MODELING TELETRAFFIC ARRIVALS BY A POISSON CLUSTER PROCESS†

GILLES FAŸ, BÁRBARA GONZÁLEZ-ARÉVALO, THOMAS MIKOSCH,
AND GENNADY SAMORODNITSKY

Abstract. In this paper we consider a Poisson cluster process N as a generating process for the arrivals of packets to a server. This process generalizes in a more realistic way the infinite source Poisson model which has been used for modeling teletraffic for a long time. At each Poisson point Γ_j, a flow of packets is initiated which is modeled as a partial iid sum process $\Gamma_j + \sum_{i=1}^{k} X_{ji}$, $k \le K_j$, with a random limit K_j which is independent of (X_{ji}) and the underlying Poisson points (Γ_j). We study the covariance structure of the increment process of N. In particular, the covariance function of the increment process is not summable if the right tail $P(K_j > x)$ is regularly varying with index $\alpha \in (1,2)$, the distribution of the X_{ji}'s being irrelevant. This means that the increment process exhibits long-range dependence. If $\mathrm{var}(K_j) < \infty$ long-range dependence is excluded. We study the asymptotic behavior of the process $(N(t))_{t\ge 0}$ and give conditions on the distribution of K_j and X_{ji} under which the random sums $\sum_{i=1}^{K_j} X_{ji}$ have a regularly varying tail. Using the form of the distribution of the interarrival times of the process N under the Palm distribution, we also conduct an exploratory statistical analysis of simulated data and of Internet packet arrivals to a server. We illustrate how the theoretical results can be used to detect distributional characteristics of K_j, X_{ji}, and of the Poisson process.

1. The model

Recent analysis of broadband measurements shows that the data sets exhibit two characteristic properties: heavy tails and long-range dependence (LRD). Traditional traffic models using independent inter-arrival times of jobs with distribution tails of job sizes which are exponentially bounded imply short range dependence in the traffic and hence are not appropriate for describing high-speed network traffic. Empirical evidence on the existence of LRD in traffic measurements can be found, for example, in Crovella and Bestavros [9], Crovella et al. [10], Leland et al. [21].

A standard model for explaining these empirically observed facts is the so-called ON/OFF model. In it, traffic is generated by a large number of independent ON/OFF sources such as workstations in a big computer lab. An ON/OFF source transmits data at a constant rate to a server if it is ON and remains silent if it is OFF. Every individual ON/OFF source generates an ON/OFF process consisting of independent alternating ON- and OFF-periods. The lengths of the ON-periods are identically distributed and so are the lengths of OFF-periods. Support for this model in the form of statistical analysis of Ethernet Local Area Network traffic of individual sources was provided in Willinger et al. [34]; the conclusions of this study are that the lengths of the ON- and OFF-periods are heavy-tailed and in fact Pareto-like with tail parameter between 1 and 2. In particular, the lengths of the ON- and OFF-periods have finite means but infinite variances. Further evidence

Date: September 2, 2005.

† A large part of this research was done with support of Institut Mittag-Leffler of the Royal Swedish Academy of Sciences when the authors participated in the Fall 2004 program on Queuing Theory and Teletraffic Theory. Mikosch's research is also partially supported by MaPhySto, the Danish research network for mathematical physics and stochastics and the Danish Research Council (SNF) Grant No 21-04-0400. Samorodnitsky's research is also partially supported by NSF grant DMS-0303493 and NSA grant MSPF-02G-183 at Cornell University. González-Arévalo's research is partially supported by BoRSF grant LEQSF(2004-2007)-RD-A-31 at the University of Louisiana at Lafayette.

is in Crovella et al. [9, 10], and Leland et al. [21] which present evidence of Pareto-like tails in file lengths, transfer times and idle times in World Wide Web traffic. It is then an immediate consequence of the heavy-tailed ON/OFF periods that the stationary ON/OFF process exhibits LRD in the sense that its covariance function is not absolutely summable, see Heath et al. [17]. Standard references to stationary processes exhibiting LRD are Brockwell and Davis [7], Section 13.2, or Samorodnitsky and Taqqu [32], Chapter 7.

A closely related model is the infinite source Poisson model. In it, transmission initiations or connections by sources happen at the points of a rate λ Poisson process. The transmission durations are iid random variables independent of the times of connection initiation. The transmission lengths have finite mean, infinite variance and heavy tails. During a transmission, a source transmits at unit rate. This model was studied in Mikosch et al. [25]. In this model, LRD of the stationary process of active sources at a given time is due to the infinite variance of the transmission lengths. This is in agreement with the ON/OFF model mentioned above.

It is the goal of the present paper to study another model which extends the infinite source Poisson and the ON/OFF models in a simple, but more realistic way. We assume that the first packets of a flow arrive at the points Γ_j of a rate $\lambda > 0$ Poisson process $\widetilde{N}$ on $\mathbb{R}$ and enumerate the positive points as follows: $0 < \Gamma_1 < \Gamma_2 < \cdots$. Each flow then consists of several packets, which arrive at the times

$$Y_{jk} = \Gamma_j + S_{jk}\,,$$

where for every j,

$$S_{jk} = \sum_{i=1}^{k} X_{ji}\,, \quad 0 \le k \le K_j\,.$$

Here X_{ji} are iid non-negative random variables and K_j are iid integer-valued random variables. We also assume that (Γ_j), (K_j) and (X_{ji}) are mutually independent. In what follows, we write K, X, etc., for generic random variables with the same distribution as K_j, X_{ji}, etc., and we also write

$$S_0 = 0\,, \quad S_n = X_1 + \cdots + X_n\,, \quad n \ge 1\,,$$

where (X_n) is an iid sequence with common distribution as X_{ji}. Notice that this model coincides with the infinite source Poisson model if $K = 0$. In the literature, this model is also referred to as Poisson cluster process (Bartlett [4]) or branching Poisson process (Lewis [22]); cf. Daley and Vere-Jones [12] and the references therein. Lewis used this model for analyzing computer failure patterns and Bartlett applied it to bunching in motor traffic. Hohn et al. [19] applied the Poisson cluster process for modeling computer traffic. In particular, they model flows arriving at the points of a Poisson process. In that paper the authors also conduct an extensive wavelet based empirical study in order to get support for their model. See also Hohn and Veitch [18] where, in particular, a cluster model with gamma intra-cluster arrivals is fitted to real-life data.

We consider the point process N on $\mathbb{R}$ given by

$$N(B) \;=\; \#\{(j,k) : j \in \mathbb{Z}\,, 0 \le k \le K_j : Y_{jk} \in B\}\,, \quad B \in \mathcal{B}\,,$$

where $\mathcal{B}$ is the Borel σ-field of $\mathbb{R}$. Notice that N is stationary due to the stationarity of the underlying Poisson process. For a Borel set $A \subset \mathbb{R}$ we will also write

$$N_A(B) \;=\; \#\{(j,k) : j \in \mathbb{Z}\,, 0 \le k \le K_j : \Gamma_j \in A \;\text{ and }\; Y_{jk} \in B\}\,, \quad B \in \mathcal{B}\,.$$

We also write for convenience

$$N(t) = N(0,t]\,, \quad t \ge 0\,.$$

Here and in what follows, we write for any measure M and intervals (a,b), $[a,b)$, etc., $M(a,b) = M((a,b))$, $M[a,b) = M([a,b))$, etc. Finally, we denote by H the counting process generated by a single cluster starting at time zero:

$$H(B) = \#\{k : 0 \le k \le K, S_k \in B\}, \quad B \in \mathcal{B}.$$

We consider a non-decreasing enumeration of the non-negative points of N:

$$0 \le T_1 \le T_2 \le \cdots$$

The *Palm measure* of the stationary point process N puts a point $T_0 = 0$ at the origin, and under this measure the interarrival times $T_n - T_{n-1}$, $n = 1, 2, \ldots$ form a stationary sequence (see e.g. Baccelli and Bremaud [2] or Daley and Vere-Jones [12].) We will denote its marginal distribution by F_0.

We start our studies of the process N in Section 2, where we characterize some of its basic properties such as its expectation, the variance as well as its covariance structure. In particular, we determine conditions under which the covariance function $\gamma_N(h) = \operatorname{cov}(N(t, t+1], N(t+h, t+h])$ of the stationary increment process $(N(t, t+1])_{t\in\mathbb{N}}$ has power law behavior in the sense that $\gamma_N(h) = L(h)h^{-\gamma}$ for some $\gamma \in (0, 1)$ and a slowly varying function L, i.e., $L(cx)/L(x) \to 1$ as $x \to \infty$ for every $c > 0$. Notice that γ_N is then not summable, i.e., the increment process exhibits LRD. An interesting observation in this context is given in Proposition 2.4: LRD is not possible unless $\operatorname{var}(K) = \infty$ and the distribution of X is not relevant in this context. However, if $P(K > x)$ is regularly varying with index $-\alpha$, $\alpha \in (1, 2)$ and $EX < \infty$, regular variation of the increment process results; see Theorem 2.5. We mention in this context that we will also say that the random variable K is regularly varying with index α and use this convention throughout the paper.

In Section 3 we prove asymptotic results for the process $(N(t))_{t\ge 0}$, including the central limit theorem with Brownian motion and stable Lévy motion limits. Both are processes with independent stationary increments.

One of the interesting questions when considering real-life teletraffic data is where the heavy tails of the random marks $S_K = \sum_{i=1}^K X_i$ come from. This question is answered in Section 4. One possible explanation is that the X_i's are heavy-tailed and the tail of K is smaller compared to $P(X > x)$. Then the tail $P(S_K > x)$ is essentially determined by $P(X > x)$; see Proposition 4.1. Alternatively, a heavy tail of $P(S_K > x)$ results from the heavy tail $P(K > x)$ if the tail $P(X > x)$ is of smaller order than $P(K > x)$; see Proposition 4.3. We also address the reverse question: given we know that $P(S_K > x)$ is regularly varying, may we conclude that K or X is regularly varying?

In Section 5 we study the Palm distribution of the process N. Under this distribution, the interarrival process $(T_i - T_{i-1})$ constitutes a stationary ergodic sequence with $T_0 = 0$ a.s. We derive the distribution of $T_i - T_{i-1}$ under the Palm distribution.

In Section 6 we use the previous results for an exploratory analysis of real-life Internet and simulated data. Assuming the data come from a Poisson cluster model, we use the strong law of large numbers of $N(t)$ and the ergodic theorem for the interarrival times $T_i - T_{i-1}$ under the Palm distribution in order to get estimates of the quantities λ and EK. Here we assume that we only observe the packet arrival times T_i and that we do not know which flow a given packet belongs to, how many packets are there in each flow, etc. Assuming that K is regularly varying with index α and that the tail of X is lighter than the tail of K, we also estimate the tail parameter α by using wavelet based estimation techniques.

Although we are aware of the fact that the Poisson cluster model is a rather simple model for real-life teletraffic, we think that it is more realistic than the commonly used ON-OFF or the infinite source Poisson models. In particular, cluster models can be extended in various ways. For example, Mikosch and Samorodnitsky [24] model traffic by a very general a stationary marked point process. At each arrival time an activity starts and the marks stand for the amount of work brought into the system at the arrival. Here many different limiting regimes are possible for the workload process.

2. Expectation, variance and covariance structure of N

Consider a marked point process obtained by marking the Poisson point Γ_j by the pair $(K_j, (X_{ji}, i \ge 1))$. The obtained process N^* is a Poisson point process in $\mathbb{E} = \mathbb{R} \times \{0, 1, \dots\} \times \mathbb{R}^\infty$ with mean measure $m^* = (\lambda\,\mathrm{Leb}) \times F_K \times F_X^\infty$, where F_K and F_X are, correspondingly, the laws of K and X. Then for every interval $(a, b]$ we have

$$N(a, b] = \int_{\mathbb{E}} f_{(a,b]}(\gamma, k, (x_i)_{i=1,2,\dots})\, N^*(d(\gamma, k, (x_i)_{i=1,2,\dots}))\,, \tag{2.1}$$

where

$$f_{(a,b]}(\gamma, k, (x_i)_{i=1,2,\dots}) = \sum_{j=0}^{k} I_{\{\gamma + \sum_{i=0}^{j} x_i \in (a,b]\}}\,. \tag{2.2}$$

From the well-known properties of integrals with respect to the Poisson random measures (just differentiate the Laplace functional in, say, Resnick [31], Proposition 3.6) we know that $N(a, b] < \infty$ a.s. if and only if

$$\int_{\mathbb{E}} \min\left(1, f_{(a,b]}(\gamma, k, (x_i)_{i=1,2,\dots})\right)\ m^*(d(\gamma, k, (x_i)_{i=1,2,\dots})) < \infty, \tag{2.3}$$

and that

$$EN(a, b] = \int_{\mathbb{E}} f_{(a,b]}(\gamma, k, (x_i)_{i=1,2,\dots})\, m^*(d(\gamma, k, (x_i)_{i=1,2,\dots})), \tag{2.4}$$

and

$$\mathrm{var}(N(a, b]) = \int_{\mathbb{E}} f^2_{(a,b]}(\gamma, k, (x_i)_{i=1,2,\dots})\, m^*(d(\gamma, k, (x_i)_{i=1,2,\dots})). \tag{2.5}$$

2.1. The expectation.

Proposition 2.1. *Assume $EK < \infty$. Then*

$$EN(t) = \lambda\, t\, (EK + 1)\,, \quad t \ge 0\,. \tag{2.6}$$

Proof. For each $k = 0, 1, \dots$, let $N^{(k)}$ be the point process of the kth points in each flow. That is, $N^{(k)}$ has as its points Y_{jk}, $j \in \mathbb{Z}$, such that $K_j \ge k$. Then by the properties of a Poisson process we immediately see that each $N^{(k)}$ is a homogeneous Poisson process on $\mathbb{R}$ with rate $\lambda P(K \ge k)$. Since $N = \sum_k N^{(k)}$, we see immediately that

$$EN(t) = \sum_{k=0}^{\infty} EN^{(k)}(t) = \sum_{k=0}^{\infty} \lambda\, P(K \ge k)\, t = \lambda\, t\, (EK + 1)\,,$$

as required. □

Interestingly, the mean of $N(t)$ (and, in particular, the fact that it is finite) does not depend on the distribution of X. This last fact extends to the finiteness of the higher moments of $N(t)$, as will be seen in the next subsection.

2.2. The covariance structure of the increment process.

We start with

Proposition 2.2. *Assume $EK < \infty$. Then for any $-\infty < a < b < \infty$, $EN^2(a, b] < \infty$.*

Proof. By stationarity it is enough to prove the claim for $a = 0$, $b = 1$. Since the mean of $N(0, 1]$ is finite by Proposition 2.1, we only need to check that the right-hand side of (2.5) is finite. We have

$$\begin{aligned} & \int_{\mathbb{E}} f^2_{(0,1]}(\gamma, k, (x_i)_{i=1,2,\ldots})\, m^*(d(\gamma, k, (x_i)_{i=1,2,\ldots})) \\ = & \ \lambda \int_{\mathbb{R}} E\left(\sum_{j=0}^{K} I_{\{\gamma+S_j\in(0,1]\}}\right)^2 d\gamma \\ \le & \ \lambda \int_0^\infty E\left(\sum_{j=0}^{K} I_{\{S_j\in(\gamma,\gamma+1]\}}\right)^2 d\gamma + \lambda\, EH^2[0,1]. \end{aligned}$$

Since the last term in the right-hand side above is, obviously, finite, it remains to prove that the first term in the right-hand side above is finite as well. For $\gamma > 0$ let $M_\gamma = \inf\{j \ge 0 : S_j > \gamma\}$. Then

$$E\left(\sum_{j=0}^{K} I_{\{S_j\in(\gamma,\gamma+1]\}}\right)^2 \le E\left[I_{\{K\ge M_\gamma,\, S_{M_\gamma}\le\gamma+1\}} EH^2(0,1]\right],$$

and so we only need to check that

$$\int_0^\infty P\left(K \ge M_\gamma,\, S_{M_\gamma} \le \gamma+1\right) d\gamma < \infty.$$

This, however, follows from the fact that by Proposition 2.1 and (2.4)

$$\begin{aligned} \infty \ > & \ \int_{\mathbb{E}} f_{(0,1]}(\gamma, k, (x_i)_{i=1,2,\ldots})\, m^*(d(\gamma, k, (x_i)_{i=1,2,\ldots})) = \lambda \int_{\mathbb{R}} E\left(\sum_{j=0}^{K} I_{\{\gamma+S_j\in(0,1]\}}\right) d\gamma \\ \ge & \ \lambda \int_0^\infty E\left(\sum_{j=0}^{K} I_{\{S_j\in(\gamma,\gamma+1]\}}\right) d\gamma \ge \lambda \int_0^\infty P\left(K \ge M_\gamma,\, S_{M_\gamma} \le \gamma+1\right) d\gamma. \end{aligned}$$

□

Interestingly, $EK < \infty$ is a sufficient condition for the *second moment* of $N(a, b]$ to be finite. A very similar argument shows that, in the case $EK < \infty$, *all moments* of $N(a, b]$ are finite. Furthermore, using (2.3) instead of (2.5) we see that, unless $EK < \infty$, $N(a, b] = \infty$ a.s. Hence the condition $EK < \infty$ is necessary and sufficient for the finiteness of $N(a, b]$ for $-\infty < a < b < \infty$, and, when finite, $N(a, b]$ has finite moments of all orders. Therefore, in the remainder of this paper we always assume, often without explicit mentioning, that $EK < \infty$.

Let U^* be a measure on $\mathbb{R}^2$ defined by

$$U^*(A \times B) = E\left[\sum_{n_1=0}^{K} I_{\{S_{n_1}\in A\}} \sum_{n_2=0}^{K} I_{\{S_{n_2}\in B\}}\right], \quad A, B \in \mathcal{B}. \tag{2.7}$$

Proposition 2.3. *Let $EK < \infty$. For any two bounded Borel sets A, B*

$$\operatorname{cov}(N(A), N(B)) = \lambda \int_{\mathbb{R}} U^*((s + A) \times (s + B))\, ds. \tag{2.8}$$

Proof. The statement is an immediate consequence of the isometry property (2.5) and the definition of U^*. □

There are several immediate conclusions from this. First of all, since we can write

$$\begin{aligned}\operatorname{cov}(N(A),N(B)) &= \lambda \int_{\mathbb{R}} ds \int_{\mathbb{R}_+^2} I_{\{y_1-s\in A\,,y_2-s\in B\}} U^*(d(y_1,y_2)) \\ &= \lambda \int_{\mathbb{R}_+^2} U^*(d(y_1,y_2)) \int_{\mathbb{R}} I_{\{u\in A\,,u+y_2-y_1\in B\}}\, du\,,\end{aligned}$$

the covariance measure $\widetilde{\gamma}_2$ of the process N (see Daley and Vere-Jones [12]) is given by

$$\widetilde{\gamma}_2 = \lambda\, U^* \circ T^{-1}\,,$$

where $T:\mathbb{R}_+^2 \to \mathbb{R}$ is defined by the relation $T(y_1,y_2)=y_2-y_1$.

Further, utilizing the definition of U^* in (2.7) and writing $\overline{G}=1-G$ for the right tail of any distribution G, we have for the covariance function $\gamma_N(h)=\operatorname{cov}(N(0,1],N(h,h+1])$ of the process $(N(t,t+1])$, if $h\ge 1$,

$$\begin{aligned}\gamma_N(h) &= \lambda \int_{-1}^{\infty} U^*((s,s+1]\times(s+h,s+h+1])\,ds \\ &= \lambda\, E\left[\sum_{n=0}^{K}\sum_{m=0}^{K}\int_{-1}^{\infty} I_{\{s<S_n\le s+1\,,s+h<S_m\le s+h+1\}}\,ds\right] \\ &= \lambda\, E\left[\sum_{n=0}^{K}\sum_{m=n+1}^{K} E\big(S_n\wedge(S_m-h)-(S_n-1)\vee(S_m-h-1)\big)_+\right] \\ &= \lambda\, E\left[\sum_{n=0}^{K}\sum_{m=n+1}^{K}\int_{(h-1,h]}(y+1-h)\,F_{S_{m-n}}(dy)+\int_{(h,h+1]}(h+1-y)\,F_{S_{m-n}}(dy)\right] \\ &= \lambda\, E\left[\sum_{n=0}^{K}\sum_{m=n+1}^{K}\left(\int_{(h-1,h]}\overline{F}_{S_{m-n}}(z)\,dz-\int_{(h,h+1]}\overline{F}_{S_{m-n}}(z)\,dz\right)\right] \\ &= \lambda\, E\left[\sum_{k=1}^{K}(K-k+1)\int_{(0,1]}\big(\overline{F}_{S_k}(x+h-1)-\overline{F}_{S_k}(x+h)\big)\,dx\right]. \qquad (2.9)\end{aligned}$$

This result gives an expression for γ_N in terms of the distributions F_{S_n} of the partial sums S_n and the distribution of K. An immediate consequence is the following result.

Proposition 2.4. *Assume $EK<\infty$. We have for the integrated covariance function γ_N of the process $(N(t,t+1])$*

$$\int_1^{\infty}\gamma_N(h)\,dh = \lambda\, E\sum_{k=1}^{K}(K-k+1)\int_0^1\big(x\wedge(2-x)\big)\overline{F}_{S_k}(x)\,dx\,,$$

and the sum on the right-hand side converges if and only if $\operatorname{var}(K)<\infty$.

This is an interesting observation insofar that LRD in the sense of non-summability of the covariance function is excluded for the increment process $(N(t,t+1])$ unless $\operatorname{var}(K)=\infty$, and this property is completely independent of the distribution of X.

The actual rate of decay of the covariance function $\gamma_N(h)$ as $h\to\infty$ when $\operatorname{var}(K)=\infty$ depends on the tail of the distribution of K, and here the distribution of X does play a role. The result below is an example of what the situation may be.

Theorem 2.5. *Assume that K is regularly varying with index $\alpha \in (1,2)$ or $\alpha = 1$ and $EK < \infty$. Assume also that X has a non-arithmetic distribution and $EX < \infty$. Then*

$$\begin{aligned}\gamma_N(h) &\sim \lambda\,(EX)^{\alpha-2}\int_h^\infty \overline{F}_K(y)\,dy\\ &\sim \lambda\,(EX)^{\alpha-2}\frac{1}{\alpha-1}\,h\,\overline{F}_K(h)\,,\quad \text{if } \alpha>1\end{aligned}$$

as $h \to \infty$.

Proof. We have by (2.9)

$$\begin{aligned}\gamma_N(h) &= \lambda\sum_{k=1}^\infty E(K-k+1)_+\int_{(0,1]}\big(\overline{F}_{S_k}(x+h-1)-\overline{F}_{S_k}(x+h)\big)\,dx\\ &= \lambda\int_h^{h+1}\left(\sum_{k=1}^\infty E(K-k+1)_+\,\big(\overline{F}_{S_k}(y-1)-\overline{F}_{S_k}(y)\big)\right)dy\,. \qquad (2.10)\end{aligned}$$

The function $E(K-k+1)_+ = \int_{x\geq k-1}\overline{F}_K(x)\,dx$ is non-increasing and, by Karamata's theorem (see Bingham et al. [6]), regularly varying with index $-\alpha+1$. Furthermore, for $\alpha > 1$,

$$E(K-k+1)_+ \sim \frac{1}{\alpha-1}k\,\overline{F}_K(k)\,.$$

By a weighted renewal function argument (see Theorem 2 of Alsmeyer [1]) it then follows that

$$\sum_{k=1}^\infty E(K-k+1)_+\,\big(\overline{F}_{S_k}(y-1)-\overline{F}_{S_k}(y)\big) \sim E(K-y+2)_+\,(EX)^{\alpha-2}\,,\quad y\to\infty\,,$$

and the statement of the theorem follows by substitution into (2.10). □

Let us go back to the case $\mathrm{var}(K) < \infty$, and notice that in this case the discrete sum

$$\sum_{h=1}^\infty \gamma_N(h) = \lambda\,E\sum_{k=1}^K (K-k+1)\int_0^1 \overline{F}_{S_k}(x)\,dx < \infty\,. \qquad (2.11)$$

This implies that, taking the limit over the integers, we obtain

$$\lim_{m\to\infty}\frac{\mathrm{var}(N(m))}{m} = \gamma_N(0) + 2\sum_{h=1}^\infty \gamma_N(h)\,. \qquad (2.12)$$

A computation similar to that in (2.9) gives us

$$\gamma_N(0) = \lambda\big(EK+1\big) + 2\lambda E\sum_{k=1}^K (K-k+1)\int_0^1 \big(1-\overline{F}_{S_k}(x)\big)\,dx\,. \qquad (2.13)$$

Substituting (2.11) and (2.13) in (2.12), one obtains

$$\lim_{m\to\infty}\frac{\mathrm{var}(N(m))}{m} = \lambda\,E[(K+1)^2]\,.$$

By the stationarity of the point process N, the asymptotically linear growth of the variance extends to the continuous limit.

Proposition 2.6. *Assume that* $\mathrm{var}(K) < \infty$. *Then*

$$\lim_{t\to\infty}\frac{\mathrm{var}(N(t))}{t} = \lambda\,E[(K+1)^2]\,.$$

3. Asymptotic results for N

3.1. The strong law of large numbers.

Proposition 3.1. *We have*

$$N(t)/t \stackrel{\text{a.s.}}{\rightarrow} \lambda\,(EK+1)\,, \quad t\to\infty\,.$$

Proof. The stationary point process N is ergodic since it is a cluster process with an ergodic parent process (the underlying Poisson process); see Westcott [33], and the statement of the proposition follows (see, e.g., Daley and Vere-Jones [12]). □

Remark 3.2. The above result also follows from Daley [11] who proved strong law of large numberss for more general so-called center point processes (not necessarily Poisson) with iid clusters.

3.2. The central limit theorems.

We first consider the situation where the cluster size $K+1$ has a finite variance.

Proposition 3.3. *Assume* $\operatorname{var}(K)<\infty$. *Then N satisfies the functional central limit theorem:*

$$\left(\frac{N(rt)-\lambda rt(EK+1)}{\sqrt{\lambda r E[(K+1)^2]}}\,,\ 0\le t\le 1\right) \Rightarrow \big(B(t)\,,\ 0\le t\le 1\big)\,, \quad \textit{as} \quad r\to\infty \tag{3.14}$$

in terms of convergence of the finite-dimensional distributions, where $\big(B(t)\,,0\le t\le 1,\big)$ *is standard Brownian motion.*

Remark 3.4. Daley [11] proves one-dimensional central limit theorems for more general (not necessarily Poisson) cluster processes with iid clusters.

Proof. Observe that the stationarity of the point process N implies that any weak (in terms of convergence of the finite-dimensional distributions) limit point of the left-hand side of (3.14) has stationary increments. It follows that, in order to prove the functional central limit theorem (3.14), it is enough to show the one-dimensional convergence

$$\frac{N(rt)-\lambda rt(EK+1)}{\sqrt{\lambda r E[(K+1)^2]}} \Rightarrow B(t) \quad \text{as} \quad r\to\infty \tag{3.15}$$

for every $t>0$. Indeed, (3.15) implies that all finite-dimensional distributions in the left-hand side of (3.14) form tight families. Any limit point must be an infinitely divisible process with Gaussian marginals, and so the whole process is Gaussian. By the stationarity of the increments and (3.15) its covariance function must coincide with that of the standard Brownian motion, and so the latter is the only possible limit point. This will imply the convergence of the finite-dimensional distributions in (3.14).

It is, clearly, enough to prove (3.15) for $t=1$. We first calculate the variance of $N_{(0,r]}(0,r]$. Using the order statistics property of the Poisson process, with $U_1,U_2,\ldots$ being i.i.d. uniform random variables in $(0,1)$ independent of all other random variables below, we have

$$\begin{aligned} \operatorname{var}(N_{(0,r]}(0,r]) &= \operatorname{var}\left(\sum_{j=1}^{\widetilde N(r)}\sum_{k=0}^{K_j} I_{\{r\,U_j+S_{jk}\le r\}}\right) \\ &= \lambda r\, E\left[\left(\sum_{k=0}^{K} I_{\{S_k\le r\,U\}}\right)^2\right] \\ &\sim \lambda r\, E[(K+1)^2] \end{aligned} \tag{3.16}$$

as $r \to \infty$. We conclude by Proposition 2.6 that $\mathrm{var}(N_{(-\infty,0]}(0,r]) = o(r)$ as $r \to \infty$ and the central limit theorem for $N(r)$ is completely determined by $N_{(0,r]}(0,r]$. Recalling the representation in law for $N_{(0,r]}(0,r]$ used in (3.16),

$$N_{(0,r]}(0,r] \stackrel{d}{=} \sum_{j=1}^{\widetilde{N}(r)} G_j \,, \tag{3.17}$$

where $G_j = \sum_{k=0}^{K_j} I_{\{S_{jk} \le rU_j\}}$. we see that the right-hand side of (3.17) is a random sum of iid random variables. Since $\widetilde{N}(r)$ is independent of (K_j), (U_j) and (X_{jk}), we may apply a standard central limit result for double arrays of independent random variables. According to Petrov [30], Theorem 4.2, it suffices to verify the following conditions

$$r\,P(|G - EG| > \varepsilon\sqrt{r}) \;\to\; 0\,, \quad \forall \varepsilon > 0\,, \quad r \to \infty\,, \tag{3.18}$$

$$\sqrt{r}\,E[G - EG] I_{\{|G-EG| \le \sqrt{r}\}} \;\to\; 0\,, \quad r \to \infty\,, \tag{3.19}$$

$$E[G^2]/E[(K+1)^2] \;\to\; 1\,, \quad r \to \infty\,. \tag{3.20}$$

Since $|G - EG| \le K + 1 + EK$ and $\mathrm{var}(K) < \infty$, it immediately follows that (3.18) holds. Moreover, the left-hand term in (3.19) can be bounded as follows

$$\left| -\sqrt{r} E[G - EG] I_{\{|G-EG| > \sqrt{r}\}} \right| \;\le\; \sqrt{r}\,E([K + 1 + EK] I_{\{K+1+EK > \sqrt{r}\}}) \to 0\,, \quad r \to \infty\,,$$

since $\mathrm{var}(K) < \infty$. Relation (3.20) follows by Lebesgue dominated convergence.

The proves the convergence of the finite-dimensional distributions. □

Next, we consider the situation described in Theorem 2.5, where the cluster size has infinite variance. Note that the assumption of non-arithmeticity of the distribution of X is not needed here.

Proposition 3.5. *Assume that K is regularly varying with index $\alpha \in (1,2)$ and that $EX < \infty$. Then N satisfies the functional central limit theorem:*

$$\left(\frac{N(rt) - \lambda rt(EK+1)}{\Theta(r)}\,,\; 0 \le t \le 1 \right) \Rightarrow \big(L_\alpha(t)\,,\; 0 \le t \le 1 \big)\,, \quad \textit{as} \quad r \to \infty\,, \tag{3.21}$$

in terms of convergence of the finite-dimensional distributions, where $\Theta : (0,\infty) \to (0,\infty)$ is a nondecreasing function such that

$$\lim_{r \to \infty} r\,P\big(K > \Theta(r)\big) = 1 \tag{3.22}$$

and $\big(L_\alpha(t)\,, 0 \le t \le 1,\big)$ is a spectrally positive α-stable Lévy motion with $L_\alpha(1) \sim S_\alpha(\sigma, 1, 0)$. Here $\sigma = 1/C_\alpha$ with

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x\, dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)} & \textit{if } \alpha \ne 1 \\ 2/\pi & \textit{if } \alpha = 1 \end{cases}.$$

Remark 3.6. It follows immediately from (3.22) that the function Θ is regularly varying at infinity with exponent $1/\alpha$. Examples of such functions can be obtained from the usual inverses of the tail $P(K > \cdot)$; see Resnick [31].

Note also that we are using the standard notation for the distribution of stable random variables, and that C_α is the multiplicative constant in the tail of an α-stable random variable; see Samorodnitsky and Taqqu [32].

Proof. We will prove the following two statements:

$$\frac{N_{(-\infty,0]}(0,r] - EN_{(-\infty,0]}(0,r]}{\Theta(r)} \to 0 \quad \text{as } r \to \infty \tag{3.23}$$

in probability and

$$\frac{N_{(0,r]}(0,r] - EN_{(0,r]}(0,r]}{\Theta(r)} \Rightarrow L_\alpha(1) \quad \text{as } r \to \infty. \tag{3.24}$$

In order to see that the statement of the proposition will then follow, take any $k \geq 1$ and $0 = t_0 < t_1 < \cdots < t_k$ and write

$$\left(\frac{N(rt_i) - \lambda r t_i(EK+1)}{\Theta(r)},\, i = 1,\ldots,k\right) = \frac{1}{\Theta(r)}(I_1,\ldots,I_k) + \frac{1}{\Theta(r)}(J_1,\ldots,J_k),$$

where

$$I_i = \sum_{j=1}^{i} \Big(N_{(t_{j-1},t_j]}(t_{j-1},t_j] - EN_{(t_{j-1},t_j]}(t_{j-1},t_j]\Big)$$

and

$$J_i = \sum_{j=1}^{i} \Big(N_{(-\infty,t_{j-1}]}(t_{j-1},t_j] - EN_{(-\infty,t_{j-1}]}(t_{j-1},t_j]\Big)$$

for $i = 1,\ldots,k$. It follows from (3.23), (3.24), stationarity and independent increments of the underlying Poisson process and regular variation of the function Θ that

$$\left(\frac{N(rt_i) - \lambda r t_i(EK+1)}{\Theta(r)},\, i = 1,\ldots,k\right)$$

$$\Rightarrow \left(t_1^{1/\alpha}R_1, t_1^{1/\alpha}R_1 + \big(t_2 - t_1\big)^{1/\alpha}R_2, \ldots, \sum_{j=1}^{k}\big(t_j - t_{j-1}\big)^{1/\alpha}R_j\right) \quad \text{as } r \to \infty,$$

where $R_1,\ldots,R_k$ are i.i.d. copies of $L_\alpha(1)$, from which (3.21) follows.

We proceed, therefore, to prove (3.23) and (3.24). We start with observing that the ratio on the left-hand side of (3.23) is an infinitely divisible random variable with characteristic function of the form

$$\psi_r(\theta) = \exp\left\{-\int_0^\infty \Big(1 - e^{i\theta x} - i\theta x\Big)\,\mu_r(dx)\right\}, \tag{3.25}$$

where the Lévy measure μ_r is given by

$$\mu_r = \lambda\,(\text{Leb} \times P) \circ T_r^{-1}, \tag{3.26}$$

and $T_r : \mathbb{R}_+ \times \Omega \to \mathbb{R}_+$ is given by

$$T_r(u,\omega) = \frac{H(u,u+r](\omega)}{\Theta(r)}.$$

To see this, simply write $N_{(-\infty,0]}(0,r]$ as an integral with respect to a Poisson random measure as in (2.1) and use, for example, Lemma 12.2 in Kallenberg [20].

It follows from Theorem 15.14 in Kallenberg [20] that (3.23) will result once we show that, as $r \to \infty$,

$$\mu_r \Rightarrow \text{o (null measure) vaguely on } \overline{\mathbb{R}} \setminus \{0\},$$

$$\int_{x \leq 1} x^2\,\mu_r(dx) \to 0 \quad \text{and} \quad \int_{x>1} x\,\mu_r(dx) \to 0.$$

All of these statements will follow once we show that for every $y > 0$

$$\int_0^\infty \left(y^2 \wedge x^2\right) \mu_r(dx) \to 0 \tag{3.27}$$

as $r \to \infty$ and that

$$\lim_{y\to\infty} \limsup_{r\to\infty} \int_{x>y} x\, \mu_r(dx) = 0. \tag{3.28}$$

Note that the left-hand side of (3.27) is

$$\lambda \int_0^\infty E\left(y^2 \wedge \left(\frac{H(u,u+r]}{\Theta(r)}\right)^2\right) du \;\; \le \;\; \lambda \int_0^\infty E\left(y^2 \wedge \left(\frac{H(u,\infty)}{\Theta(r)}\right)^2\right) du$$

and so by the dominated convergence theorem (3.27) will follow once we show that

$$\int_0^\infty E\left(y^2 \wedge (H(u,\infty))^2\right) du < \infty.$$

This is, however, clear because the integral above cannot exceed

$$y^2\, ES_K = y^2\, EX\, EK < \infty.$$

Furthermore,

$$\begin{aligned}
\int_{x>y} x\, \mu_r(dx) &= \lambda \int_0^\infty E\left(\frac{H(u,u+r]}{\Theta(r)} I_{\left\{\frac{H(u,u+r]}{\Theta(r)}>y\right\}}\right) du \\
&\le \frac{\lambda}{\Theta(r)} E\left[I_{\{K>y\,\Theta(r)\}} \int_0^\infty H(u,u+r]\, du\right] \\
&\le \frac{\lambda\, r}{\Theta(r)} E\left[K I_{\{K>y\Theta(r)\}}\right],
\end{aligned}$$

and so by Karamata's theorem

$$\limsup_{r\to\infty} \int_{x>y} x\, \mu_r(dx) \le \lambda\, \frac{\alpha}{\alpha-1} y^{-(\alpha-1)}.$$

Now (3.28) follows, and so we have established (3.23).

We now switch to proving (3.24). Once again, the ratio on the left-hand side of (3.24) is an infinitely divisible random variable with characteristic function of the form (3.25), with Lévy measure μ_r given by (3.26), but this time $T_r : [0,r] \times \Omega \to \mathbb{R}_+$ is given by

$$T_r(u,\omega) = \frac{H(0,u](\omega)}{\Theta(r)}.$$

By appealing once again to Theorem 15.14 in Kallenberg [20] we see that (3.24) will follow once we check that, as $r \to \infty$,

$$\mu_r(y,\infty) \to \lambda\, y^{-\alpha} \tag{3.29}$$

for all $y > 0$,

$$\int_0^1 x^2\, \mu_r(dx) \to \lambda \int_0^1 x^2\, \alpha x^{-(\alpha+1)}\, dx \tag{3.30}$$

and

$$\int_1^\infty x\, \mu_r(dx) \to \lambda \int_1^\infty x\, \alpha x^{-(\alpha+1)}\, dx. \tag{3.31}$$

Note that

$$\begin{aligned}\mu_r(y,\infty) &= \lambda \int_0^r P\left(H(0,u] > y\,\Theta(r)\right)\,du \\ &= \lambda\,E\left[I_{\{K>y\,\Theta(r)\}}\left(r - S_{[y\,\Theta(r)]}\right)_+\right] \\ &= \lambda\,P\left(K > y\,\Theta(r)\right)\,E\left(r - S_{[y\,\Theta(r)]}\right)_+ .\end{aligned}$$

Since

$$r - [y\Theta(r)]EX \le E\left(r - S_{[y\,\Theta(r)]}\right)_+ \le r,$$

we recall (3.22), the regular variation of the tail of K, and the fact that Θ is regularly varying with exponent $1/\alpha$, to obtain (3.29).

Furthermore, for any ϵ, $\delta > 0$ we have by the regular variation of K,

$$\begin{aligned}\int_0^\epsilon x^2\,\mu_r(dx) &= \lambda \int_0^r E\left[\left(\frac{H(0,u]}{\Theta(r)}\right)^2 I_{\{H(0,u]\le\epsilon\Theta(r)\}}\right]\,du \\ &\le \lambda\frac{r}{\Theta(r)^2}E\left(K^2 I_{\{K\le\delta\Theta(r)\}}\right) + \lambda\frac{1}{\Theta(r)^2}\int_0^r E\left[(H(0,u])^2 I_{\{K>\delta\Theta(r),\,H(0,u]\le\epsilon\Theta(r)\}}\right]\,du \\ &\le \lambda\frac{r}{\Theta(r)^2}\Big(C\big(\delta\Theta(r)\big)^2 P\big(K>\delta\Theta(r)\big)\Big) + \lambda\epsilon^2 r P\big(K>\delta\Theta(r)\big).\end{aligned}$$

Here C is a finite positive constant. Therefore, using (3.22) and the regular variation of the tail of K we obtain

$$\limsup_{r\to\infty}\int_0^\epsilon x^2\,\mu_r(dx) \le \lambda\,\left(C\delta^2+\epsilon^2\right)\delta^{-\alpha}.$$

Letting first $\epsilon \to 0$ and then $\delta \to 0$ we see that

$$\lim_{\epsilon\to 0}\limsup_{r\to\infty}\int_0^\epsilon x^2\,\mu_r(dx) = 0.$$

Using (3.29), we now obtain (3.30).

Finally, the same argument that led to (3.28) also shows that the statement holds with the new Lévy measures μ_r and, as before, (3.31) follows from (3.29). This proves (3.24), and so completes the proof of the proposition. □

4. The tail behavior

In real-life teletraffic one observes heavy tails in various disguises, for example for transferred file lengths, interarrival times of transmissions, etc. In this paper we are merely interested in the periods of activity of one source given by the Poisson arrivals Γ_j of a packet which initiates a flow composed of packets with arrival times $\Gamma_j + S_{jk}$, $0 \le k \le K_j$.

A natural question in this context arises as to where the observed heavy tails of distributions of interarrival times come from — from the interarrival times X_{ji} or from the number of packets K_j initiated at the Poisson arrival Γ_j. It is the goal of this section to answer this question by studying the right tail of the random sum

$$S_K = \sum_{i=1}^K X_i\,,$$

where, as usual, (X_i) and K are independent. We are interested in conditions on the distributions of X and K under which S_K is regularly varying. We start with the case when X has heavier tail than K.

Proposition 4.1. *Assume X is regularly varying for some $\alpha > 0$, $EK < \infty$ and $P(K > x) = o(P(X > x))$. Then, as $x \to \infty$,*

$$P(S_K > x) \sim EK\, P(X > x)\,. \tag{4.32}$$

Remark 4.2. The tail of S_K has been studied even for the class of subexponential distributions of X under the assumption that the tail $P(K > x)$ decays exponentially fast. Then (4.32) remains valid. The subexponential distributions include the regularly varying ones as a subclass, see Embrechts et al. [13], Chapter 1 and Appendix A3, for the definition and properties of subexponential distributions, their relationship with regularly varying distributions and the mentioned result about the tail of S_K. Proposition 4.1 above gives (4.32) under very weak conditions on the tail of K.

Proof. By independence of K and (X_i),

$$\frac{P(S_K > x)}{P(X > x)} = \sum_{k=1}^{\infty} P(K = k)\, \frac{P(S_k > x)}{P(X > x)}\,.$$

By subexponentiality of the distribution of X for every fixed k_0,

$$\sum_{k=1}^{k_0} P(K = k)\, \frac{P(S_k > x)}{P(X > x)} \to \sum_{k=1}^{k_0} P(K = k)\, k\,.$$

Choose $b_k = ck$, where

$$\begin{cases} c > 0 & \text{is arbitrary if } EX = \infty\,, \\ c = EX & \text{if } EX < \infty\,. \end{cases}$$

Choose $\epsilon \in (0,1)$. Then

$$\begin{aligned} \sum_{k=k_0+1}^{\infty} P(K = k)\, \frac{P(S_k > x)}{P(X > x)} &= \sum_{k=k_0+1}^{\infty} P(K = k)\, \frac{P(S_k - b_k > x - b_k)}{P(X > x)} \\ &= \left(\sum_{k_0 < k, k \le \epsilon x} + \sum_{k_0 < k, k > \epsilon x}^{\infty} \right) P(K = k)\, \frac{P(S_k - b_k > x - b_k)}{P(X > x)} \\ &= I_1 + I_2\,. \end{aligned}$$

Then

$$I_2 \le \frac{P(K > \epsilon x)}{P(X > x)} = o(1)$$

since $P(K > x) = o(P(X > x))$ and $P(X > x)$ is regularly varying with index $-\alpha$. For $k \le \epsilon x$, $x - b_k \ge k(\epsilon^{-1} - c) > 0$ for ϵ small. By the large deviation results of A.V. Nagaev [27, 28] (cf. S.V. Nagaev [29]) for $\alpha > 2$ and Cline and Hsing [8] for $\alpha \le 2$ it follows for any $\delta > 0$ that

$$\sup_{y > \delta k} \left| \frac{P(S_k - b_k > y)}{k\, P(X > y)} - 1 \right| \to 0\,. \tag{4.33}$$

Hence, for some positive constant C,

$$\limsup_{x \to \infty} I_1 \le C \sum_{k=k_0+1}^{\infty} P(K = k)\; k \to 0\,, \quad k_0 \to \infty\,.$$

This proves the proposition. □

Real-life teletraffic interarrivals X_{ji} are often not very heavy-tailed due to technological restrictions on the file length and the transmission times. However, it is observed that S_K has tails which are well approximated by power laws. The next result gives an explanation of this phenomenon. It makes plausible that the heavy tail of S_K comes from the heavy tail of K.

Proposition 4.3. *Assume K is regularly varying with index $\beta \geq 0$. If $\beta = 1$, assume that $EK < \infty$. Moreover, let (X_i) be an iid sequence such that $EX < \infty$ and $P(X > x) = o(P(K > x))$. Then, as $x \to \infty$,*

$$P(S_K > x) \sim P(K > (EX)^{-1}\, x) \sim (EX)^{\beta}\, P(K > x)\,. \tag{4.34}$$

Before we give the proof of this result we provide an auxiliary result.

Lemma 4.4. *Assume $0 < h(x) \to 0$ as $x \to \infty$. Then there exists a slowly varying function L such that $L(x) \to \infty$ and $L(x)h(x) \to 0$ as $x \to \infty$.*

Proof. We construct L as follows. Let $L(x) = 1$ for $x \in [0, x_0]$, where $x_0 = \sup\{y : h(y) > 1\}$. Then set $L(x) = 2$ for $x \in (x_0, x_1]$, where $x_1 = \sup\{\sup\{y : h(y) > 2^{-2}\}, 2x_0\}$. Then set $L(x) = 3$ for $x \in (x_1, x_2]$, where $x_2 = \sup\{\sup\{y : h(y) > 3^{-2}\}, 3x_1\}$ and continue this construction in the straightforward way. The function L is slowly varying since for any $c > 0$ and sufficiently large x, cx and x are either in the same interval $(x_k, x_{k+1}]$ or in two neighboring intervals of this form. The fact that $L(x) \to \infty$ is obvious, as well as the fact that $L(x)h(x) \to 0$ since $L(x)h(x) \leq 1/k$ for $x > x_k$. This concludes the proof. □

Proof of Proposition 4.3. Notice that for $\epsilon \in (0, 1)$, by the law of large numbers,

$$\begin{aligned} P(S_K > x) &= \sum_{k=1}^{\infty} P(K = k)\, P(S_k > x) \\ &\geq P(K > (1+\epsilon)(EX)^{-1}x)\, P(S_{[(1+\epsilon)(EX)^{-1}x]} > x) \\ &\sim P(K > (1+\epsilon)(EX)^{-1}x) \sim (1+\epsilon)^{-\beta}\, P(K > (EX)^{-1}x)\,. \end{aligned}$$

Letting $\epsilon \downarrow 0$, this proves the lower bound in (4.34). As to the upper bound, for $\epsilon \in (0, 1)$,

$$P(S_K > x) \;\leq\; P(K > (EX)^{-1}(1-\epsilon)x) + \sum_{k \leq (EX)^{-1}(1-\epsilon)x} P(K = k)\, P(S_k > x)\,.$$

By regular variation,

$$P(K > (EX)^{-1}(1-\epsilon)x) \;\sim\; (1-\epsilon)^{-\beta}\, P(K > (EX)^{-1}x)\,.$$

From this relation the lower bound in (4.34) follows by letting $\epsilon \downarrow 0$, provided one can show that

$$p(x) = \sum_{k \leq (EX)^{-1}(1-\epsilon)x} P(K = k)\, P(S_k > x) = o(P(K > x))\,. \tag{4.35}$$

We will consider several cases. Suppose first that $\beta \geq 1$. In particular, $EK < \infty$. Let (X_i') be an iid sequence of regularly varying random variables with the property that $P(X > x) \leq P(X' > x) = o(P(K > x))$ and that X' is regularly varying with index β. Such a construction is possible by virtue of Lemma 4.4. Write $S_k' = X_1' + \cdots + X_k'$. Then we have $P(S_k > x) \leq P(S_k' > x)$ for all $x \geq 0$. An application of (4.33) yields

$$p(x) \;\leq\; C \sum_{k \leq (EX)^{-1}(1-\epsilon)x} P(K = k)\, k\, P(X' > x) \leq C\, EK\, P(X' > x) = o(P(K > x))\,.$$

for some positive C, as required.

Next we consider the case $\beta < 1$. Let $\delta > 0$ be a small positive number. We have

$$\begin{aligned} p(x) &\le \sum_{k\le\delta x} P(K=k)\,P(S_k>x) + \sum_{\delta x<k\le (EX)^{-1}(1-\epsilon)x} P(K=k)\,P\,(S_k>x) \\ &:= p_1(x;\delta)+p_2(x;\delta). \end{aligned}$$

Note that for every $\delta x < k \le (EX)^{-1}(1-\epsilon)x$, by the law of large numbers

$$P(S_k > x) \le P(S_{[(1-\epsilon)(EX)^{-1}x]} > x) \to 0$$

as $x \to \infty$. Therefore, for any δ as above,

$$p_2(x;\delta) \le o(1)P(K>\delta x) = o\Big(P(K>x)\Big)$$

by the regular variation of K. Furthermore, since $P(S_k > x) \le kEX/x$, we have

$$p_1(x;\delta) \le EX\frac{1}{x}\sum_{k\le\delta x} k\,P(K=k) \le EX\frac{1}{x}\sum_{k\le\delta x} P(K>k).$$

By Karamata's theorem

$$\sum_{k\le\delta x} P(K>k) \sim \frac{1}{1-\beta}\delta x\,P\,(K>\delta x) \sim \frac{1}{1-\beta}\delta^{1-\beta}x\,P(K>x).$$

Therefore,

$$\lim_{\delta\to 0}\limsup_{k\to\infty}\frac{p_1(x;\delta)}{P(K>x)} \le \lim_{\delta\to 0}\frac{EX}{1-\beta}\delta^{1-\beta} = 0,$$

from which (4.35) follows. □

Remark 4.5. One case has been left out in the assumptions of Proposition 4.3, the case where $\beta = 1$ and $EK = \infty$. Somewhat surprisingly, the statement of the proposition fails, in general, in this case, as the following example shows. Let $d_n = e^{n^4}$ for $n \ge 1$ and let the distribution of X be concentrated on the set $\{d_1, d_2, \ldots\}$ with $P(X = d_n) = c_1/\big(d_n(\log(d_n))^{1/2}\big)$, $n = 1, 2, \ldots$ for some $c_1 > 0$. Notice that $EX < \infty$. If we let $P(K = k) = c_2/k^2$, $k \ge 1$ for some $c_2 > 0$, we see that $P(X > x) = o(P(K > x))$. We have for every $k \le x$

$$\begin{aligned} P(S_k>x) &\ge P\left(\max_{i=1,\ldots,k} X_i > x\right) \\ &\ge kP(X>x) - \frac{k(k-1)}{2}\,(P(X>x))^2 \\ &\ge kP(X>x)\,[1-x\,P(X>x)] \\ &\ge \frac{1}{2}kP(X>x) \end{aligned}$$

for x large enough since $EX < \infty$ implies $x\,P(X > x) \to 0$ as $x \to \infty$. Therefore for x large enough,

$$\begin{aligned} P(S_K > x) &\geq \sum_{k \leq x} P(K = k)\,P(S_k > x) \\ &\geq \frac{1}{2}\left(\sum_{k \leq x} k\,P(K = k)\right) P(X > x) \\ &\sim \frac{c_2}{2} \log x\,P(X > x) \end{aligned}$$

as $x \to \infty$. Letting $x \to \infty$ over the set $\{d_1, d_2, \ldots\}$ we see that $\limsup_{x\to\infty} P(S_K > x)/P(K > x) = \infty$.

On the other hand, if in the case $\beta = 1$ and $EK = \infty$ one assumes that $xP(X > x) = o(P(K > x))$, then the conclusion on the proposition still holds. See the proof of Proposition 4.9 below.

Remark 4.6. Under more restrictive assumptions the statement of Proposition 4.3 is in the PhD thesis of Hansen [16], Lemma 3.3.5.

Remark 4.7. A modification of the proof of Proposition 4.1 shows that in the case when $P(K > x) \sim cP(X > x)$ for some $c > 0$ and if X is regularly varying with index $\alpha \geq 1$ and $EX < \infty$, then

$$P(S_K > x) \sim (EK + c\,(EX)^\alpha)P(X > x)$$

(suggested by Daryl Daley (personal communication)). It is also likely that certain results in this section can be extended from the regularly varying case to the more general subexponential case.

In the rest of this section we are interested in the reverse problem: given we know that S_K has regularly varying tail, what can we say about the tails of X and K? In view of Propositions 4.1 and 4.3 above it is clear that this question cannot be answered without additional conditions. Indeed, these lemmas suggest that one has to specify in advance which of the tails $P(X > x)$ or $P(K > x)$ is asymptotically heavier.

We consider a situation where K has a sufficiently light tail. Notice the similarity with Proposition 4.1.

Proposition 4.8. *Assume S_K is regularly varying with index $\alpha > 0$ and $EK^{1\vee(\alpha+\delta)} < \infty$ for some positive δ. Then X is regularly varying with index α and $P(S_K > x) \sim EK\,P(X > x)$.*

Proof. First assume $\alpha \in (0, 1)$. An application of Karamata's Tauberian theorem (see Bingham et al. [6]) yields

$$s^\alpha L(1/s) \sim 1 - E\mathrm{e}^{-sS_K} = E[1 - (E\mathrm{e}^{-sX})^K]\,, \quad s \downarrow 0\,,$$

for some slowly varying function L. An application of Lebesgue dominated convergence and $EK < \infty$ imply that

$$E[1 - (E\mathrm{e}^{-sX})^K] \sim E[K\,(1 - E\mathrm{e}^{-sX})] = EK\,(1 - E\mathrm{e}^{-sX})$$

as $s \downarrow 0$. Another application of Karamata's Tauberian theorem proves that X is regularly varying with index α.

Now assume $\alpha \geq 1$. We will prove that, under the assumptions of the proposition,

$$P\left(X_1^2 + \cdots + X_K^2 > x\right) \quad \text{is regularly varying with exponent } \alpha/2. \tag{4.36}$$

Then, proceeding by induction we may conclude for $\alpha \in [2^l, 2^{l+1})$, $l \geq 0$, that $X_1^{2^{l+1}} + \cdots + X_k^{2^{l+1}}$ is regularly varying with index $\alpha/2^{l+1} \in (0, 1)$. By the first part of the proof, $X_1^{2^{l+1}}$ is regularly varying with index $\alpha/2^{l+1}$ and hence X_1 is regularly varying with index α. It remains, therefore, to prove (4.36).

Notice that for $\varepsilon \in (0,1)$,

$$P(X_1^2+\cdots+X_K^2>x) \le P(S_K^2>x) \le P(X_1^2+\cdots+X_K^2>(1-\varepsilon)x) + P\left(2\sum_{1\le i<j\le K} X_i\,X_j > \varepsilon x\right).$$

We will show that

$$P\left(2\sum_{1\le i<j\le K} X_i\,X_j > \varepsilon x\right) = o\left(P(S_K^2>x)\right). \tag{4.37}$$

If this is true, then we have

$$\begin{aligned}(1-\varepsilon)^{\alpha/2} &= \lim_{x\to\infty}\frac{P(S_K^2>(1-\varepsilon)^{-1}x)}{P(S_K^2>x)} \le \liminf_{x\to\infty}\frac{P(X_1^2+\cdots+X_K^2>x)}{P(S_K^2>x)}\\ &\le \limsup_{x\to\infty}\frac{P(X_1^2+\cdots+X_K^2>x)}{P(S_K^2>x)} \le 1\,.\end{aligned}$$

Letting $\varepsilon \downarrow 0$, (4.36) follows from regular variation of S_K^2.

We prove now (4.37), which will follow once we show that for some $\beta > \alpha/2$,

$$E\left(\sum_{1\le i<j\le K} X_i\,X_j\right)^{\beta} < \infty\,. \tag{4.38}$$

Suppose first that $1 \le \alpha < 2$. Choose $\alpha/2 < \beta < \min(1, (\alpha+\delta)/2)$. Since $P(X>x) \le P(S_K>x)$, we know that $EX_1^\rho < \infty$ for any $0<\rho<1$. Choose $\beta<\rho<1$, $\rho > 2\beta/(\alpha+\delta)$ and observe that, in the obvious notation,

$$\begin{aligned}E\left(\sum_{1\le i<j\le K} X_i\,X_j\right)^{\beta} &= E_K\left\{E_X\left[\left(\sum_{1\le i<j\le K} X_i\,X_j\right)^{\rho}\right]^{\beta/\rho}\right\}\\ &\le E_K\left\{\left[E_X\left(\sum_{1\le i<j\le K} X_i^\rho\,X_j^\rho\right)\right]^{\beta/\rho}\right\}\\ &= (EX_1^\rho)^{2\beta/\rho}\,E\left(\frac{K(K-1)}{2}\right)^{\beta/\rho} < \infty\end{aligned}$$

by the choice of ρ, proving (4.38). The argument in the case $\alpha \ge 2$ is similar: choose $\alpha/2 < \beta < \min\big((\alpha+\delta)/2, \alpha\big)$ and recall that the inequality $P(X>x) \le P(S_K>x)$ implies that $EX_1^\beta < \infty$. We have

$$\begin{aligned}E\left(\sum_{1\le i<j\le K} X_i\,X_j\right)^{\beta} &\le E\left[\left(\frac{K(K-1)}{2}\right)^{\beta-1}\sum_{1\le i<j\le K}(X_i\,X_j)^\beta\right]\\ &= \left(EX_1^\beta\right)^2 E\left(\frac{K(K-1)}{2}\right)^{\beta} < \infty\end{aligned}$$

by the choice of β. This proves (4.38) in all cases and, hence, completes the proof. □

Next we consider the situation where the tail of X is sufficiently light. Notice the similarity with Proposition 4.3.

Proposition 4.9. *Assume S_K is regularly varying with index $\alpha > 0$. Suppose that $EX < \infty$ and $P(X > x) = o(P(S_K > x))$ as $x \to \infty$. In the case $\alpha = 1$ and $ES_K = \infty$, assume that $xP(X > x) = o(P(S_K > x))$ as $x \to \infty$. Then K is regularly varying with index α and*

$$P(S_K > x) \sim (EX)^\alpha \, P(K > x) \,. \tag{4.39}$$

Proof. The upper bound

$$\limsup_{x\to\infty} \frac{P(K > x)}{P(S_K > x)} \le \frac{1}{(EX)^\alpha}$$

follows in the same way as in Proposition 4.3. To prove the corresponding lower bound, rule out, first, the case $\alpha = 1$ and $ES_K = \infty$. As in the proof of Proposition 4.3, we need to show that for every $0 < \epsilon < 1$

$$p(x) = \sum_{k \le (EX)^{-1}(1-\epsilon)x} P(K = k) \, P(S_k > x) = o(P(S_K > x)) \,. \tag{4.40}$$

If we prove that

$$\limsup_{x\to\infty} \frac{P(K > x)}{P(S_K > x)} < \infty \,, \tag{4.41}$$

then the above statement will follow in the same way as in Proposition 4.3. To prove (4.41), choose $\theta > 0$ such that $P(X > \theta) > 0$. Write $Y = \sum_{i=1}^n I_{\{X_i > \theta\}}$ and observe that $Y \sim \text{Bin}(n, P(X > \theta))$. Notice that

$$\begin{aligned} P(S_K > x) &\ge P\left(K > \frac{2x}{\theta P(X > \theta)}\right) \inf\left\{P\left(Y > \frac{n}{2} P(X > \theta)\right) : \quad n \ge \frac{2x}{\theta P(X > \theta)}\right\} \\ &\sim P\left(K > \frac{2x}{\theta P(X > \theta)}\right) \end{aligned}$$

as $x \to \infty$, from which (4.41) follows because of the regular variation of the tail of S_K.

Finally, consider the case $\alpha = 1$ and $ES_K = \infty$, where we assume that $xP(X > x) = o(P(S_K > x))$. To prove (4.40) in this case, notice that for a (small) $\delta > 0$ we have

$$\begin{aligned} p(x) &\le P\left(S_{[(EX)^{-1}(1-\epsilon)x]} > x\right) \\ &\le P\left(X_i > \delta x \text{ for some } i \le [(EX)^{-1}(1-\epsilon)x]\right) \\ &\quad + P\left(\sum_{i \le [(EX)^{-1}(1-\epsilon)x]} X_i I_{\{|X_i| \le \delta x\}} > \epsilon x\right) \\ &:= p_1(x) + p_2(x) \,. \end{aligned}$$

Notice that

$$p_1(x) \le (EX)^{-1}(1-\epsilon)x \, P(X > \delta x) = o(P(S_K > x))$$

because of the regular variation of the tail of S_K. It remains to consider the second term in the right-hand side above. For $\theta > 0$ denote $\mu_\theta = EXI_{\{X \le \theta\}}$ and notice that μ_θ is uniformly bounded in θ. Therefore, for x so large that $\mu_\theta \le \epsilon x/2$ we have

$$p_2(x) \le P\left(\sum_{i \le [(EX)^{-1}(1-\epsilon)x]} Y_i > \frac{\epsilon x}{2}\right) \,,$$

where $Y_i = X_i I_{\{X_i \le \delta x\}} - \mu_{\delta x}$, $i = 1, 2, \ldots$.

Notice that $Y_1, Y_2, \dots$ are zero mean random variables, and for x large enough we have $|Y_i| \le \delta x$. Using Prokhorov's inequality (see Petrov [30], e.g. Lemma A3.6 in Mikosch and Samorodnitsky [23]) we have

$$p_2(x) \le \exp\left\{ -\frac{\epsilon}{8\delta} \text{arcsinh}\left(\frac{\epsilon \delta x^2}{4\,\text{var}\left(\sum_{i \le [(EX)^{-1}(1-\epsilon)x]} Y_i\right)} \right) \right\}.$$

It follows from the assumption $xP(X > x) = o(P(S_K > x))$ that

$$\text{var}\left(\sum_{i \le [(EX)^{-1}(1-\epsilon)x]} Y_i \right) = o(1)x \int_0^{\delta x} P(S_k > z)\, dz = o(1)x^2 P(S_k > x)$$

as $x \to \infty$. Since $\text{arcsinh}(y) \sim \log y$ as $y \to \infty$, we see that, by selecting $\delta > 0$ small enough relative to ϵ, we will also have

$$p_2(x) = o(P(S_K > x)),$$

completing the proof of the proposition. □

5. The distribution of the interarrival times under the Palm measure

In this section we derive an explicit expression for the tail of the distribution function of the interarrival times of the process N under the Palm measure. We will use it in the next section as a tool to fit the parameters of a model to data.

Recall that F_0 denotes the distribution of the interarrival times under the Palm measure. Under this measure $T_0 = 0$ a.s. and the interarrival times of the non-decreasing enumeration $0 \le T_1 \le T_2 \le \cdots$ of the non-negative points of N constitute a stationary ergodic process; see Daley and Vere-Jones [12]. By the Palm-Khinchin formula (Baccelli and Bremaud [2], p. 20 or p. 24) we have

$$\lambda(EK + 1) \int_t^\infty \overline{F}_0(x)\, dx = P(T_1 > t), \quad t \ge 0,$$

where $\overline{F}_0 = 1 - F_0$ denotes the right tail of F_0.

For any cluster the fact whether or not any of its point belongs to the interval $[0, t]$ depends only on the arrival time of the cluster and not on any other clusters. Therefore, by the time-dependent thinning of a Poisson process

$$\begin{aligned}
P(T_1 > t) &= P(N(0, t] = 0) \\
&= \exp\Big\{ -E\Big(\text{number of clusters that contribute at least one point to } [0, t]\Big)\Big\} \\
&= \exp\Big\{ -\lambda \int_{-\infty}^t P\Big(\text{a cluster arriving at time } s \text{ contributes} \\
&\qquad \text{at least one point to } [0, t]\Big)\Big\}\, ds \\
&= \exp\left\{ -\lambda \left[t + \int_{-\infty}^0 P\big(W(-s, -s+t] > 0\big)\, ds \right] \right\} \\
&= \exp\left\{ -\lambda \left[t + \int_0^\infty P\big(W(x, x+t] > 0\big)\, dx \right] \right\},
\end{aligned}$$

where for any Borel set A,

$$W(A) = \sum_{i=1}^K I_{\{S_i \in A\}}.$$

Now,

$$\begin{aligned}\int_0^\infty P(W(x,x+t]>0)\,dx &= E\left(\int_0^\infty I_{\{W(x,x+t]>0\}}\,dx\right)\\ &= E\left(\sum_{i=1}^{K}\int_{S_i}^{S_{i+1}} I_{\{W(x,x+t]>0\}}\,dx\right).\end{aligned}$$

Now, for every $i=0,1,2,\dots$ and $x\in(S_i,S_{i+1}]$,

$$W(x,x+t]=0 \quad \text{if and only if} \quad x+t<S_{i+1}\,.$$

Hence

$$\int_{S_i}^{S_{i+1}} I_{\{W(x,x+t]>0\}}\,dx = X_{i+1}I_{\{X_{i+1}\le t\}} + t\,I_{\{X_{i+1}>t\}} = X_{i+1}\wedge t\,.$$

This implies

$$\int_0^\infty P(W(x,x+t]>0)\,dx = E\left(\sum_{i=1}^{K}(X_i\wedge t)\right) = EK\,E(X\wedge t)\,.$$

Therefore

$$\begin{aligned}P(T_1>t) &= \exp\{-\lambda(t+EK\,E(X\wedge t))\}\\ &= \exp\left\{-\lambda(t+EK\int_0^t P(X>x))\,dx)\right\},\end{aligned}$$

and so

$$\int_t^\infty \overline{F}_0(x)\,dx = \frac{1}{\lambda(EK+1)}\exp\left\{-\lambda(t+EK\int_0^t P(X>x)\,dx)\right\}.$$

By differentiating this expression, we obtain

$$\overline{F}_0(t) = \frac{1}{EK+1}\,(1+EKP(X>t))\,\exp\left\{-\lambda(t+EK\int_0^t P(X>x)\,dx)\right\}. \tag{5.42}$$

Under somewhat more restrictive assumptions this expression was also obtained in Bartlett [4].

6. Fitting the model to Internet traffic data

In this section we fit the proposed Poisson cluster model to real-life Internet data. In order to do this we need some parametric assumptions on the distributions of K and X. Empirical Internet studies suggest that the number of packets in a flow is long-range dependent due to heavy tails of the distribution of K rather than heavy tails of the distribution of X. We assume that the integer-valued random variable K is regularly varying with unknown index α. We also assume the simplifying condition that X is exponentially distributed. If $\alpha\in(1,2)$, it follows from Proposition 4.3 that the flow lengths are regularly varying with index α and the increment process of N has slowly decaying correlations, see Theorem 2.5.

With these assumptions on the distributions of X and K we want to get some exploratory statistical information about the following quantities of interest.

- The arrival rate of the Poisson process: λ.
- The mean of the exponential distribution of X: $1/\theta$.
- The tail index of the distribution of K: α.
- The expected value of K: EK.

6.1. **Description of the data set.** The data set consists of the times of packet arrivals to a server at the University of North Carolina. These traces were taken on Sunday, April 20, 2003, starting at 7:00pm. Since the data set is extremely large we will only use the first 10 million observations, accounting for approximately 245 seconds. In this very short time period the assumption of stationarity of the process N is acceptable. The data may be downloaded from

`http://www-dirt.cs.unc.edu/ts_len/2003_Apr_20_Sun_1900.ts_len.gz`.

This data set has two columns: packet arrival time stamp and packet length. The latter was ignored in our analysis.

6.2. **Estimation of λ.** As mentioned before, we cannot discriminate between the Poisson arrivals Γ_j and the points Y_{ji}. Therefore the estimation of the Poisson arrival rate λ is a major problem. We will extract some information about it by using the results of Section 5 about the distribution of the interarrival times under the Palm distribution. Recalling (5.42), we have

$$\log(\overline{F}_0(t)) \sim -\lambda t\,, \quad t \to \infty\,. \tag{6.43}$$

Now replace in this asymptotic relation F_0 by its empirical counterpart F_n calculated from the sample of the interarrival times $T_i - T_{i-1}$. Under the Palm distribution, this sequence constitutes a stationary ergodic process and therefore $\sup_x |F_n(x) - F_0(x)| \stackrel{\text{a.s.}}{\to} 0$. The latter limit relation and (6.43) encourage one to perform a linear regression of $\log(\overline{F}_n)$ on t, for large values of t. We will ignore the very large values of t since, for those values, the empirical distribution function $F_n(t)$ is a poor estimator of $F_0(t)$. However, we do not expect this restriction to have significant impact on our analysis since the sample size is very large.

In Figure 6.1 we can see that t ranges between 0 and 1.2×10^{-3} seconds, but the empirical distribution function $F_n(t)$ is unreliable after 0.7×10^{-3} seconds. Therefore, we will fit a regression line for $t \in [0.4 \times 10^{-3}, 0.7 \times 10^{-3}]$, where $\log(\overline{F}_n)$ is close to a straight line.

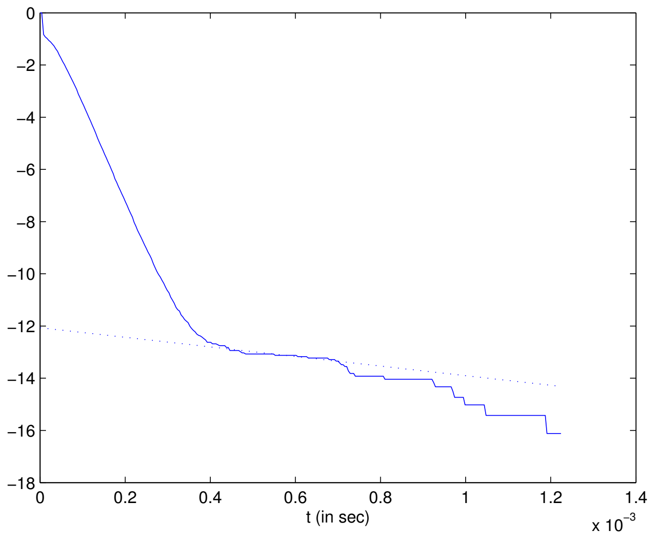

Figure 6.1. *The continuous line shows the natural logarithm of the empirical distribution function,* $\log(\overline{F}_n)$*, and the dotted line shows the fitted regression line for the points where* $t \in [0.4\times10^{-3}, 0.7\times 10^{-3}]$*. The negative of the slope of this regression line is our estimate of* λ*.*

In Figure 6.1 we can see this fitted regression line. The negative of the slope of this regression line is our estimate for λ. That is, $\widehat{\lambda} \approx 1,837$.

6.3. **Estimation of θ.** We conclude from Section 5 that the mean excess function of F_0 is given by

$$E_{F_0}(Y-t \mid Y>t) = \frac{\int_t^\infty \overline{F}_0(x)\,dx}{\overline{F}_0(t)} = \frac{1}{\lambda(1+EKP(X>t))} \to \frac{1}{\lambda}, \quad t\to\infty\,. \tag{6.44}$$

Replacing F_0 by its sample version F_n, we have as $n\to\infty$,

$$E_{F_n}(Y-t \mid Y>t) \stackrel{\text{a.s.}}{\to} E_{F_0}(Y-t \mid Y>t) \tag{6.45}$$

by the ergodic theorem, uniformly on t-compact sets. The asymptotic relations (6.44) and (6.45) encourage one to estimate $P(X>t)$. Since we have a very large sample size n we expect that

$$r_n(t) = \frac{1}{E_{F_n}(Y-t \mid Y>t)} \approx \lambda\,(1+EK\,P(X>t))\,. \tag{6.46}$$

Under the assumption that X has an exponential distribution with mean $1/\theta$, a reorganization of (6.46) yields

$$\log\left[\frac{1}{EK}\left(\frac{1}{\lambda}r_n(t)-1\right)\right] \approx -\theta t\,. \tag{6.47}$$

This approximation is inaccurate for too large t-values. Therefore, we use moderately large values of t to fit a regression to the left-hand side of (6.47) on t. We interpret the negative of the slope of this regression line as an estimate of θ.

In Figure 6.2 we see that the left-hand side of (6.47) is roughly linear for $t\in[0.2\times10^{-3}, 0.4\times10^{-3}]$, where we fit the linear regression, shown with a dotted line. This gives us an estimate of θ of $\widehat{\theta}\approx 16,239$.

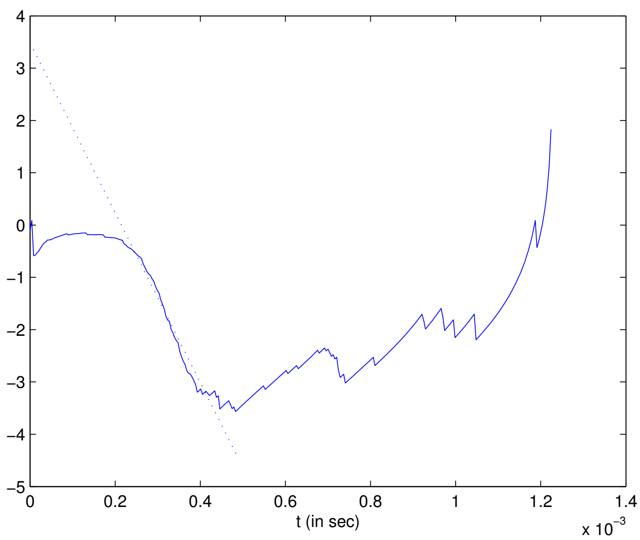

Figure 6.2. *The continuous line shows* $\log\left[\frac{1}{EK}\left(\frac{1}{\lambda}r_n(t)-1\right)\right]$*, and the dotted line shows the fitted regression line for* $t\in[0.2\times10^{-3}, 0.4\times10^{-3}]$*. The negative of the slope of this line is our estimate of* θ*.*

6.4. **Estimation of α.** We know from Theorem 2.5 that the covariance function γ_N of the increment process N is regularly varying with index $1-\alpha$ provided $\alpha \in (1,2)$. Figure 6.3 suggests that this is a reasonable assumption for our data set. Thus we may apply commonly used methods for estimating the so-called Hurst parameter $H=(3-\alpha)/2$. Several of these methods can be classified as Fourier or wavelet based methods. Both have high computational efficiency which is necessary for large data sets encountered in telecommunications. In this application we will use wavelet based methods.

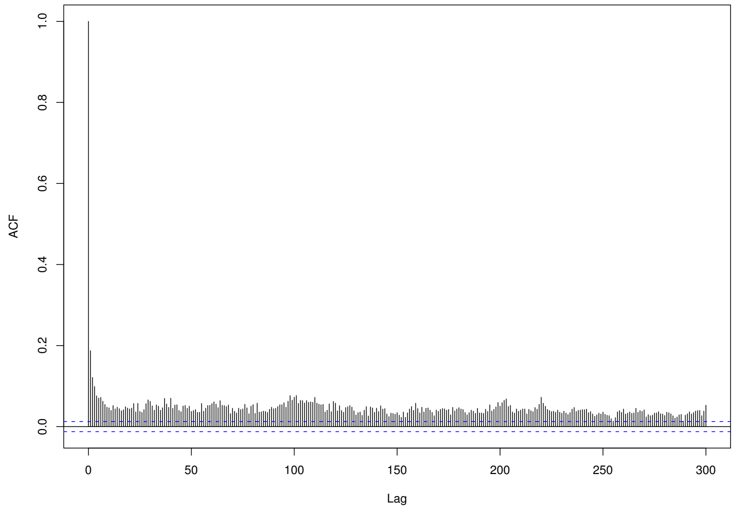

Figure 6.3. *Sample autocorrelation function (ACF) for the increment process* $(N(t10^{-3},(t+1)10^{-3}])_{t\in\mathbb{Z}}$.

We give a short description of the method; for details we refer to Bardet et al. [3]. Let $d_{j,k}$, $j,k \in \mathbb{Z}$, denote the wavelet coefficients of the increment process of N with respect to the Haar mother wavelet. For a variety of models and under appropriate conditions including regular variation of the covariance function one has

(1) (*within-scale behavior*) the continuous or discrete wavelet transform performs as an approximate whitening of the data, *i.e.* the wavelet coefficients at a given scale form a stationary weakly dependent time series
(2) (*between-scale behavior*) the variance of the wavelet coefficients are related to the scale j as a power law with an exponent depending linearly on α as $j \to \infty$:

$$v_j := \operatorname{var}(d_{j,k}) \approx 2^{(2-\alpha)j}. \tag{6.48}$$

There exists a large number of papers where α is estimated via a linear regression of $\log(\hat{v}_j)$ on j. Here $\hat{v}_j$ is the empirical variance of the wavelet coefficients at scale j. Most of the results are obtained in the Gaussian case, see Bardet et al. [3]. An empirical wavelet based application to the cluster Poisson model can be found in Hohn et al. [19]. Using the above properties of the wavelet coefficients, in particular (6.48), an alternative estimation procedure was suggested by Moulines et al. [26]; see also Wornell and Oppenheim [35]. They propose to estimate α by minimization of a local Whittle contrast, defined as follows:

$$\hat{\alpha} = \arg\min_{\alpha'} \left\{ \log\Big(\sum_{(j,k)\in\Delta} \frac{d_{j,k}^2}{2^{(2-\alpha')j}} \Big) + \log 2(2-\alpha') \sum_{(j,k)\in\Delta} j \right\}, \tag{6.49}$$

where we choose Δ as follows: assume that $d_{j,k}$ is observable for $J_{\min} \le j \le J_{\max}$ and $0 \le k < 2^{J_{\max}-j}$. Then we take

$$\Delta = \{j, k | J_0 \le j \le J_{\max}\,,\, 0 \le k < 2^{J_{\max}-j}\}\,,$$

for some $J_0 \in [J_{\min}, J_{\max})$. Faÿ et al. [14] show that this procedure leads to a consistent estimate of the Hurst parameter for the infinite source Poisson model under the following conditions

$$J_0 \to \infty\,, \quad J_{\max} - J_0 \to \infty \quad \text{and} \quad \limsup J_0/J_{\max} < 1/\alpha\,.$$

Notice that $J_0 = [J_{\max}/2]$ is a possible choice satisfying the above conditions for $\alpha \in (1, 2)$. The choice of J_0 corresponds to a bias-variance trade-off, similarly to the choice of the frequency interval for the classical regression estimators based on the periodogram. Simulations indicate that wavelet Whittle estimators outperform the classical Fourier methods when α is close to 1.

Since theoretical results for this model are not available at present, we conducted a small Monte-Carlo study to confirm acceptable performance of the wavelet estimator (6.49) for the Poisson cluster model, before we applied it to the telecommunication data. We simulated $N_{\mathrm{MC}} = 50$ traces of the Poisson cluster process with parameters $\lambda = 1800$ and $\theta = 16000$ and α in $\{1.2, 1.4, 1.6, 1.8\}$. The duration of each trace is 2.4 seconds. The increment process is computed by dividing each path into 2^{13} intervals of equal length, corresponding to $J_{\max} = 13$. According to (6.49), the estimator $\hat{\alpha}_k$ is computed from the wavelet coefficients, for all possible choices of $J_0 \in \{1, \dots, 12\}$. Finally, the empirical mean-square error $\frac{1}{N_{\mathrm{MC}}} \sum_{k=1}^{N_{\mathrm{MC}}} (\hat{\alpha}_k - \alpha)^2$ is also computed. In Figure 6.4 the bias-variance trade-off is illustrated. Moreover, it is indicated in Figure 6.5 that the optimal choice of J_0 is not very sensitive with respect to the value of α: the value $J_0 = 5$ is efficient for the chosen values α. Thus one can avoid sophisticated adaptive procedures for the estimation of α from the data given that the remaining parameters are fixed.

Back to the data set, we sampled the increasing process at 2^{19} points, and performed the estimation on the 32 disjoint subsamples of size 2^{13}, with $J_0 = 5$. Averaging the estimates, we get $\hat{\alpha} \simeq 1.66$.

6.5. **Estimation of EK.** We proceed by estimating the mean number of packets in a flow, that is, $EK + 1$, or, equivalently, EK, exploiting Proposition 3.1. In Figure 6.6 we see that $N(t)/t$ approaches a constant for increasing t, as we expect from the theory. Thus we take the estimate of $\lambda(EK + 1)$ to be $N(T)/T = \frac{10{,}000{,}000}{245} \approx 40{,}816$, where T is the total length of time when the measurements were taken.

From this estimate and the estimate $\widehat{\lambda}$ of the rate of the Poisson process (see Section 6.2) we get an estimate for the expected value of K: $\widehat{EK} = 21$.

Acknowledgments. This research started at the Mittag-Leffler Institute in Stockholm during the thematic period on Queuing Theory and Teletraffic Theory in Fall 2004. We take pleasure to thank the organizers for their hospitality and for providing optimal conditions for our research. We are grateful to Daryl Daley and Allan Gut for various stimulating discussions. In particular, we learnt from Daryl about his paper [11]. We thank Sid Resnick and Philippe Soulier for drawing our attention to the papers of Darryl Veitch and his co-authors.

References

[1] Alsmeyer, G. (1992) On generalized renewal measures and certain first passage times. *Ann. Probab.* **20**, 1229–1247.

[2] Baccelli, F. and Brémaud, P. (2003) *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences.* 2nd edition. Springer, New York.

[3] Bardet, J. M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long-range dependent processes. *Stat. Inference Stoch. Process.* **3** 85–99. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).

[4] Bartlett, M.S. (1963) The spectral analysis of point processes. *J. Roy. Statist. Soc., Series B* **25**, 264–296.

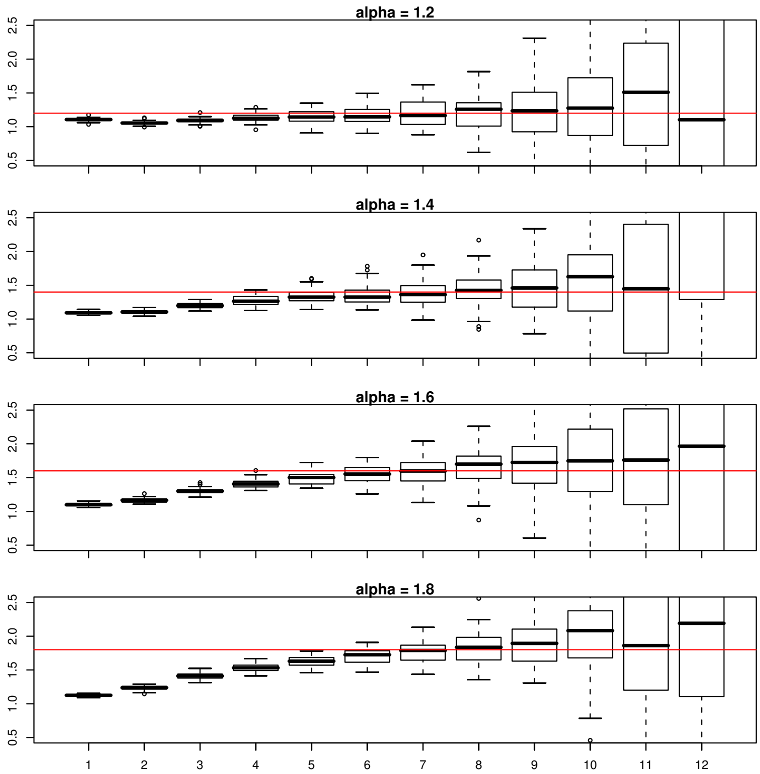

Figure 6.4. *The distributions of $\hat{\alpha}$ for different values of J_0 and α*

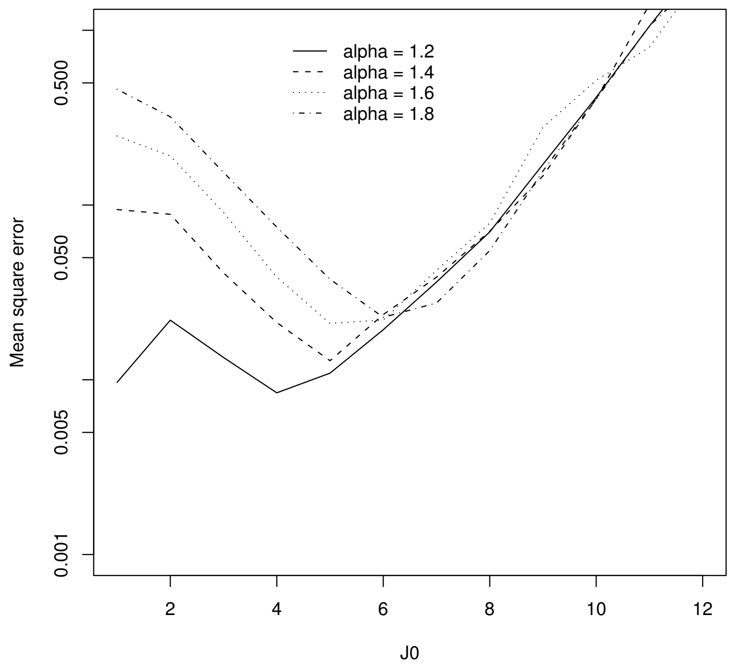

Figure 6.5. *The mean-square error of the estimation of α for different values of J_0 and α*

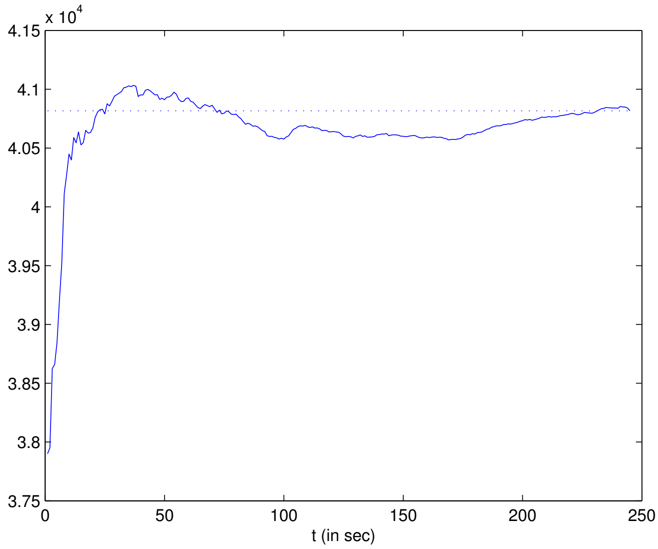

Figure 6.6. *The continuous line shows the value of* $\frac{N(t)}{t}$*, and the dotted line the estimated value of* $\lambda(EK+1)$*.*

[5] Billingsley, P. (1968(*Convergence of Probability Measures.* Wiley, New York.

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

[7] Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

[8] Cline, D.B.H. and Hsing, T. (1998) Large deviation probabilities for sums of random variables with heavy or subexponential tails. Technical Report, Texas A& M University.

[9] Crovella, M. and Bestavros, A. (1996) Self-similarity in world wide web traffic: evidence and possible causes. *Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems* **24**, 160–169.

[10] Crovella, M., Bestavros, A. and Taqqu, M.S. (1996) Heavy-tailed probability distributions in the world wide web. Preprint, 1996.

[11] Daley, D. (1972) Asymptotic properties of stationary point processes with generalized clusters. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **21**, 65–76.

[12] Daley, D. and Vere-Jones, D. (1988) *An Introduction to the Theory of Point Processes.* Springer, New York.

[13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[14] Faÿ, G., Roueff, F. and Soulier, P. (2005) Estimation of the memory parameter of the stationary on nonstationary infinite source Poisson model. Preprint.

[15] Gut, A. (1988) *Stopped Random Walks.* Springer, New York.

[16] Hansen, N.R. (2003) *Markov Controlled Excursions, Local Alignment and Structure.* PhD thesis, Institute of Mathematics, University of Copenhagen.

[17] Heath, D., Resnick, S. and Samorodnitsky, G. (1998) Heavy tails and long range dependence in ON/OFF processes and associated fluid models. *Math. Oper. Res.* **23**, 145–165.

[18] Hohn, N. and Veitch, D. (2003) Inverting sampled traffic. *ACM/SIGCOMM Internet Measurement Conference,* Miami, USA, November 2003, pp. 222-233

[19] Hohn, N., Veitch, D. and Abry, P. (2003) Cluster processes: a natural language for network traffic. *IEEE Trans. Signal Process.* **51** 2229–2244.

[20] Kallenberg, O. (2001) *Foundations of Modern Probability.* 2nd Edition. Springer, New York.

[21] Leland, W.E., Taqqu, M.S., Willinger, W. and Wilson, D.V. (1993) On the self-similar nature of Ethernet traffic. *ACM/SIGCOMM Computer Communications Review* 183–193.

[22] Lewis, P.A.W. (1964) A branching Poisson process model for the analysis of computer failure patterns. *J. Roy. Statist. Soc., Series B* **26**, 398–456.

[23] Mikosch, T. and Samorodnitsky, G. (2000) The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.* **10**, 1025–1064.

[24] Mikosch, T. and Samorodnitsky, G. (2005) Scaling limits for workload processes. Work in progress.

[25] Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. (2002) Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12**, 23–68.

[26] Moulines, E., Roueff, F. and Taqqu, M. (2005) A wavelet Whittle estimator of the memory parameter of a Gaussian process. Preprint.

[27] Nagaev, A.V. (1969a) Limit theorems for large deviations where Cramér's conditions are violated (in Russian). *Izv. Akad. Nauk UzSSR, Ser. Fiz.–Mat. Nauk* **6**, 17–22.

[28] Nagaev, A.V. (1969b) Integral limit theorems for large deviations when Cramér's condition is not fulfilled I,II. *Theory Probab. Appl.* **14**, 51–64, 193–208.

[29] Nagaev, S.V. (1979) Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745–789.

[30] Petrov, V.V. (1995) *Limit Theorems of Probability Theory.* Oxford University Press, Oxford.

[31] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[32] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman & Hall, New York.

[33] Westcott, M. (1971) On existence and mixing results for cluster point processes. *J. Royal Stat. Soc. Series B* **33**, 290–300.

[34] Willinger, W., Taqqu, M.S., Sherman, R. and Wilson, D. (1995) Self-similarity through high variability: statistical analysis of ethernet lan traffic at the source level. Proceedings of the ACM/SIGCOMM'95, Cambridge, MA. *Computer Communications Review* **25**, 100–113.

[35] Wornell, G.W. and Oppenheim, A.V. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing* **40**, 611–623

Laboratoire Paul-Painlevé, Université Lille 1, 59655 Villeneuve d'Ascq cedex, France
E-mail address: gilles.fay@univ-lille1.fr

Mathematics Department, University of Louisiana at Lafayette, 217 Maxim D. Doucet Hall, P.O.Box 41010, Lafayette, LA 70504-1010, U.S.A.
E-mail address: barbara@louisiana.edu

Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark
E-mail address: mikosch@math.ku.dk , www.math.ku.dk/∼mikosch

School of Operations Research and Industrial Engineering, Cornell University, 220 Rhodes Hall Ithaca, NY 14853, U.S.A.
E-mail address: gennady@orie.cornell.edu