; as remarked above this distinction is irrelevant for $\theta \neq 0$).

Theorem 2 shows that Equation (2) enables the definition of a regularized power-spectrum for a large class of heavy-tailed ON/OFF processes. The idea is to determine the behavior of $\mathbf{E}\left[|\tilde{X}_\epsilon(\theta)|^2\right]$, as $\epsilon \to 0$, and extract its leading order part by re-scaling by ϵ^ν for some $\nu > 0$. This leading order part, $S^{\mathrm{reg}}(\theta)$, is defined to be the regularized power-spectrum of the process for $\theta \neq 0$, and as Equation (3) shows, this procedure yields the correct power-spectrum for a WSS process (with exponent $\nu = 1$).

Theorem 2 applies to ON/OFF processes for which the OFF periods have finite mean and variance. The ON periods may be heavy-tailed, as occurred for the output of the balanced queue. The ON/OFF process, $X(t)$, is defined for $t \geq 0$ as follows:

$$X(t) = \begin{cases} 1 & \text{during an ON period;} \\ 0 & \text{during an OFF period.} \end{cases} \tag{4}$$

Without loss of generality, the traffic rate is normalized to be 1 during an ON period. There is no assumption of stationarity for the process, which always starts with an ON period at $t = 0$.

For the purposes of Theorem 2, it is convenient to describe the heavy-tailed distribution of the ON periods by specifying the asymptotic behavior of their density function. The duration of an ON period is denoted B and its density denoted f_B. Hence, for example,

$$P(B > t) = \int_t^\infty f_B(u)du. \tag{5}$$

For some $\alpha > 0$, f_B is assumed to satisfy

$$\lim_{u\to\infty} u^{\alpha+1} f_B(u) = c > 0 \tag{6}$$

or, more compactly, $f_B(u) \sim u^{-\alpha-1}$. Using (5), this implies that $P(B > t) \sim t^{-\alpha}$.

For $\alpha > 2$, both the mean and variance are finite, thus B is short range. For $1 < \alpha \leq 2$, the mean is finite but the variance is infinite. For $0 < \alpha \leq 1$ both the mean and variance are infinite.

Theorem 2 *Let $X(t)$ be the ON/OFF process defined in (4) and define ν by:*

$$\nu := \begin{cases} 1 & \textit{for } \alpha > 1; \\ \alpha & \textit{for } 0 < \alpha < 1. \end{cases}$$

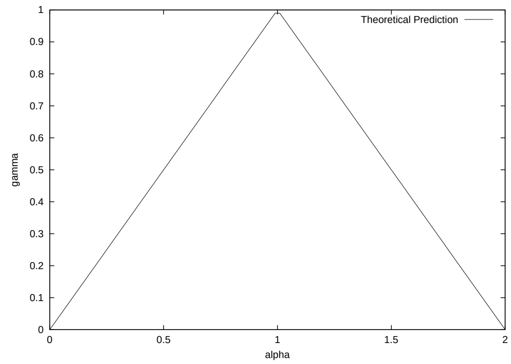

Figure 1: Theoretical prediction for the power-spectrum divergence, γ, against ON period power-tail strength, α.

Then, for all $\theta \neq 0$, the following limit exists and is not identically zero:

$$S^{\text{reg}}(\theta) = \lim_{\epsilon \to 0} \epsilon^{\nu} \ \mathbf{E}\Big[|\tilde{X}_{\epsilon}(\theta)|^2\Big]. \tag{7}$$

Furthermore, the following asymptotic behavior holds:

$$S^{\text{reg}}(\theta) \sim \theta^{\gamma} \quad \text{as } \theta \to 0, \tag{8}$$

where

$$\gamma := \begin{cases} 0 & \text{for } \alpha > 2, \\ \alpha - 2 & \text{for } 1 < \alpha < 2, \\ -\alpha & \text{for } 0 < \alpha < 1. \end{cases} \tag{9}$$

The exponent, γ, describes the long range dependence of $\{X(t)\}$. It is related to the Hurst parameter by $H = (1-\gamma)/2$. A graph of γ vs. α is shown in Figure 1. The shape of this graph for $\alpha > 1$ can be deduced from Theorem 4.3 of [17] and Theorem 1.7.1 of [20]. One of the main contributions of this paper is to determine the relationship for $0 < \alpha < 1$. This demonstrates that the "same" LRD may arise from different heavy-tailed distributions.

4 Simulation

The purpose of the simulations in this section is to answer the following question: if the systems described in Theorems 1 and 2 are simulated and the divergence of the power-spectrum estimated, is there agreement with the theoretical prediction? As it is the divergence of the power-spectrum that

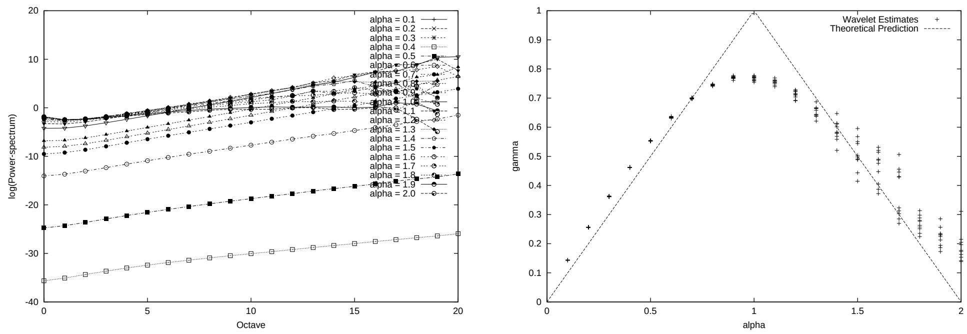

Figure 2: Wavelet estimator for $0 < \alpha < 2$.

we estimate, we first verify Equation (9) of Theorem 2 in Section 4.1, before verifying Theorem 1 in Section 4.2.

Many methodologies (see Beran's book [21]) can be employed to determine the "strength" of a time-series' long-range dependence. For a comparison of a selection of those invented prior to 1995, see [22]. One of the most successful approaches, and the one that we use, is described by Abry and Veicht [14].

Essentially, it consists of a wavelet-based estimator for the power-spectrum defined in Equation (1). This enables us to determine the divergence, γ, of $S(\theta)$ at the origin by the following procedure: for a range of values of n, plot $\log(S(1/2^n))$ against n; if the graph is of constant slope going to $-\infty$, then the time-series is constant; if the graph is of positive slope, then the value of the slope is the estimate of γ; if the graph raises steeply and then levels to a flat line, the time-series is short-range dependent.

4.1 Wavelet Estimates for $0 < \alpha < 2$

To test Theorem 2, a variety of ON/OFF sources of the form described in Equation (4) were simulated. These had exponentially distributed OFF periods, and ON periods distributed as $P(B > x) \sim x^{-\alpha}$, for a range of values of $\alpha \in [0, 2]$.

Veicht has made available free `matlab` code for a wavelet-based power-spectrum estimator, [23]. However, we do not use it in this section as the infinite expected burst length associated with $0 < \alpha \leq 1$ makes a simple implementation impractical. Instead, an implementation was designed to exploit the known response of the wavelet estimator to long, constant bursts. C-code for the burst-based estimator can be found online [24].

In the simulated source, the probability an OFF period is longer than l is 2^{l-1}. ON periods last for $u^{-1/\alpha}$, where u is uniformly distributed in $[0, 1]$.

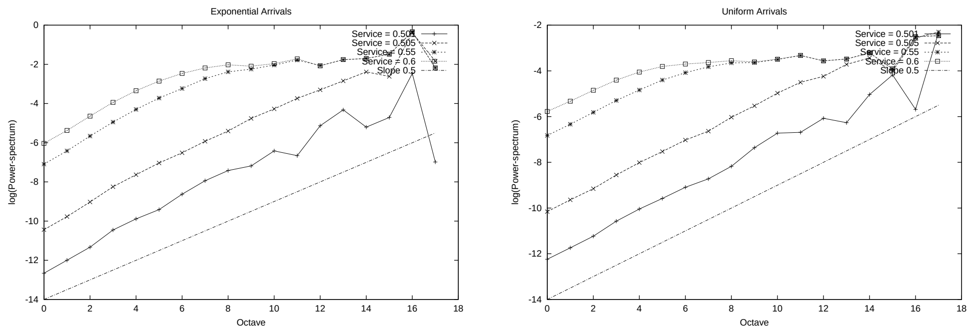

Figure 3: Wavelet log-log plots for exponential and uniform i.i.d. arrivals.

Each source comprises 1 000 000 ON/OFF cycles.

The graph on the left of Figure 2 consists of log-log plots produced by the wavelet estimator, for a range of values of α. As the graphs form straight lines, the traffic exhibits long-range dependence. The divergence of the power-spectrum at the origin is measured by the slope of these curves.

The graph on the right of Figure 2 shows least-squares fitted slopes, fit between octaves 3 and 15, for different values of α. 10 sources with different seeds are presented for each α. Theorem 2 predicts the hat function, which is shown for comparison.

4.2 Departures from a balanced queue

To verify Theorem 1, a slotted time queue with i.i.d. arrivals is simulated. In all experiments, the mean-rate of arrival is 0.5 units per clock-tick. At each clock-tick, a random amount of work is placed in the queue, taken from a given distribution. The service-rates considered are: 0.6; 0.55; 0.505 and 0.501. The departing process is analyzed using the standard wavelet-based power-spectrum estimator.

Log-log plots of the estimated power-spectrum for the queue output are shown in Figure 3. In the plot on the left, the volume of traffic arriving at each clock-tick is taken from an exponential distribution. In the plot on the right, the volume is uniformly distributed. A reference-line with slope 0.5 is also shown.

When the service-rate is above the mean arrival-rate, the graphs show initial curvature and then level out, indicating short-range dependence. As the service-rate moves towards the mean arrival-rate, the queue moves into balance and the slope of the graphs move towards 0.5, indicating long-range dependence as predicted in Theorem 1.

The proof of Theorem 1 suggests that finite variance of the arrival process

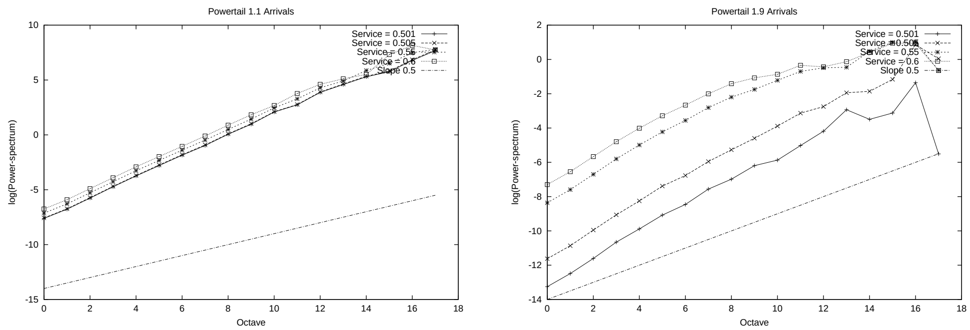

Figure 4: Wavelet log-log plots for power-tail i.i.d. arrivals.

may be necessary to deduce $\gamma = 0.5$ for the departure process. Simulation of queues fed with power-tail distributed arrivals per clock-tick, shown in Figure 4, support this notion. In some cases, the departure process does appear to exhibit long-range dependence, but does so with $\gamma \neq 0.5$.

5 Experiments

Sections 2, 3 and 4 demonstrate that in theory and simulation, balanced queues cause long-range dependent features in the queue's departure process. In this section, experiments are reported that use real TCP and UDP implementations. In particular, as one of the primary goals of TCP is to obtain maximum network throughput without causing over-load, its feedback mechanisms may produce a balanced queue at the bottle-neck link along its path.

5.1 Experimental setup

The setup consists of two computers running FreeBSD, connected with 100Mbps Ethernet. To introduce a bottle-neck link of variable size and buffer-space into the system, Rizzo's `dummynet` [25] is used. `dummynet` simulates the effects of finite queues, bandwidth limitations and communication delays. It works by intercepting communications via firewall rules that redirect packets through the simulated system, prior to their transmission on the physical interface. A minor change to `dummynet` was made to expand its range of valid buffer-sizes.

A mix of clients, in proportions similar to that observed in real traffic, is used. 2 clients produce UDP traffic and 6 produce TCP traffic. All clients have a mean-rate of 0.5Mbps. UDP clients wait for an exponentially distributed amount of time (mean 20ms) and then transmit an exponentially distributed amount of data in UDP datagrams, with a payload of up to 1KB. Once the data is transfered, the UDP client waits again.

The TCP clients are more complex. At exponentially distributed times (mean 200ms), a new TCP connection is opened to transfer an exponentially distributed amount of data. In contrast to the UDP client, new connections can be opened before the transfer is completed (up to a maximum concurrency of 4 per client).

In each experiment, the mean-rate of the traffic is 4Mbps plus protocol overhead. The traffic is fed through `dummynet` pipes of bandwidth 2, 3, 3.5, 4, 4.1, 4.2, 4.3, 4.5, 5 and 8Mbps, to gauge behavior when the link is overloaded, close to balance and under-loaded.

The use of an infinite buffer in Theorem 1 is fundamental to the proof. As infinite buffers do not occur in networks, the importance of this approximation is investigated by setting the `dummynet` packet-buffer to 1000, 100 and 10 packets for each bandwidth. An alteration to `dummynet` is necessary to allow large packet buffers. Each configuration was run for 1 hour.

Traffic is recorded with `tcpdump` after it traverses the `dummynet` pipe and 100Mbps Ethernet. Sequences of arrival volumes at 10ms intervals are recorded for TCP, UDP and their aggregate. These are analyzed with the standard wavelet estimator described in Section 4.

5.2 Experimental Results

The results for 1000 packet and 100 packet buffers are similar, so only the results for 1000 and 10 are included. Figure 5 presents power-spectrum graphs for the departure processes of these experiments. Results for 1000 packet buffer are described first.

For over-loaded links (2, 3, 3.5Mbps) the TCP and UDP plots are flat, indicating little long-range correlation. The aggregate plot is diverging toward $-\infty$, indicating a constant departure stream. The TCP plot and UDP plot show similar power-spectrum structure. Perhaps this is not surprising as they are two processes that sum to a constant.

For a range of bandwidths around balance (4–4.5), the aggregate traffic shows a strong linear response in octaves 2 to 13. As the bandwidth increases, the TCP and UDP plots separate. The UDP plot becomes flat, exhibiting short-range dependence. The TCP plot, however, shows an increasing linear section similar to the aggregate trace.

For under-loaded links (5 and 8Mbps) the UDP plots exhibit short-range dependent characteristics. The linear response seen in both the TCP and aggregate plots begins to flatten out. At 5Mbps, this occurs by the 10th

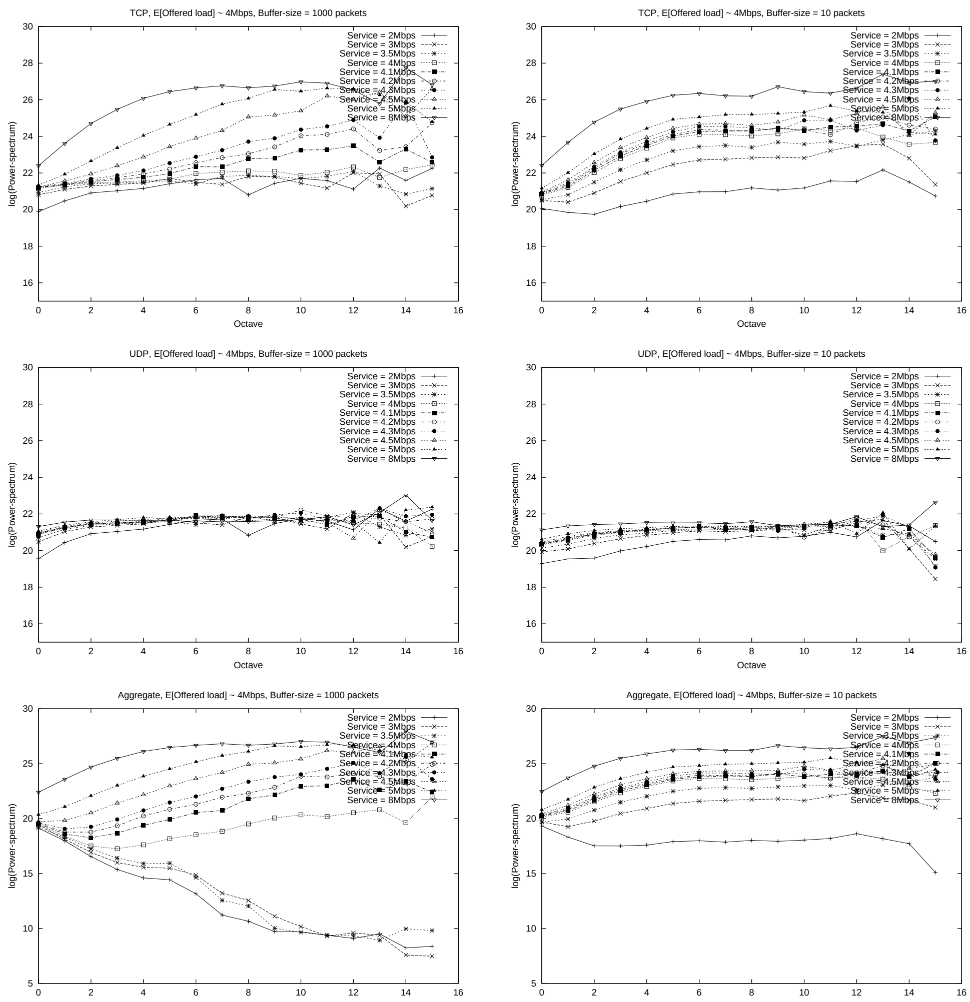

Figure 5: Experimental wavelet log-log plots for a mix of TCP and UDP traffic. Plots on the left are for a 1000 packet buffer and on the right for a 10 packet buffer.

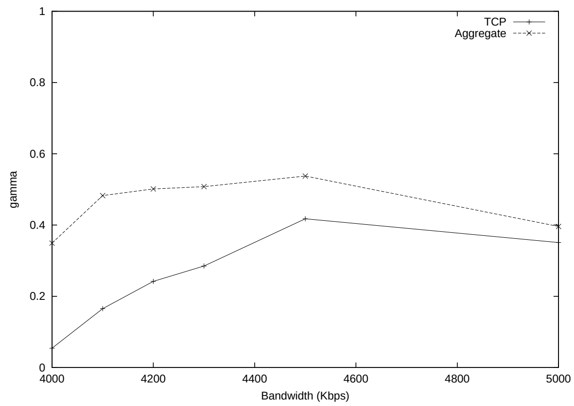

Figure 6: Slope estimates against bandwidth for the departures from a 1000 packet buffer.

octave; at 8Mbps, this occurs by the 6th octave. This implies short-range dependence, but over a longer time-scale than the UDP traffic.

For bandwidths close to balance, Figure 6 shows slope estimates over octaves 2 to 13 for TCP and the aggregate. The aggregate exhibits values of γ between 0.35 and 0.5, corresponding to values of the Hurst parameter between 0.675 and 0.75.

With a 10 packet buffer, short-range dependent behavior is observed when the link is under-loaded. Around balance, the linear response is over a smaller number of octaves, indicating short-range dependence. The buffer-space is not large enough to induce long-range correlation structure. As there is little buffering space, the TCP traffic backs off, so that the departure process is not constant until the system is totally over-loaded with a 2Mbps pipe.

In summary: for packet buffers of size 100 or 1000, in a queue near balance, the departure process exhibits long-range dependence; when the packet buffer is small, no such effect is observed.

6 Conclusions

This paper proves that the departure process from a balanced queue can exhibit long-range dependence when the queue is fed with short-range dependent traffic. Moreover, the relationship between the divergence of the power-spectrum for this "new" class of processes is determined and shown to be indistinguishable from a well known class of processes.

That these effects are not an oddity of the model employed is demonstrated through simulation and experiment. In particular, we investigate the appearance of LRD in queues which are servicing TCP traffic. Our results are consistent with the notion that TCP traffic is elastic, and tries to make its bottle-neck link a balanced queue.

Queueing occurs throughout networks: at end-host socket buffers; hardware buffers on network cards; within routers. Packetisation also behaves like queueing and sequences of packet-sizes have been shown to exhibit long-range dependence [14]. That long-range dependence arises as a consequence of queueing within networks may explain the persistent observation of LRD in traffic-traces, despite changes in technology, protocols and user-behavior.

A Proofs

Proof of Theorem 1: As well as results from [18], the proof uses the following Tauberian theorem due to Karamata:

Lemma 1 (Bingham et. al. [20], Theorem 1.7.1) *Let $F(x) = \mathbf{P}(X \leq x)$ be the distribution of a positive random variable X, and let $\phi(t) = \mathbf{E}[e^{-tX}]$ be its Laplace transform. Then each of the following statements implies the other:*

$$1 - \phi(t) \sim t^{1-a} L\left(\frac{1}{t}\right) \quad \text{as } t \to 0;$$

$$1 - F(x) \sim \frac{1}{\Gamma(a)} x^{a-1} L(x) \quad \text{as } x \to \infty;$$

where $a \geq 0$ and $L(x)$ is a slowly varying function. That is, a measurable function with the asymptotic behavior

$$\frac{L(tx)}{L(t)} \to 1 \quad \text{as} \quad t \to \infty, \qquad \text{for every fixed } x.$$

Recall that A and B are the durations of ON periods of the arrival and departure processes respectively. Define their Laplace transforms

$$\gamma(u) = \mathbf{E}[e^{-uA}] \quad \text{and} \quad \pi(u) = \mathbf{E}[e^{-uB}].$$

Let $r' = r/s$, then the following formula is derived in Theorem 3.6 of [18] (though in different notation):

$$\pi(u) = \gamma\Big(r'u + \lambda(r' - 1)(1 - \pi(u))\Big). \tag{10}$$

Its deviation is based on the observation that a busy period, B, of the output process can be viewed as the following succession of events: first an ON period, A, during which an amount of traffic α ($= rA$) arrives; then an idle period of length, T_1 (during which the queue partly empties), at the end of which there remains an amount, β_1, in the queue; then the start of another arrival period that adds more traffic to the buffer. The total time until the amount of traffic in the buffer drops again to β_1 is denoted B_1. This is followed by another period, T_2, during which traffic leaves the buffer, until another arrival period begins, at which point the buffer contains the amount β_2. The time until the buffer again reaches β_2 is denoted B_2. This continues until the buffer empties, which signals the end of the busy period B.

This leads to the representation

$$B = A + T_1 + B_1 + T_2 + B_2 + \cdots + T_K.$$

The key observations now are: first, $\sum T_i = \alpha/s = r'A$, since this is the time needed to empty the traffic which arrived during the first ON period, A; second, each interval, B_i, is distributed as an ON period of the departure process, and these periods are all independent; third, the number of departure periods, K, has the Poisson distribution with rate $\lambda r'A$. Equation (10) now follows easily.

Defining

$$\kappa = \mathbf{E}[A]\sqrt{\frac{2(\mathbf{E}[A] + 1/\lambda)}{\mathbf{E}[A^2]}},$$

a straightforward calculation using (10) shows that

$$\lim_{u \to 0+} \left(\frac{1 - \pi(u)}{\sqrt{u}} \right) = \kappa.$$

Applying Lemma 1 with $a = 1/2$ and $L = \kappa$ gives the result:

$$\lim_{x \to \infty} x^{1/2} \mathrm{P}(B > x) = \frac{2\kappa}{\sqrt{\pi}}.$$

QED

Proof of Theorem 2: the ON/OFF process is simple enough to allow an explicit calculation of its regularized Fourier transform defined in Equation (2). For $k = 1, 2, \ldots$, let B_k denote the duration of the k^{th} busy period. Suppose that the k^{th} period starts at time σ_k and ends at time τ_k. From the definition (4), it follows that $\sigma_1 = 0$. Defining

$$z = i\theta - \epsilon, \tag{11}$$

it follows that

$$\tilde{X}_\epsilon(\theta) = \frac{1}{z}\sum_{k=1}^{\infty}\left(e^{z\tau_k} - e^{z\sigma_k}\right).$$

Hence, the regularized power is computed using

$$\mathbf{E}\left[|\tilde{X}_\epsilon(\theta)|^2\right] = \frac{1}{|z|^2}\sum_{k,l=1}^{\infty}\mathbf{E}\left[(e^{\overline{z}\tau_k} - e^{\overline{z}\sigma_k})(e^{z\tau_l} - e^{z\sigma_l})\right]. \tag{12}$$

As busy and idle periods of the process are independent, this expectation can be computed in terms of the following generating functions for a typical ON period B and OFF period C:

$$\phi(w) = \mathbf{E}[e^{wB}] \quad \text{and} \quad \psi(w) = \mathbf{E}[e^{wC}],$$

where these are well-defined when w has a non-positive real part. Recalling Equation (11), the result is

$$\begin{aligned}\mathbf{E}\left[|\tilde{X}_\epsilon(\theta)|^2\right] \;=\; & \frac{\phi(z+\overline{z}) - \phi(z) - \phi(\overline{z}) + 1}{|z|^2(1-\phi(z+\overline{z})\psi(z+\overline{z}))} \\ & + \frac{(\phi(\overline{z})-1)(\phi(z+\overline{z}) - \phi(\overline{z}))\psi(\overline{z})}{|z|^2(1-\phi(\overline{z})\psi(\overline{z}))(1-\phi(z+\overline{z})\psi(z+\overline{z}))} \\ & + \frac{(\phi(z)-1)(\phi(z+\overline{z}) - \phi(z))\psi(z)}{|z|^2(1-\phi(z)\psi(z))(1-\phi(z+\overline{z})\psi(z+\overline{z}))}.\end{aligned} \tag{13}$$

In order to compute $S^{\text{reg}}(\theta)$, it is necessary to extract the leading order part from Equation (13) as $\epsilon \to 0$. For $\theta \neq 0$, the only singularities on the right side of (13) arise from the term $(1 - \phi(z+\overline{z})\psi(z+\overline{z}))$ in the denominators, since $z \to i\theta$, as $\epsilon \to 0$, implies means that $z + \overline{z} \to 0$.

The behavior of this term depends on the tail behavior of the ON period, B. If $\alpha > 1$ in Equation (6), then $\mathbf{E}[B] < \infty$, which implies that

$$1 - \phi(z+\overline{z})\psi(z+\overline{z}) \sim 2\epsilon(\mathbf{E}[B] + \mathbf{E}[C]).$$

Therefore, to leading order in this case, (13) behaves as

$$\begin{aligned}\mathbf{E}\left[|\tilde{X}_\epsilon(\theta)|^2\right] \;\sim\; & \frac{1}{2\theta^2\epsilon(\mathbf{E}[B]+\mathbf{E}[C])}\left[2 - \phi(i\theta) - \phi(-i\theta)\right. \\ & + \frac{(\phi(-i\theta)-1)(1-\phi(-i\theta))\psi(-i\theta)}{(1-\phi(-i\theta)\psi(-i\theta))} \\ & \left. + \frac{(\phi(i\theta)-1)(1-\phi(i\theta))\psi(i\theta)}{(1-\phi(i\theta)\psi(i\theta))}\right].\end{aligned} \tag{14}$$

The right side diverges as ϵ^{-1}; this establishes (7) for $\alpha > 1$.

For $\alpha \leq 1$, the mean of B is infinite and more care is needed. Defining g by

$$\mathbf{E}[e^{-2\epsilon B}] = 1 - g(\epsilon),$$

the leading order part of (13) becomes

$$\begin{aligned}\mathbf{E}\Big[|\tilde{X}_\epsilon(\theta)|^2\Big] \quad \sim \quad & \frac{1}{\theta^2(2\epsilon\mathbf{E}[C] + g(\epsilon))}\Bigg[2 - \phi(i\theta) - \phi(-i\theta) \\ & + \frac{(\phi(-i\theta) - 1)(1 - \phi(-i\theta))\psi(-i\theta)}{(1 - \phi(-i\theta)\psi(-i\theta))} \\ & + \frac{(\phi(i\theta) - 1)(1 - \phi(i\theta))\psi(i\theta)}{(1 - \phi(i\theta)\psi(i\theta))}\Bigg].\end{aligned} \qquad (15)$$

Applying Lemma 1 with $\alpha \leq 1$ gives $g(\epsilon) \sim \epsilon^\alpha$ as $\epsilon \to 0$. Hence, $2\epsilon\mathbf{E}[C] + g(\epsilon) \sim \epsilon^\alpha$ and, thus, Equation (7) follows for this case.

The behavior of $S^{\text{reg}}(\theta)$ as $\theta \to 0$ is determined in a similar way. If $\alpha > 2$, both the ON and OFF periods have finite mean and variance. Evaluating the expression (14) at $\theta = 0$ yields a finite constant, establishing (8) for this case.

For $1 < \alpha < 2$, defining

$$h(\theta) := \phi(i\theta) - 1 - i\theta\mathbf{E}[B] = \mathbf{E}[e^{i\theta B} - 1 - i\theta B],$$

a similar analysis shows that the leading order part of $S^{\text{reg}}(\theta)$ as $\theta \to 0$ goes as $\text{Re}(h(\theta))/\theta^2$. To determine the asymptotic behavior of $h(\theta)$ as $\theta \to 0$, write

$$\begin{aligned} h(\theta) &= \int_0^\infty (e^{i\theta x} - 1 - i\theta x) f_B(x)\,dx \\ &= \int_0^\infty \left(\frac{e^{i\theta x} - 1 - i\theta x}{x^{\alpha+1}}\right) x^{\alpha+1} f_B(x)\,dx \\ &= \theta^\alpha \int_0^\infty \left(\frac{e^{iu} - 1 - iu}{u^{\alpha+1}}\right)\left(\frac{u}{\theta}\right)^{\alpha+1} f_B\left(\frac{u}{\theta}\right) du. \end{aligned}$$

Since, for $1 < \alpha < 2$, the function $(e^{iu} - 1 - iu)/u^{\alpha+1}$ is integrable on $[0, \infty)$, the dominated convergence theorem and (6) imply

$$\lim_{\theta \to 0} \theta^{-\alpha}\, h(\theta) = c \int_0^\infty \left(\frac{e^{iu} - 1 - iu}{u^{\alpha+1}}\right) du,$$

establishing (8) for $1 < \alpha < 2$.

For the remaining case, $0 < \alpha < 1$, define

$$q(\theta) = 1 - \phi(i\theta), \quad r(\theta) = 1 - \psi(i\theta) \quad \text{and} \quad k = \lim_{\epsilon \to 0} \epsilon^{-\alpha} g(\epsilon).$$

Renormalising and rewriting Equation (15) gives

$$\frac{1}{\theta^2 k} \mathrm{Re}\Big[\frac{q(\theta) r(\theta)}{r(\theta) + (1 - r(\theta)) q(\theta)}\Big] \qquad \text{as } \epsilon \to 0. \tag{16}$$

Since, as $\theta \to 0$, $q(\theta) \sim \theta^{\alpha}$ and $r(\theta) \sim i\theta$, it follows that the leading order part of (16) is

$$\frac{1}{\theta^2 k} \mathrm{Re}\Big[- \frac{r(\theta)^2}{q(\theta)}\Big],$$

which establishes (8) for $0 < \alpha < 1$.
QED

References

[1] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. On the self-similar nature of Ethernet traffic. In Deepinder P. Sidhu, editor, *ACM SIGCOMM*, pages 183–193, San Francisco, California, 1993.

[2] Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V . Wilson. On the self-similar nature of Ethernet traffic [extended version]. *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.

[3] Vern Paxson and Sally Floyd. Wide-area traffic: The failure of poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[4] Mark E. Crovella and Azer Bestavros. Self-similarity in world wide web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.

[5] Walter Willinger, Murad Taqqu, Robert Sherman, and Daniel Wilson. Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Transactions on Networking*, 5(1):71–86, 1997.

[6] N. G. Duffield. Economies of scale in queues with sources having power-law large deviation scalings. *Journal of Applied Probability*, 33:840–857, 1996.

[7] I. Norros. A storage model with self-simmilar input. *Queueing systems*, 16:387–396, 1994.

[8] N. G. Duffield and N. O'Connell. Large deviations and overflow probabilities for the general single server queue, with applications. *Mathematical Proceedings of the Cambridge Philosophical Society*, 118(2):363–374, 1995.

[9] Z. Liu, P. Nain, D. Towsley, and Z. L. Zhang. Asymptotic behavior of a multiplexer fed with a long-range dependent process. *Journal of Applied Probability*, 36(1):105–118, 1999.

[10] L. Massoulie and A. Simonian. Large buffer asymptotics for the queue fed with FBM input. *Journal of Applied Probability*, 36(3):894–906, 1999.

[11] Ken Duffy, John T. Lewis, and Wayne G. Sullivan. Logarithmic asymptotics for the supremum of a stochastic process. *To appear in Annals of Applied Probability*, 2003.

[12] H. E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–808, 1951.

[13] Patrice Abry, Paulo Gonçalvès, and Partick Flandrin. *Wavelets, spectrum analysis and 1/f processes*, volume 103 of *Lecture notes in Statistics*, pages 15–30. Springer Verlag, 1995.

[14] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2–15, 1998.

[15] K. Claffy, G. Miller, and K. Thompson. The nature of the beast: recent traffic measurements from an internet backbone. In *INET 98*, Washington, DC, 1998.

[16] Bellcore. Ip traces. Available: `http://ita.ee.lbl.gov/html/contrib/BC.html`.

[17] D. Heath, S. I. Resnick, and G. Samorodnitsky. Heavy tails and long range dependence in on/off processes and associated fluid models. *Mathematics of Operations Research*, 23(1):145–165, 1998.

[18] O. J. Boxma and V. Dumas. Fluid queues with long-tailed activity period distributions. *Computer Communications*, 21(17):1509–1529, 1998.

[19] Alberto Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addison Wesley, 1989.

[20] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

[21] Jan Beran. *Statistics for long-memory processes*. Chapman & Hall, New York, 1994.

[22] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3(4):785–798, 1995.

[23] P. Veicht. Matlab wavelet code. Available: `http://www.emulab.ee.mu.oz.au/~darryl/secondorder_code.html`.

[24] Ken Duffy, Christoper King, and David Malone. Wavelet C-code. Available: `http://www.cnri.dit.ie/wavelet/`.

[25] Luigi Rizzo. Dummynet: a simple approach to the evaluation of network protocols. *ACM Computer Communication Review*, 27(1):31–41, 1997.