Revisiting an old friend: On the observability of the relation between Long Range Dependence and Heavy Tail

Patrice Abry, Pierre Borgnat, Fabio Ricciato, Antoine Scherrer, Darryl Veitch

HAL Id: ensl-00338215

https://hal-ens-lyon.archives-ouvertes.fr/ensl-00338215

Submitted on 12 Nov 2008

Revisiting an old friend: On the observability of the relation between Long Range Dependence and Heavy Tail

Patrice Abry · **Pierre Borgnat** · **Fabio Ricciato** · **Antoine Scherrer** · **Darryl Veitch**

Abstract Taqqu's Theorem plays a fundamental role in Internet traffic modeling, for two reasons: First, its theoretical formulation matches closely and in a meaningful manner some of the key network mechanisms controlling traffic characteristics; Second, it offers a plausible explanation for the origin of the long range dependence property in relation with the heavy tail nature of the traffic components. Numerous attempts have since been made to observe its predictions empirically, either from real Internet traffic data or from numerical simulations based on popular traffic models, yet rarely has this resulted in a satisfactory quantitative agreement. This raised in the literature a number of comments and questions, ranging from the adequacy of the theorem to real world data to the relevance of the statistical tools involved in practical analyses. The present contribution aims at studying under which conditions this fundamental theorem can be actually seen at work on real or simulated data. To do so, numerical simulations based on standard traffic models are analyzed in a wavelet framework. The key time scales involved are derived, enabling a discussion of the origin and nature of the difficulties encountered in attempts to empirically observe Taqqu's Theorem.

1 Motivation

More than a decade of research works have shown that modeling the statistical properties of Internet traffic is challenging. Indeed, traffic traces are characterized by non trivial statistical

Patrice Abry, Pierre Borgnat, Antoine Scherrer
CNRS, Laboratoire de Physique ENS Lyon, Université de Lyon, 46 allée d'Italie, 69007 Lyon, France
E-mail: {firstname.name}@ens-lyon.fr

Fabio Ricciato
University of Salento, Lecce, Italy and Forschungszentrum Telekommunikation Wien (FTW), Austria
E-mail: ricciato@ftw.at

Darryl Veitch
ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN), an affiliated program of National ICT Australia (NICTA), Dept. of E&E Engineering, University of Melbourne, Australia
E-mail: dveitch@unimelb.edu.au

properties, the two most prominent being: Long Range Dependence (LRD), or asymptotic self-similarity, of the time-series counting the 'volume' of communication over time (aggregated traffic counts) [1–4]; Heavy-tailness (HT) of important traffic statistics such as the size of the objects sent through the Internet (be they images, movies, WWW files or emails) [5,6], or the size of computer activity sessions, measured either by duration or in packet or byte count, or flow lengths [7–9]. From the first disclosures of LRD in Internet traffic [1], mechanisms explaining its origin have been sought.

A theorem due to Taqqu and collaborators related LRD and HT via an infinite superimposition of On/Off processes [1,2,10]. It was immediately to play a fundamental role for Internet traffic modeling as its theoretical formulation closely matched quantities which are meaningful in networking terms, such as flow sizes. Furthermore using objects with HT size distributions as the input of some queueing system yields output traffic times series with LRD. Moreover, the theorem is appealing as it provides a simple closed form relation $H(\alpha) = (3 - \alpha)/2$ between the LRD and HT parameters, respectively H and α. A precise and detailed formulation is postponed to Sec. 2.

The theoretical relation above opened the door to a large number of works whose goal was its empirical validation. Prominently, seminal works on this question [5,6] obtained results in partial and qualitative agreement with the theoretical prediction, first for WWW traffic, and later for more general types of traffic. However, numerous difficulties arose: obtaining and storing sufficiently long traffic traces was non-trivial, and the values of parameters found in real traffic, being fixed by the data, do not cover the full range allowed for by theory. Practitioners then resorted to numerical simulations to generate synthetic traffic based either on popular time series models or on traffic simulators such as NS-2. This offered the possibility of controlling and tuning the traffic parameters, while preserving a number of realistic features of actual traffic. An orthogonal issue was the fact that statistical tools for HT and LRD measurement and analysis had their own problems. For instance, LRD measured via R/S or variogram procedures have been shown since to suffer from many drawbacks and to have poor performance (responsible, for instance, of the bias of Fig. 3 in [5]). Currently, the use of the wavelet methodology, proposed in [3], for the analysis of LRD is well-assessed and documented, and intensively used. Yet, again, validation has turned out to be far from straightforward and quantitative satisfactory agreements have rarely been reported. Often, a significant mismatch between the wavelet based estimated LRD exponent and the HT index are observed (for example see the detailed and motivating contributions in [11,12], and the analysis of [13]). This lead practitioners to raise a number of questions: Is the Theorem misunderstood? misinterpreted? misused? Is the wavelet based LRD analysis procedure incorrect? or not operating in this framework? Is the simulation set-up at fault?

The goal of the present contribution is obviously not to call Taqqu's Theorem into question. Instead, it consists in analyzing the conditions under which it can actually and practically be involved in explaining the extent to which long range dependence (or self-similarity) in real Internet traffic time series can be related to heavy-tailness. The present contribution aims at clarifying the issues which have been raised over time. It focuses on a careful analysis and precise quantification of the time scales involved in practical in relation to the Theorem, as well as on a correct use and interpretation of the wavelet based LRD analysis. Hence, the objective is twofold: First, to provide practitioners with a better understanding of the conclusions that can be drawn (or not) from the wavelet methodology applied to traffic traces; Second, to clarify under which practical conditions the relation between LRD and HT can actually be seen.

The theoretical background needed for the statement of Taqqu's Theorem is recalled in Section 2, while the wavelet based LRD analysis procedure is surveyed in Section 3.1.

Taqqu's Theorem is then rephrased and reread in this wavelet framework in Section 3.2. To understand the issue in the empirical validation of the $H(\alpha)$ relation, we make use of synthetic traffic traces simulated from standard traffic models (M/G/N and M/G/∞ models); the simulation set-ups are described in 4. The careful analysis of the ranges of scales, formulated in the wavelet framework for M/G/∞ models, is detailed in Sec. 5; this enables us to precisely state the actual conditions under which the $H(\alpha)$ can be practically (and satisfactorily) observed. Section 6 discusses what these results imply for other traffic models and network simulations. Finally, discussions and comments on the use and applicability to real Internet data is given in Section 7.

2 Theoretical Background

2.1 Heavy tail

A random variable (RV) W is said to possess a heavy tail distribution with tail exponent $\alpha > 0$ (and will be denoted α-HT), if [14, 15]:

$$P(W > w) = 1 - F_W(w) \sim cw^{-\alpha} \ \text{ for } w \to \infty, \tag{1}$$

where F_W denotes its cumulative distribution. The key feature of a α-HT distribution is that its moments are infinite for orders larger or equal to α. The range $1 < \alpha < 2$ corresponds to RV with finite mean but infinite variance, while the range $0 < \alpha \leq 1$ implies infinite mean. This latter range will not be addressed here, as considered not relevant for Internet traffic modeling (see, a contrario, [11, 16] for wireless traffic). The former range is the focus of this contribution as it both matches experimental observation of real Internet data [5, 6] and is at the core of Taqqu's Theorem formulation.

A canonical example of α-HT distribution is that of Pareto:

$$F_W(w) = 1 - \left(\frac{k}{w+k} \right)^{\alpha}, \tag{2}$$

with $k > 0$. Its mean reads: $\mathbb{E}W = k/(\alpha - 1)$, when $\alpha > 1$ and its variance is: $\text{Var } W = k^2 \frac{\alpha}{(\alpha-1)^2(\alpha-2)}$, when $\alpha > 2$. In the following, we use use Pareto random variables in simulations, but results are valid for other heavy tail random variables.

2.2 Long range dependence

Long range dependence is classically modeled via the power law divergence of the spectral density at origin (cf. e.g., [4, 17, 18]): The second order stationary process Y, with covariance $C_Y(\tau)$ and spectrum $\Gamma_Y(\nu)$, is said to be LRD if and only if there is a constant $\gamma \in]0, 1[$ and:

$$\Gamma_Y(\nu) \sim D|\nu|^{-\gamma}, \text{ for } |\nu| \to 0. \tag{3}$$

Equivalently, this can be rewritten as $C_Y(\tau) \sim D'|\tau|^{-(1-\gamma)}$, when $\tau \to +\infty$. Essentially, LRD implies that no characteristic correlation time (or no characteristic frequency) can be singled out in the data and that the sum of the covariance function diverges. The major practical consequence of LRD is to significantly degrade estimation procedure performance.

Well known examples of LRD processes are the increment processes of finite variance self-similar process with stationary increments. Self-similar processes are characterized by finite dimensional distributions that are covariant under dilations (cf. e.g. [15,19]) and their properties are controlled by a sole exponent $H \in]0,1[$, referred to as the Hurst parameter. Fractional Brownian motion (FBM), B_H is the most prominent only Gaussian member of that class, and actually appears in Taqqu's Theorem. The increment process $Y(t) = X(t+1) - X(t)$ of any finite variance self similar processes with stationary increments is characterized by a covariance of the form [19],

$$\mathbb{E}\{Y(t+\tau)Y(t)\} \sim \sigma^2 H(2H-1)|\tau|^{-(2-2H)}, \text{ for } \tau \to +\infty \tag{4}$$

and is hence LRD whenever $1/2 < H < 1$, with the correspondence $\gamma = 2H - 1$. Parameter H is hence classically used as the common scaling exponent for both LRD and self-similarity. By abuse of notations, it is also often used in the literature to refer to as a scaling exponent be it related to LRD or self-similarity only.

2.3 A fundamental result in Self-Similar traffic modeling

2.3.1 Stating the Theorem

Let us now formulate Taqqu's Theorem dealing with limits of the process known as the On/Off model [4]. Let N denote a strictly positive integer. Let $\{Z_i(t), t \in \mathbb{R}\}_{i=1,\dots,N}$ be a collection of independent binary i.e., taking values 0 or 1, reward-renewal processes with independent activation periods. Let us assume that the On periods, τ_{on}, consist of i.i.d. random variables drawn from a α-HT distribution. The i.i.d. Off periods are characterized either by β-HT distribution or by some light tailed distribution. Let us consider the superposition $Y_N(t) = \sum_{i=1}^N Z_i(t)$ and its cumulated sum,

$$X_N(tT) = \int_0^{Tt} Y_N(u)\mathrm{d}u = \int_0^{Tt} \left(\sum_{i=1}^N Z_i(u) \right) \mathrm{d}u. \tag{5}$$

The theorem states that there exists a positive constant C such that

$$\lim_{T \to +\infty} \lim_{N \to +\infty} \frac{X_N(tT) - \frac{\mathbb{E}\tau_{\text{on}}}{\mathbb{E}\tau_{\text{on}}+\mathbb{E}\tau_{\text{off}}}NTt}{C\sqrt{N}T^H} = B_H(t), \tag{6}$$

where B_H is the fractional Brownian motion with parameter H satisfying:

$$H = \frac{3 - \alpha^*}{2}, \text{ with } \alpha^* = \min(\alpha, \beta, 2). \tag{7}$$

Essentially, this theorem states that, in addition to a linear trend, fractional Brownian motion models the fluctuations of the (cumulated sums of) aggregated traffic time series, at least in the asymptotic limits of infinitely many users and of infinitely long observation duration.

The popularity of the Theorem stems from its mimicking the nature of real traffic, consisting of the superimposition of a large number of flows within which IP packets are emitted. Indeed, its construction in essence amounts to a definition of a fluid traffic model, with constant 'On' activities Z_i standing for independent traffics, such as flows, with random durations τ_{on}. The aggregated traffic $Y_N(t)$ results from their additive superimposition, valid for links without congestion. A schematic realization of the model in terms of network is sketched in Fig. 1.
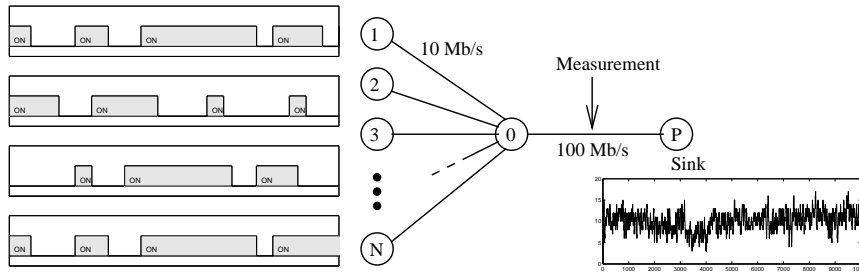
Fig. 1 On/Off model: Network equivalence of the M/G/N reward renewal On/Off model. This model network is used in the NS-2 simulations of Section 6.2.

2.3.2 Reading the Theorem

The correct understanding of the theorem calls for a number of comments. First, Eq. (7) reveals that LRD emerges from the infinite variance nature of the flow size distribution, rather than the power-law behavior of its tail as such. Indeed, should the On/Off duration sizes be heavy tailed with finite variance ($\alpha > 2$ and $\beta > 2$), then $\alpha^* = 2$ and hence $H = 1/2$. Heavy tails with finite variance hence leads asymptotically to ordinary Brownian motion, and consequently to the absence of LRD.

Second, the order of the limits (first infinitely many users, second infinitely long observation duration) plays a crucial role. Very different limit processes are obtained if these are reversed (as detailed for example in [4] and references therein).

Third, Taqqu's Theorem is an asymptotic one, formulated as a limit $T \rightarrow +\infty$. Often, this is misleadingly read as follows: that the self-similar nature of X would be observed when the observation is *long* enough, after some transient has been discarded. The removal of transients is particularly important in numerical simulation details but is another issue entirely. Instead, a proper reading of the asymptotic limit is that that $X_N(tT)$ actually corresponds to FBM $B_H(t)$ only when analyzed $X_N(tT)$ *from far*, *at coarse scales*, *globally*, or when *zooming out*, without paying attention to the details at small scales. Referring to $X_N(tT)$ as a self-similar process is hence confusing, as its distributions are not covariant under dilation of any *fine* dilation factors, but only for *coarse* ones. Consequently, traffic times series correspond much more to LRD than to self-similar processes. This issue is revisited in Section 3.2 in relation to wavelet analysis.

3 Wavelet analysis of LRD

3.1 Wavelet transform

Wavelet transforms [20] have been shown to be effective for the analysis of LRD and hence for the estimation of the Hurst parameter in [21]. In the context of Internet time series analysis, this procedure has been studied and proved efficient in [3,22–24].

Let ψ_0 denote a reference pattern, called the mother-wavelet. It is an oscillating function whose energy is concentrated on a small support both in the time and frequency domains. It is characterized by an integer $N_\psi \geq 1$, referred to as the number of vanishing moments and defined as: $\forall k = 0, ..., N_\psi - 1, \int_\mathbb{R} t^k \psi_0(t) dt \equiv 0$ and $\int_\mathbb{R} t_\psi^N \psi_0(t) dt \neq 0$. Moreover, ψ_0 is chosen such that the $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k), k \in \mathbb{Z}, j \in \mathbb{Z}^{*,+}\}$ form a basis

of $L^2(\mathbb{R})$. The (Discrete) Wavelet Transform (DWT) coefficients of a process X are defined as: $d_X(j,k) = \langle \psi_{j,k}, X \rangle$.

Let Y denote a second order stationary stochastic process. It has been shown (cf. [3,21]) that the variance of the $d_X(j,k)$ mimics the spectral behavior of Y:

$$\mathbb{E}|d_Y(j,k)|^2 = \int_{\mathbb{R}} \Gamma_Y(\nu) 2^j |\tilde{\Psi}_0|^2 d\nu, \tag{8}$$

with $\tilde{\Psi}_0$ the Fourier transform of ψ_0. When Y is a LRD process, this yields an asymptotic power law behavior for $\mathbb{E}|d_Y(j,k)|^2$ in the limit of coarse scales:

$$\mathbb{E}|d_Y(j,k)|^2 \sim C 2^{j(2H-1)} \text{ for } 2^j \to +\infty. \tag{9}$$

Eq. (9) suggests that the analysis of LRD should be conducted in $\log_2 S(j)$ vs. $\log_2 2^j = j$ plots, where the time average reads

$$S(j) = \frac{1}{n_j} \sum_k |d_Y(j,k)|^2, \tag{10}$$

n_j being the number of $d_Y(j,k)$ available at scale 2^j. It can be understood as an estimator for the ensemble average $\mathbb{E}|d_X(j,k)|^2$ [3,21]. Such plots are often referred to as Logscale Diagram (LD) plots, and a weighted linear regression provides an estimation of the Hurst parameter [22,23]:

$$\hat{H} = \frac{1}{2} \left(1 + \sum_{j=j_1}^{j_2} w_j \log_2 S(j) \right), \tag{11}$$

where the weights w_j satisfy the constraints of linear regression: $\sum_{j=j_1}^{j_2} j w_j \equiv 1$ and $\sum_{j=j_1}^{j_2} w_j \equiv 0$. The weights can be written as $w_j = b_j \frac{V_0 j - V_1}{V_0 V_2 - V_1^2}$ with $V_i = \sum_{j_1}^{j_2} j^i b_j$, $i = 0, 1, 2$. The positive numbers b_j can be chosen as the confidence granted to $\log_2 S(j)$. Following [3,21], we use $b_j = 1/n_j$.

In applications involving LRD in real data, it is crucial for practically meaningful analysis to be equipped with confidence intervals. Confidence intervals for $\log_2 S(j)$ (derivations can be found in [22] are used in the LDs, for example Figs 2 and 3), throughout this paper. More generally, the performance of this estimation procedure have been both theoretically and practically quantified in depth and shown to be satisfactory if compared to the best parametric maximum likelihood techniques available (that come with a heavy computational cost), and it also provides extra robustness thanks to the possibility of varying N_ψ [3,22]. In the present context, confidence intervals will be of importance to determine whether the relation predicted by Taqqu's Theorem is observed in practice.

3.2 Taqqu's Theorem in the wavelet world

Let us now use wavelet decomposition to illustrate how the asymptotic limit $T \to +\infty$ of Taqqu's theorem translates in practice. Let $d_{X_N^T}(j,k)$ denote the DWT coefficients of the cumulative process $X_N(tT)$. Stating that $X_N(tT)$ is essentially $B_H(t)$ would imply that the power law behavior of its wavelet coefficients is valid for all scales 2^j: $\mathbb{E}|d_{X_N^T}(j,k)|^2 =$

$\sigma^2 2^{j(2H+1)}$ (cf. [21,23,25]). However let us examine the following calculation (where we set $T = 2^{j_0}$ for simplicity):

$$
\begin{aligned}
d_{X_N^T}(j,k) &= \int 2^{-j/2} \psi_0(2^{-j}t - k)X_N(tT)dt, \\
&= \int 2^{-j/2} \psi_0(2^{-j}t - k)X_N(t2_0^j)dt, \\
&= 2^{-j_0/2} \int \psi_0(2^{-j-j_0}t - k)X_N(t)dt, \\
&= 2^{-j_0/2} d_{X_N^1}(j + j_0, k),
\end{aligned}
$$

yielding

$$
\lim_{T \to +\infty} \lim_{N \to +\infty} d_{X_N^T}(j,k) = \lim_{j_0 \to +\infty} \lim_{N \to +\infty} d_{X_N^1}(j + j_0, k). \tag{12}
$$

This amounts to saying that analyzing $X_N(tT)$, for a given time position t and letting $T \to +\infty$ is equivalent to analyzing the behavior of the wavelet coefficients $d_{X_N^1}(j + j_0, k)$, with $t = 2^{-j-j_0}k$, in the limit of coarser and coarser scales, i.e., $j_0 \to +\infty$. Therefore, because $X_N(tT)$ is essentially $B_H(t)$ only in the limit of large T, the power law behavior of its wavelet coefficients is valid only in the limit of coarse scales: $\mathbb{E}\{d_{X_N^T}(j,k)^2\} \sim \sigma^2 2^{j(2H+1)}$ for $j \to +\infty$. The limit $T \to +\infty$ is a central fact that implies that the equivalence between $X_N(tT)$ and FBM will be observed only in the limit of the largest scales.

From a practical perspective, it is more *intuitive* to perform the wavelet analysis over the aggregated count time series $Y_N(t)$, rather than on the cumulative series $X_N(T)$. The analysis above holds equivalently for $Y_N(t)$ and the practical relation in the wavelet analysis reads:

$$
S(j) = \frac{1}{n_j} \sum_k |d_Y(j,k)|^2 \sim \sigma^2 2^{j(2H-1)} \text{ for } j \to +\infty. \tag{13}
$$

4 Models, numerical simulations and parameter setting

To investigate observability in practice of the $H(\alpha)$ relation, numerical experiments are conducted. Synthetic traffic are generated (in Matlab) consisting of time series of counts based on two models. The first is the superposition of N On/Off sources described in the formulation of Taqqu's Theorem, before any limits are taken. The second is a particular well known limiting form of this superposition (not the fBM limit of the theorem but another), known as the M/G/∞ model. The name derives from the fact that this discrete state space continuous time model is equivalent to that given by the number of active sources in an M/G/∞ queueuing system where each arrival has a dedicated server. For consistency with this notation (but a small abuse of Kendall notation), we will denote the first model by M/G/N. For each of these two models the relation $H = (3 - \alpha)/2$ holds.

4.1 M/G/N On/Off model.

The M/G/N On/Off superposition model is in direct correspondence with the renewal-reward stochastic process from the statement of Taqqu's Theorem. Synthesizing it amounts to generating $N > 0$ traffic sources, with On activity periods of value 1, and Off inactivity periods taking the value 0, whose durations are drawn at random, and independently, from a HT distribution. Pareto distributions, with parameters $(\alpha, \mu_{\text{on}})$ and $(\beta, \mu_{\text{off}})$ respectively, are used.

The count times series $Y_N(t)$ results from the superposition of the traffic generated from these N sources.

The parameters $(N, (\alpha, \mu_{\text{on}}), (\beta, \mu_{\text{off}}))$ are called intrinsic as they depend only on the modeled traffic. The objective is to probe for this simulated reward-renewal process statistical properties predicted by Taqqu's Theorem. Notably, we will study the impacts of varying α in the range 1.1 to 4, and of increasing N.

4.2 M/G/∞ model.

The M/G/∞ model is a queuing system that models arrivals of fluid flows on a server [4, 26]. It differs from the M/G/N model insofar as it assumes *a priori* an infinite number of independent sources, with flow arrival times $\{t_i, i \in \mathbb{N}\}$ distributed according to a Poisson renewal process with intensity λ. The model assumes a single On period for every Poisson source. The On durations $\{\tau_i, i \in \mathbb{N}\}$ are i.i.d. positive RVs. The count traffic time series results from the superimposition of the activities of all flows: $Y_\infty(t) = \sum_{i \in \mathbb{N}} Z_0 \left((t - t_i)/\tau_i \right)$, where Z_0 denotes the indicator function of the interval $[0, 1]$.

When the flow durations $\{\tau_i, i \in \mathbb{N}\}$ are drawn from a α-HT distribution (with mean μ_{on}), a result recalled in [4] and analogous to Eqs. (6) and (7), holds for the cumulated sum $X_\infty(tT) = \int_0^{Tt} Y_\infty(u) \mathrm{d}u$. The $\lambda \to \infty$ asymptotic for the M/G/∞ model plays the role of the $N \to \infty$ limit for the M/G/N model. This is further discussed in Sec. 6.1.

For the M/G/∞ models, λ and $(\mu_{\text{on}}, \alpha)$ are the intrinsic parameters that specifiy the exact traffic conditions that are to be simulated. Both because the shape of its LD is easier to interpret and its numerical simulation is easier and faster, the M/G/∞ model is first analyzed in Sec. 5, while the M/G/N On/Off model is later investigated in Sec. 6.1.

4.3 Numerical simulations and parameter setting

For numerical simulation, the required number of random variables are drawn independently according to the laws mentioned above, using the respective intrinsic parameters of the models. A realization of the process $Y_\infty(t)$ or $Y_N(t)$ is calculated with an aggregation period Δ (defining also the actual sampling time). A long transient D_0 is a posteriori discarded, to ensure the stationarity of the resulting process.

To take into account the asymptotic nature of the theorem, the observation duration D and the aggregation level Δ are introduced as extrinsic parameters that can be controlled. This observation duration D is the parameter that allows an increase of the coarsest scale available for the analysis, hence making it possible to probe for the conditions of observability of the LRD predicted by the Theorem. In numerical simulations, both intrinsic and extrinsic parameters will be tuned so as to understand under which conditions the theoretical $H(\alpha)$ relation can be observed.

In Sec. 5, a complete study of the M/G/∞ process is conducted. Simulations of *nbreal* = 400 independent realizations are made, with model (or intrinsic) parameters: $\mu = 3.2$s, $\lambda = 5$. Parameter α is varied from 1.1 to 4, with 16 different values (the range $\alpha \in (1, 2.2]$ being densely scanned, so as to probe the transition from processes with to without LRD). The observation (or extrinsic) parameters are set to $D = 8 \cdot 10^5$s and $\Delta = 0.1$s. Burn out time is set to $D_0 = 2 \cdot 10^4$s. Such parameters are chosen so as to match as much as feasible those observed in modern network traffic. Scaling and LRD analyses are conducted as detailed in Sec. 3.1, using a standard Daubechies $N_\psi = 3$ mother-wavelet (cf. e.g., [20]).
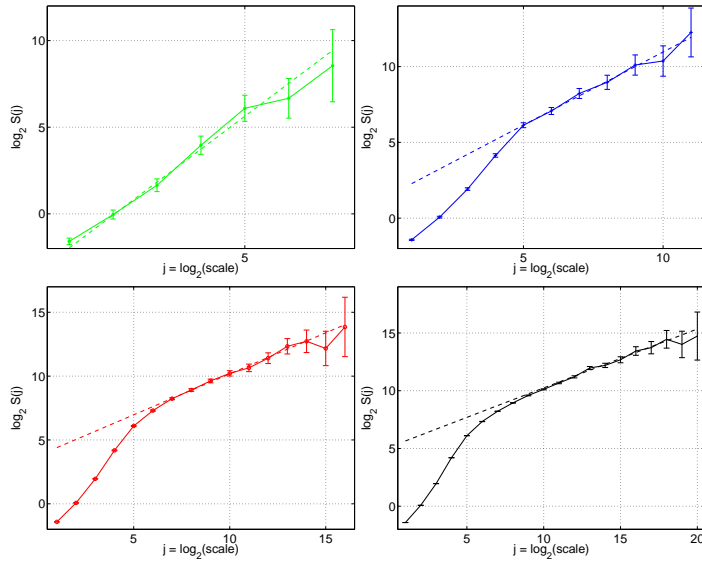
Fig. 2 Empirical Logscale Diagrams. LDs for data simulated from the M/G/∞ model (identical parameters, $\alpha = 1.5$), with 4 different durations (short to long, from top left to bottom right). For all 4 times series, the aggregation level is identical: $\Delta = 0.1$s. For each LD, scaling behaviors are observed, over different ranges of scales and with different scaling exponents. The question is: What are the relations of these different scaling with respect to the heavy tail parameter α ?

In Sec. 6.1, the M/G/N model is explored in a comparable manner. Again, α is changed from 1.1 to 4. The Pareto distributions have parameters $\alpha = \beta$ and $\mu_{\mathrm{on}} = \mu_{\mathrm{off}} = 2$s. Extrinsic parameters are set to: $D = 10 \cdot 10^6$s and $\Delta = 0.1$s (to observe specifically the coarse scales) with a burn out time of $D_0 = 2 \cdot 10^3$s. The number of sources N is 50, 100 or 200. Especially, the study will show how varying the parameter N permits to explore the practical asymptotic nature of the *infinite number of* sources, involved in Taqqu's Theorem.

5 Taqqu's theorem at work on the M/G/∞ model

5.1 Choice of scaling range

To understand the key issues controlling the observation of the connection between heavy tail and long range dependence, we begin with the analysis of a specific toy example.

5.1.1 Toy example

Time series of aggregated traffic are simulated according to the M/G/∞ model, with intrinsic parameters $(\lambda, \mu, \alpha) = (5, 3.2\text{s}, 1.5)$ (the true H is $H = 0.75$), and for 4 different durations chosen in geometric progression, with factor 20 : $D = 10^2, 2 \cdot 10^3, 4 \cdot 10^4, 8 \cdot 10^5$s. The aggregation level Δ is identically set to 0.1s. Fig. 2 shows the LDs calculated from each of these 4 time series.

Examining the top left plot, calculated from the time series with the shortest duration, one observes a satisfactory scaling behavior over the range of available scales $(j_1, j_2) =$
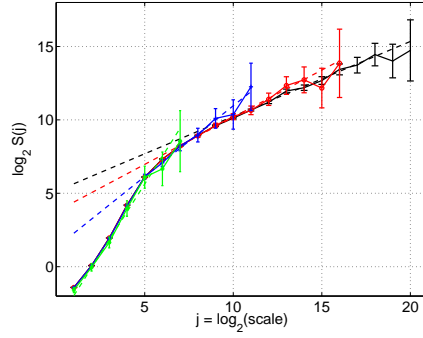
Fig. 3 Empirical Logscale Diagrams. Same LDs as in Fig. 2 superimposed one onto the other. A clear scaling at fine scales $(j_1, j_2) = (1, 6)$ can be observed, together with an asymptotic coarse scales scaling starting above $j = 9$. This latter scaling is related to the LRD phenomenon, while the former is not.

D	$\log_2 D/\Delta$	$\hat{H}$	$\pm \Delta H$	j_1, j_2
$1 \cdot 10^2$	10	1.50	± 0.17	(1,6)
$2 \cdot 10^3$	14	0.93	± 0.09	(6,12)
$4 \cdot 10^4$	18	0.82	± 0.04	(7,15)
$8 \cdot 10^5$	22	0.77	± 0.05	(12,19)

Table 1 Estimated scaling parameter. Estimated H with their 95% confidence intervals for M/G/∞ data, with 4 different durations. The theoretical scaling exponent predicted from Taqqu's Theorem is $H = 0.75$.

$(1, 6)$. Here, *satisfactory* means that a statistical test would not reject the validity of the linear fit and would hence validate the power law behavior of $S(j)$. Such a test has been described in [27] and strongly relies on the fact that confidence intervals, for the $\log_2 S(j)$, are computed along with LDs (Sec. 3). However, the estimate of the scaling parameter, namely the slope of the straight line, yields: $\hat{H}_1 = 1.50 \pm 0.17$ (95% confidence interval, cf. Table 1) in clear disagreement with the theoretical LRD parameter $H = 0.75$. To understand such a severe discrepancy, let us analyze the 3 other LDs, obtained from longer time series. The scaling behavior over the $(j_1, j_2) = (1, 5)$ range remains valid, yet another valid scaling behaviors appear over coarser ranges of scales. Fig. 3 reveals that these scaling superimpose well over each other at large scale and that they reach coarser and coarser scales as duration increases. This validates the asymptotic coarse scale nature of this scaling and hence its correspondence to the LRD phenomenon and hence to the heavy tail. Estimated scaling exponents and scaling ranges are reported in Table 1. It shows that, despite scaling behaviors that any practitioners would accept and that pass goodness-of-fit tests, the estimated H values for the second and third largest duration time series are statistically inconsistent with the value predicted by Taqqu's Theorem. Only the measure obtained from the longest duration time series matches the prediction of Eq. (7).

5.1.2 Theoretical Logscale Diagrams

To further illustrate the difficulties related to the use of LDs to observe practically the relation $H(\alpha)$ (Eq. (7)), LDs are calculated from the *nbreal* M/G/∞ time series, simulated as described in Sec. 4. For each value of α, the *theoretical* LD, $\log_2 \mathbb{E}\{S(j)\}$, is estimated by a Monte-Carlo simulation as $\log_2 \hat{E}_{MC} S(j)$, where $\hat{E}_{MC}$ denote Monte Carlo averages, i.e.,
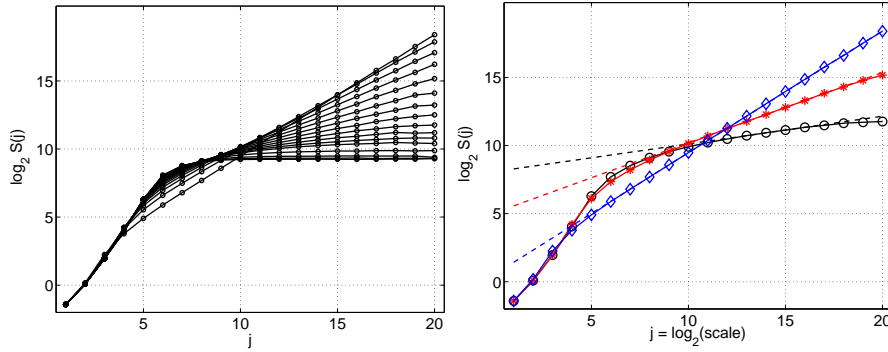
Fig. 4 *Theoretical* **Logscale Diagrams.** Left, superimposed *Theoretical* LDs for 16 different values of $\alpha \in (1,4]$, obtained from Monte Carlo average over *nbreal* copies of the same M/G/∞process. Right, *Theoretical* LDs for $\alpha = 1.1$ (⋄), $\alpha = 1.5$ (∗), $\alpha = 1.9$ (o). The LRD scaling behavior takes on at higher j_1 when α increases within the range $\alpha \in (1,2)$.

means over the *nbreal* independent realizations. These *theoretical* LDs (superimposed in Fig. 4) clearly show that:

i) two valid scaling exist over two different ranges of scales,
ii) the slopes of the fine scales scaling do not depend on α,
iii) the slopes of the coarse scale scaling do vary with α,
iv) the location of the cross-over between these two scaling ranges increases when α is increased within the range $\alpha \in (1,2)$.

5.1.3 Interpretations

The analyses of both the toy set of time series and the *theoretical* Logscale Diagrams yield a clear conclusion: the crucial issue for practically verifying Taqqu's relation between LRD and HT relies on correctly selecting the range of scale devoted to the estimation of H. This issue hides two different pitfalls. When data are too short (only a range of fine scales are given), a statistically perfectly valid scaling behavior may be observed, yet turns out to be related neither to LRD nor to heavy tail. In that case, there is no way in which the relations between H and α can be investigated. The nature of the apparent scaling will be discussed in Sec. 7.3. More intricate is the analysis of data with longer observation durations. For such cases, a clear scaling range appears in the limit of the coarsest (available) scales in consistence with the asymptotic LRD. However, quantitative assessments are needed to check that the data observation duration is large enough for the asymptotic implied in the Taqqu's Theorem be actually reached. In particular, it could prohibit an accurate validation of the theoretical $H(\alpha)$ relation.

The remainder of this section aims at quantifying such comments and concentrates on the quantitative determination of the coarse scale range $[j_1, j_2]$ where the expected relationship between H and α can be found.

5.2 Selecting the coarse scale j_2

The choice of the coarse scale j_2 appears simple at first: the asymptotic formulation of the Theorem seems to indicate that it should be chosen as large as observation duration permits,

i.e., $j_2 = j^D$, with:

$$2^{j^D}\Delta = D. \tag{14}$$

5.2.1 The mystery of missing scales

There exists, however, an extra difficulty, already pointed out in [13,28] from the specific example of LRD induced by a Fractal point process [29] and referred to as *the mystery of missing scales*. Essentially, it consists of a classical difficulty associated to the manipulation of heavy tail random variables (RV): A correct exploration, by means of a *draw at random* procedure, of the farthest part of the tail of the distribution, is made very difficult and unprecise, because of the intrinsic slow hyperbolic decrease of this distribution.

One can quantify this statement, using a classical argument from statistical physics literature [30]: Let M denote the number of random samples drawn independently from a given heavy tail distribution with heavy tail parameter α, then the order of magnitude of the largest observed sample w_M reads: $1 - F_W(w_M) \simeq 1/M$. For a Pareto distribution, this becomes $w_M \simeq \mu(\alpha-1)(M^{1/\alpha}-1)$. In the context of the M/G/∞ model, this shows that the largest flow duration that can be (on average) observed actually depends both on intrisic parameters α, μ, λ and on the (average) number of flows $M = \lambda D$, related to the extrinsic parameter D. An intuitive way to understand Taqqu's Theorem consists of the idea that LRD is created over scales $2^j\Delta$ by the occurrence of flows with typical size $\tau \simeq 2^j\Delta$. This leads to conclude that if no flow with duration larger than w_M occurs, LRD cannot be observed for scales larger than

$$2^{j^M}\Delta = \mu(\alpha-1)((\lambda D)^{1/\alpha}-1), \tag{15}$$

and that, the time series will appear as quasi uncorrelated over larger time scales.

5.2.2 Quantitative study

To quantify the validity of the theoretical prediction in Eq. (15), the set of numerical simulations described above is used. Fig. 5 shows *theoretical* LDs (those shown in Fig. 4, left plot), for $\alpha = 1.1, 1.5, 1.9$, to which the theoretical power law behavior $(2H-1)j$ has been subtracted. This clearly shows that, at coarse scales, plots departs from the desired flat behaviour: For $\alpha = 1.9$, it is clearly visible for $j \geq 17 \simeq j^M = 16.5$, for $\alpha = 1.5$, it occurs for $j \geq 19 \simeq j^M = 19$ while it can not be seen for $\alpha = 1.1$ because of the small observation duration D (hence, $j^D = 20 \leq j^M = 21.5$).

For the range $\alpha \in (1,2)$, careful measurements of the octaves j for which departures from scaling are observed show a striking agreement with the prediction of Eq. (15), and hence corroborate a crucial information: $j^M \sim (1/\alpha)\log_2 \lambda D$ is a (drastically) decreasing function of α (this is illustrated in Fig. 8, dashed black line).

For $\alpha \geq 2$ the analysis, that one might intuitively have expected to be easier as variances are now finite, actually turns out to be more involved. First, the arguments leading to Eq. (15) remain valid. Therefore, at scales $j \geq j^M$, data should appear as quasi-uncorrelated, hence leading to an estimate of H close to $1/2$. Second, finite variance predicts a scaling behavior in the range $j_* \ll j \leq j^M$, with exponent $H = 1/2$. However, there is no reason a priori why those two scaling ranges, $j \geq j^M$ and $j_* \ll j \leq j^M$ should have the same energy level, and transition between two such energy levels would manifest as a range of scales around j^M where the wavelet spectrum is neither flat, nor the signature of any kind of scaling behaviour. Moreover, j^M diminishes when α increases so that the range $[j_*, j^M]$ may become very narrow and even quasi-inexistent, therefore eliminating the possibility that the $H = 1/2$
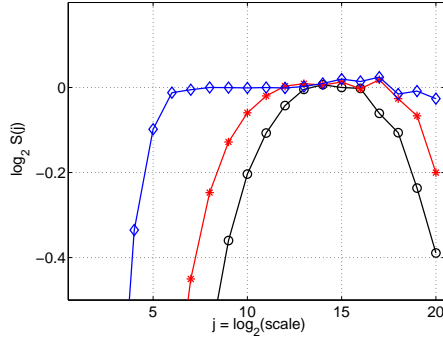
Fig. 5 *Theoretical* **Logscale Diagrams.** For $\alpha = 1.1$ ($\diamond$), $\alpha = 1.5$ ($*$), $\alpha = 1.9$ (o): *Theoretical* LDs to which the estimated linear behavior $(2\hat{H} - 1) \cdot j + \hat{C}$ has been substracted. These plots show clear departure from the LRD scaling behavior for $j \geq j^*(\alpha)$, for $\alpha = 1.9$, (o) with $j^M = 16.5$ and $\alpha = 1.5$, ($*$), with $j^M = 18.5$. For $\alpha = 1.1$, ($\diamond$), $j^M = 21.5$ and hence cannot be observed as $j^D = 20$.

induced by the heavy tail with $\alpha \geq 2$ can actually be observed. In a practical sense therefore, and counter-intuitively, the role of j^M is now reversed: It acts as lower limit for the actual range $j^M \leq j \leq j^D$, where $H = 1/2$ can be measured, though actually corresponding to the missing scale effect, rather than to the manifestation of the HT vs. LRD relation !

Therefore, the *mystery of the missing scales* is an intricate issue: theoretically, it acts as upper limit for the range of scales where HT actually induces LRD scaling. Practically, for infinite variance, $1 < \alpha < 2$, it plays the role of a upper limit, yet, for finite variance, $2 \leq \alpha$, it turns out to play the role of a lower limit when trying to recover the predicted $H = 1/2$.

Our analysis here of the *mystery of missing scales* effect is in the same spirit as that of [13, 28]. The observations regarding the $\alpha > 2$ regime are new.

5.2.3 Conclusion on selecting j_2

The analyses reported in this section indicate that the upper limit j_2 of the scaling range where estimation of the LRD parameter H is to be performed needs to be chosen as the minimum j^* between:

$$
\begin{aligned}
j_2 \leq j^* &= \min(j^D, j^M), \quad \text{when } \alpha \in (1, 2], \\
j_2 \leq j^* &= j^D, \qquad\qquad\quad \text{when } \alpha > 2.
\end{aligned}
\tag{16}
$$

This shows that j_2 does not only depend on the observation duration D (as intuitively expected) but also on α. Fig. 8 indicates that, for a given D, the decrease of j_2 may be drastic when α increases from 1 to 2. Moreover, the coarsest available scale for LRD analysis, j^M, grows asymptotically as $1/\alpha \log_2 D/\Delta$, hence slower than the duration $j^D = \log_2 D/\Delta$. The benefits of increasing D in terms of estimation range extension depend on α.

5.3 Selecting the fine scale j_1

5.3.1 Theoretical analysis

Let us now analyze the arguments entering the choice of j_1, assuming that a large j_2 is set a priori and not changed. Choosing a large j_1 (yet smaller than j_2) enables to better

satisfy the asymptotic requirement and hence to produce less biased estimations of H. However, increasing j_1 also diminishes the number, $j_2 - j_1 + 1$, of octaves involved in the linear regression. Also, confidence intervals of the quantity $\log_2 S(j)$ (over which the linear regression is performed) at coarse scales are larger (than those at finer scales), cf. [22]. These two facts together obviously produce an increase of the size of the corresponding confidence intervals for the estimate of H. Conversely, diminishing j_1 decreases the variance of $\hat{H}$ at the price though of an increased bias given that regression is performed over scales where the departure from the asymptotic scaling behavior becomes significant. Therefore, the choice of j_1 results from a bias-variance trade-off. This has been extensively discussed in a more general context in [3, 22, 24].

Stating that j_1 should be large amounts to choosing it larger than any characteristic scale of time intrinsic to the data or to the model producing them. In the framework of a M/G/∞ model, there exists obviously only two such scales, μ and $1/\lambda$, yielding:

$$2^{j_*}\Delta = \max(\mu, 1/\lambda) \ll 2^{j_1}\Delta. \tag{17}$$

It is important to note that the inequalities defining j_1 (cf. Eq. 17) and j_2 (cf. Eq. 16) are of different natures. For the latter, the scaling hold exactly for all $j \leq j^*$, hence the use of the $\leq$ symbol ; while the former corresponds to an asymptotic behavior, scaling holds when j_1 is much (yet unprecisely) larger than j_*, hence the use of the $\ll$ symbol.

From the discussion above, when $\alpha \geq 2$, the *missing scale effect* acts a lower limit and hence also needs to be accounted for:

$$j_* \ll j_1, \text{ and } j^M \lesssim j_1, \text{ if } \alpha \geq 2. \tag{18}$$

5.3.2 Statistical analysis

We use the set of numerical simulations described in Sec. 4.3 to quantify the bias-variance trade-off that leads to the choice of the optimal j_1. The simulation parameters yield $j_* = 5$ and $j^D = 20.5$. Also, j^M is calculated from Eq. (15) above, where M consists of the actual number of flow duration drawn at random for each realization. For the linear regression range, we set $j_2 = j^*(\alpha)$ (with j^* calculated according to Eq. (16)) while j_1 is varied: $j_1 = 5, 6, ..., 13$. Therefore for each realization $k = 1, ..., nbreal$ and each α, we calculate the estimates $\hat{H}(\alpha, k; j_1, j_2)$ (using Eq. 8). From this set of estimates, biases, variances and mean-square errors (MSE) are measured as:

$$B(\alpha; j_1, j_2) = \hat{\mathbb{E}}_{MC}\left\{\hat{H}(\alpha, k; j_1, j_2) - H(\alpha)\right\}, \tag{19}$$

$$\text{Var } \hat{H}(\alpha; j_1, j_2) = \hat{\mathbb{E}}_{MC}\left\{(\hat{H}(\alpha, k; j_1, j_2)^2\right\} - (\hat{\mathbb{E}}_{MC}\{\hat{H}(\alpha, k; j_1, j_2)\})^2, \tag{20}$$

$$MSE(\alpha; j_1, j_2) = (B^2(\alpha; j_1, j_2) + \text{Var } \hat{H}(\alpha; j_1, j_2))^{1/2} \tag{21}$$

where $\hat{\mathbb{E}}_{MC}$ stands for the Monte Carlo average (average over realizations).

Obviously the bias $B(\alpha; j_1, j_2)$ should decrease when j_1 increases (as can be seen in Fig. 6) while variances increase. Therefore, the MSE necessarily possesses a minimum (as shown in Fig. 7, for various αs), at an optimal choice j_1^o of j_1. The function $j_1^o(\alpha)$ estimated from this procedure is represented in Fig. 8.

For α in the range $1 < \alpha < 2$, examination of the curve $j_1^o(\alpha)$ reveals that the inequality $j_1 \gg j_*$ is much more difficult to satisfy when α is close to 2 than for $\alpha \simeq 1$. For example, in our simulation set up, we observe $j_1^o(1.1) = 6$ and $j_1^o(1.9) = 13$, and hence $2^{j_1^o}\Delta \simeq 2 \cdot 2^{j_*}\Delta$ for $\alpha = 1.1$ while $2^{j_1^o}\Delta \simeq 250 \cdot 2^{j_*}\Delta$ for $\alpha = 1.9$. This is consistent with the LDs shown in
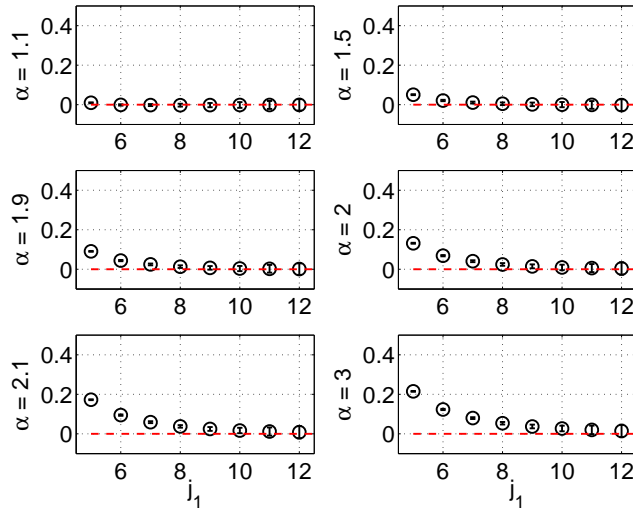
Fig. 6 Bias. Bias of $\hat{H}(\alpha, k; j_1, j_2)$ for fixed j_2 as a function j_1, for different αs, and the Monte-Carlo confidence interval of the bias (small bars inside the circles)
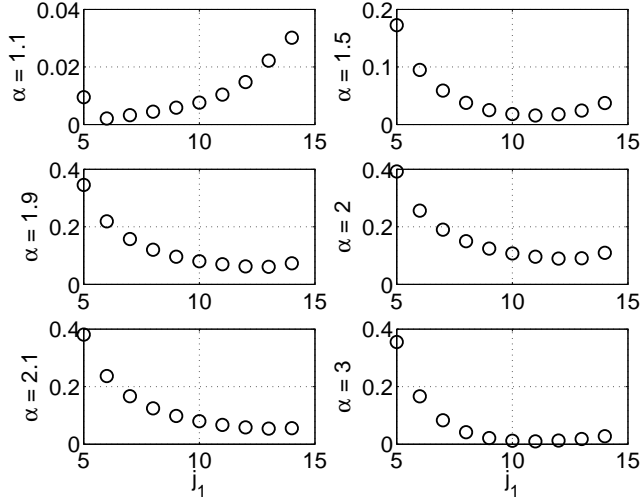

Fig. 7 Mean Square Error. MSE of $\hat{H}(\alpha, k; j_1, j_2)$ for fixed j_2 as a function j_1, for different αs. The minimum of the MSE(j_1) curves yields j_1^o minimizing the bias-variance trade-off.

Fig. 4 where it can obviously be seen that the closer α is to 1, the smaller the j_1 at which the LRD scaling range starts. This is also consistent with our interpretation of the *missing scale* effect. In other words, as α approaches 2, $j_1^o - j_*$ increases notably, a fact of crucial importance for the observability of the $H(\alpha)$ relation.

For finite variance $\alpha > 2$, Fig. 8 shows that the shape of $j_1^o(\alpha)$ matches that of $j^M(\alpha)$. This illustrates that the *missing scale effect* $j^M(\alpha)$ drives the lower limit of the estimation scale range and hence unambiguously validates the analysis of the complex and various roles of $j^M(\alpha)$, reported in Sec. 5.2.2.
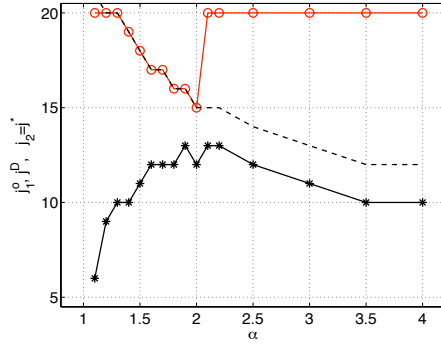
Fig. 8 Scaling range versus α. Functions j_1^o (black solid line with ∗), j^M (dashed black line) and j^* (red solid line with 'o'), with respect to α. The scaling range $[j_1, j_2] = [j_1^o, j^*]$ is narrowing when α increases within $\alpha \in (1, 2)$ as both j_1^o increases and j^* decreases. This explains why the estimation of H turns out to be so difficult for α reaches 2 from below. For $\alpha > 2$, both j_1^o and j^* only weakly depend on α.

Moreover, it is important to note that the bias $B(\alpha; j_1, j_2)$ is a function of the theoretical LD and hence depends on the intrinsic parameters μ, λ, α (any residual dependence observable in our estimated theoretical LD should vanish in the limits *nbreal* $\to +\infty$ and $D \to +\infty$). Conversely, Var $\hat{H}(\alpha; j_1, j_2)$ can be shown to be quasi independent of H (hence of α) and to mostly behave as $2^{j_1}\Delta/D$ (cf. [22, 24]). It is therefore quasi independent of the intrinsic parameters of the model, and mostly varies with the observation duration D. As a result of the bias-variance trade-off, j_1^o depends both on the intrinsic (μ, λ, α) and extrinsic (D, Δ) parameters and cannot hence be determined analytically unless the model underlying the data is precisely known. This is further discussed in Sec. 7.

5.4 Validation of the $H(\alpha)$ relation

Using the same set of numerical simulation as above, H is estimated using a weighted linear regression [3, 22] over the ranges of scales $[j_1 = j_1^o, j_2 = j^*]$, that depend on α, as shown in Fig. 8. The average estimates for H,

$$\hat{\mathbb{E}}_{MC}\left\{\hat{H}(\alpha; j_1^o, j^*)\right\} \pm 1.96\sqrt{\text{Var }\hat{H}(\alpha; j_1^o, j^*)/nbreal}, \qquad (22)$$

together with their Monte Carlo 95% confidence intervals, are plotted in Fig. 9 as a function of α. The $H(\alpha)$ curve predicted from Taqqu's Theorem is superimposed for comparison, showing a very satisfactory agreement: the theoretical $H(\alpha)$ curve is within confidence intervals for all α (but those close to 2). Clearly, for α close to 2, estimation of H is severely impaired by the very narrow scaling range described above. Despite our comfortable parameter settings, for $\alpha = 1.9$, the regression range is as narrow as: $[j_1, j_2] = [13, 16]$, hence yielding large variance and biased estimates for H. For $\alpha \geq 2$, the estimated j_1^o turns out to be smaller than the calculated j^D, yielding a visible bias. Increasing j_1 to j^D reduces the bias (and hence improve agreement between the theoretical $H(\alpha)$ and the average estimated H) at the price though of a substantial increase in the variance of the estimation (not shown).

A careful examination of the literature shows that the estimated curves $\hat{H}(\alpha)$ never satisfactorily match the theoretical form for α around 2. This is largely explained by the aforementioned difficulties in correctly selecting the scaling range, which have often been
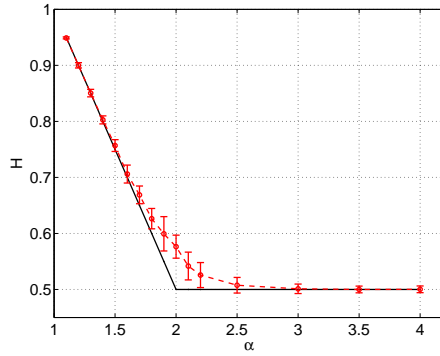
Fig. 9 Relation $H(\alpha)$. Estimates, and their 95% confidence intervals $\hat{\mathbb{E}}_{MC}\hat{H}(\alpha; j_1^o, j_2) \pm \Delta H$ as a function of α (dashed red line with 'o') compared to the theoretical $H(\alpha)$ curve predicted by Taqqu's Theorem (solid black line). Agreement is satisfactory for most values of α (but those close to 2).

very poorly handled. To the best of our knowledge (see however [28]) such an agreement between experimental and theoretical $H(\alpha)$ curves had barely been obtained, neither from real traces nor even from numerical simulations involving Internet time series modeling or network simulations. Here agreement has only been achieved thanks to a careful analysis of the various arguments governing the choice of the range of scales over which estimation is to be performed: the choices of both j_1 and j_2 result from an interplay between intrinsic (or model) and extrinsic (or observation) parameters.

6 Taqqu's Theorem at work: M/G/N model and NS-2 simulations

6.1 Analyses on the M/G/N On/Off model

Let us now return to the original formulation of the Theorem: the M/G/N source described in 2.3. Discussions related to the M/G/∞ model shed light on the asymptotic nature of the Theorem, and have implications on the choice of the scales j_1 and j_2 yielding the correct LRD parameter $H(\alpha)$. The purpose of the current section is to show that, once the meaning of *infinite number of sources* is understood in practice, identical conclusions are drawn both from the M/G/N and M/G/∞ models.

Theoretically, the argument of Sec. 5.2 on the choice of the upper limit j_2 applies: j_2 should be lower than j^M. For a M/G/N process, the average number of samples drawn reads $M = N/(\mu_{on} + \mu_{off})$, and can, hence, be used in Eq. (15) to calculate j^M: $2^{j^M}\Delta = \mu(\alpha - 1)((N/(\mu_{on} + \mu_{off}))^{1/\alpha} - 1)$. The correspondence between the M/G/N and M/G/∞ models stems from the equivalence $\lambda = N/(\mu_{on} + \mu_{off})$.

Let us now turn to the lower limit j_1, resulting from the bias-variance trade-off discussed in Sec. 5.3. In the framework of the M/G/N process, the characteristic scales of time are controlled by μ_{on} and μ_{off} only for the activity of a single source, and by $(\mu_{on} + \mu_{off})/N$ for the source renewal process. The number N of sources being at least 1, the largest characteristic time scale then clearly depends on $(\mu_{on} + \mu_{off})$ alone:

$$2^{j_*}\Delta = (\mu_{on} + \mu_{off}) \ll 2^{j_1}\Delta. \tag{23}$$

Simulations are conducted (in Matlab) in the same manner as for the M/G/∞ process, (cf. Sec. 4.2). Fig. 10 displays the experimental LDs obtained from different α, with the
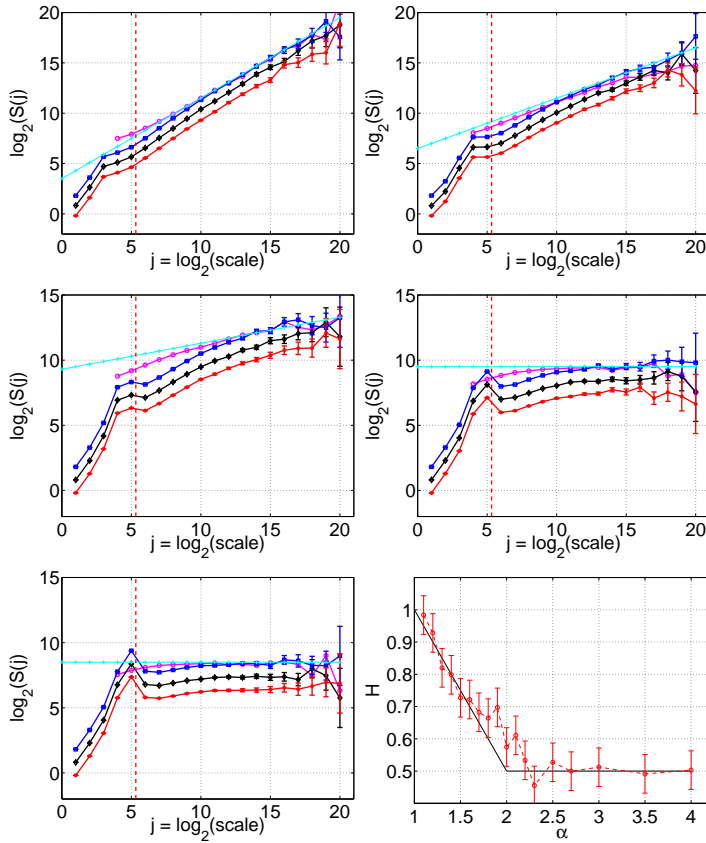
Fig. 10 LDs and $H(\alpha)$ relation for M/G/N. From left to right, top to bottom: $\alpha = 1.2, 1.5, 1.8, 2.5, 3$. $D = 10 \cdot 10^6$s, $\Delta = 0.1s$ and $\mu_{on} = \mu_{off} = 2$s. LDs are obtained from different number of sources: $N = 50$ (star, red), $N = 100$ (square, blue) or $N = 200$ (diamond, black). The vertical dashed (red) line indicates the position of the characteristic scale $j_* = \log_2((\mu_{on} + \mu_{off})/\Delta)$. The Magenta curve (circles) corresponds to the LD obtained from a M/G/∞ process with same parameters. The solid '+' straight lines correspond to the scaling estimated in the appropriate range of range $[j_1, j_2]$ for $N = 200$. Bottom right plot is the empirical $H(\alpha)$ relation obtained in that case, compared to the relation (solid black line) predicted by Taqqu's Theorem.

corresponding j_*. Note that the shift in level is directly attributable to the change of energy, and this is proportional to N: when N is doubled, $S(j)$ is shifted up by 1. The LD of a M/G/∞ process with the same parameters is superimposed for comparison. Notably, λ is chosen to ensure a comparable average number of active flows M, as that for M/G/N with $N = 200$.

When N increases (here from 50 to 200), the LD converges (albeit slowly) to an asymptotic shape. The fine scale range does not evolve much when N increases and remains different from the fine scale range shape of the LD for M/G/∞, where an energy peak associated to the mean periodicity of activity of each source at $j_* = \log_2((\mu_{on} + \mu_{off})/\Delta)$, is clearly apparent for the M/G/N process. This emphasizes the necessity to seek LRD at scales much larger than this j_*. For coarser scales $j \geq j_1$ (as given by (Eq. 23)), the shape exactly matches that of the M/G/∞ model, up to the large cut-off scale j^M. One observes a perfect superposition of the LD of M/G/N with $N = 200$ and of M/G/∞ in the LRD range of scales.
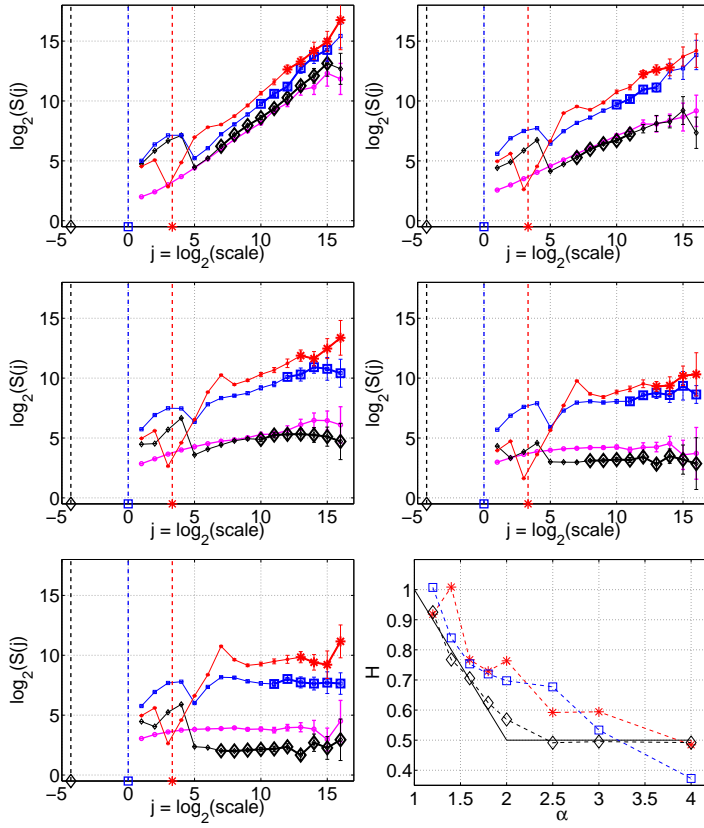
Fig. 11 LDs and $H(\alpha)$ relation for NS-2 simulation. From left to right, top to bottom: $\alpha = 1.2, 1.6, 2, 2.5, 3$. $D = 7200s$, $\Delta = 0.1s$, $N = 50$. Mean flow duration times are $\mu_{on} = \mu_{off} = \{5, 100, 1000\}$ms for (respectively) curves in black (diamond), blue (squares) and red (stars). The vertical dashed line (with corresponding color and symbol) indicates the position of the characteristic scale $j_* = \log_2(2\mu_{on}/\Delta)$. The Magenta curve (circles) corresponds to the LD for a M/G/∞ process with parameters corresponding to the $\mu_{on} = 5$ms NS-2 simulation. Bottom right plot corresponds to the empiral $H(\alpha)$ relation obtained from the NS-2 simulations, when selecting the appropriate range of scales. One sees that only when $\mu_{on} \ll D$ is obtained a large enough range of scales to find an empirical $H(\alpha)$ relation that compares well to the theoretical relation (solid black line).

In conclusion, up to the effect of fine scale model differences (such as the signature of some average periodicity), the observability of Taqqu's Theorem on the M/G/N process follows exactly the same rules as the M/G/∞ process, in terms of the range of scales to be selected and with an equivalence of the respective asymptotic limits $N \to \infty$ and $\lambda \to \infty$. In those conditions, an empirical relation $H(\alpha)$ is obtained (see Fig. 10, bottom right) that compares well to the relation predicted by Taqqu's Theorem.

6.2 NS-2 simulation of On/Off model

Data simulated from traffic time series models do not enable practitioners to explore the potential impacts of actual network mechanisms. To address this issue, in this section we

generate traffic traces using NS-2, a standard event-based network simulator, commonly used to test communication behaviors over networks [31].

We use a configuration inspired from the M/G/N model, with the simple topology depicted in Fig. 1: N computers emit TCP (or UDP) traffic towards a router that transmits them to a bottleneck link, terminated by a sink node. Traffic is collected on this link, is aggregated and analyzed. This network situation closely resembles that considered in [5]. The distributions of activity and inactivity period durations of the sources consist of Pareto laws with parameters α and μ_{on} for active periods, and β and μ_{off} for inactive periods. Alternatively, for inactive periods, exponential distributions (with mean μ_{off}) can be used. Additional parameters are needed to put up a complete network framework. The throughput of each transmitter is set to a constant, and packet sizes are either constant, or uniformly distributed between two limiting sizes. The mean RTT (around 10 ms) is unchanged from one simulation to another. Such hypotheses are quite crude approximations of real traffic distributions. However, they offer sufficient versatility to investigate the impacts of in-flow packet structures driven by actual protocols (TCP, UDP) on the measured LRD.

The simulations replicate the experiments on M/G/N processes. The theoretical discussion at the coarse scale end remains valid. For smaller time scales corresponding to those characteristics of the flow or packet arrivals however, the LD changed. Relevant characteristic scales are μ_{on} (mean flow duration) and the mean interarrival time between packets (for the in-flow structure). With UDP, those are the only relevant time-scales. For TCP flows, the congestion control mechanism adds the RTT and the length of the congestion control window as possible time-scales that intervene in the choice of the optimal j_1.

Fig. 11 displays LDs obtained from different μ_{on}, using TCP flows. It shows how the LD fine scale range is changing when the mean length of the flows is varied, impacting both the coarse scale cut-off, through the *missing scale effect*, and the fine scale cut-off. However, a satisfactory overlap is observed over the coarse scale range (up to the *missing scale* scale) with a M/G/∞ process having same simulation parameters as the NS-2 experiment used here. Simulations (not shown here) show that UDP behaves in the same manner at coarse scales but with a less strigent requirement for the choice of the fine scale limit, due to the absence of a congestion avoidance mechanism. Note that the specific shape at fine scales is driven by a complex interplay between the RTT, the size of the congestion window in TCP, the average duration of the flows; this is not detailed here (see, e.g. [4, 8, 32]).

In conclusion, specific network mechanisms do not change the relevance of the discussion of Sec. 5. What changes is that in-flow network dynamics add additional characteristic time scales that influence j_* and hence lower end of the scale range used in H estimation. Fig. 11 (bottom right) shows that, provided $D \gg \mu_{\text{on}}$, it is possible to obtain an empirical relation $H(\alpha)$ (measured in the appropriate range) that follow closely Taqqu's Theorem.

7 Discussion and comments

7.1 Reading LDs: 3 scale ranges

Here we provide a guide to the reading of LD's arising from traffic time series, be they real or synthetic. The goal here is not to attempt a comprehensive discussion of detailed, modelling issues such as the impact of in-flow packet structure and protocol mechanisms, but rather to describe generic features with a focus on the issue of scale ranges.

Most Internet traffic is subject to two (major) characteristic time scales: a small scale μ_P, corresponding to the interarrival time distribution of IP packets; and a coarse scale μ_F,

related to flow characteristics (e.g., duration). Each are influenced, at least partially, by protocol mechanisms. In the fluid models studied thus far, the parameter μ_F is the only one which relates to flows (for instance, $\mu_F = \mu$ in the case of M/G/∞). These two scales naturally define three scale ranges delineated by octaves $j_P = \log_2 \mu_P / \Delta$ and $j_F = \log_2 \mu_F / \Delta$.

In the limit of coarse scales, $j \to \infty$ (in practice $j \gg j_F$), the LD is characterized by a power law scaling behavior corresponding potentially to LRD, with some H in the range $1/2 \le H < 1$. In the limit of fine scales, $j \to 0$, in practice, $j \ll j_P$, the analysis almost resolves individual packets, and at such scales the arrival process cannot be distinguished from a simple Poisson process. The LD betrays this fact with a power law scaling behavior, characterized by the trivial and non tunable scaling parameter $H = 1/2$. The third range of scales, in practice $j_P \le j \le j_F$, often referred to as the intermediate range of scales, essentially connects these two asymptotic scaling behaviors. There is often no true scaling behaviour existing over this intermediate range. However, as described in [13], there may be a true scaling arising as the signature of sharp discontinuities, which yields $H = 1.5$ (recall Fig. 2 and see also [28]). This is discussed further in Sec. 7.3.

7.2 What can actually be seen on real data ?

Whether the three scale ranges described above can be clearly identified and separated on LDs calculated from empirical data depends on the orders of magnitude of the intrinsic attributes (μ_P, μ_F) in relation to the extrinsic parameters (observed duration D and aggregation level Δ). In the numerical simulations that enable us obtain the empirical $\hat{H}(\alpha)$ curves the duration was set to $D = 8 \cdot 10^5$s, that is 222 hours or roughly 9 days! This is not typically realistic with respect to actual network data. Let us quantify these issues.

Clearly, LRD scaling at coarse scales is visible if and only if $D \gg \mu_F$. According to the literature on Internet traffic metrology, the average estimated mean flow sizes typically range from 10^0s to 10^1s — assuming "flows" are defined based on the canonical IP 5-tupla, i.e. as sequences of packets with the same protocol identifier and address/port end-points. Also, it is common to consider that Internet traffic can be regarded as stationary over time windows of width around 1 to 3 hours, or 10^3 to 10^4s. To examine the implications of this, let us fix $D = 10^3$s, $\Delta = 100$ms, $\mu_F = 3$s, and $\mu_P = 0.1$s. This yields, $j_* \simeq 5$ and $j^* \simeq 13$, implying that for $\alpha = 1.1$ the regression can be performed over the range $[j_1, j_2] = [7, 12]$, while for $\alpha = 1.9$ it narrows to $[11, 12]$, resulting in poor estimate.

The LRD parameter for Internet time series is often measured at around $H = 0.8$, corresponding to $\alpha = 1.4$, and is typically measured over a range $[9, 12]$. This is a rough indication that data (if stationary) collected over 10^3s is sufficient to observe the LRD asymptotic regime. Yet, the quantitative analyses reported in Sec. 5 also clearly indicate that, for $D \simeq 10^3$s, H is overestimated as the asymptotic LRD regime has not yet been reached. They also suggest that, as a rule of thumb, the LRD exponent H is correctly estimated on condition that, whatever the aggregation level, the scales $2^{j_*}\Delta$ and $2^{j^*}\Delta$ differ by at least 4 orders of magnitude,

$$\frac{2^{j^*}\Delta}{2^{j_*}\Delta} \gtrsim 10^4, \qquad (24)$$

i.e., j_* and j^* are separated by at least 13 to 14 octaves. With typical parameters this implies at least from 3 to 5 hour long traces collected under stationary conditions, a challenge for actual operating networks (cf. the experiments conducted on a real Grid platform to validate the LRD/HT relation [33]).

The range associated to the *fine scale range* appears if and only if $\mu_P \gg \Delta$. In the Internet traffic metrology literature, commonly reported values range from 2^{-10}s to 2^{-3}s, which is often not seen in LD plots as practitioners rarely choose small enough aggregation levels Δ. In M/G/∞ simulations, the setting chosen is one of a fluid model (without packets) with $\Delta \sim \mu_{on}$. This explains why the *fine scale ranges* does not appear on the LDs shown in Fig. 4, 10. It barely appears on Fig. 11 because $\Delta = 100$ms is larger than $\mu_P \simeq 5$ms, the mean packet interarrival times of our NS-2 simulations.

The *intermediate scale range* can be seen on the LD for scales $j \sim j_F \gg j_P$, that is when both intrinsic scales are well separated. In such cases, if there is a separation of more than 3 octaves, the *intermediate scale range* are sometimes considered as a scaling range by practitioners, whose exponent is quoted as potentially related to flow shape [8, 32]. However, there is no generic expectation that such a scaling should be obtained.

It is rare to observe all three ranges in a single LD from real data: Indeed, this would require long traces (to see the coarse asymptotic LRD) aggregated at a low level (to see in-flow packet structure), hence resulting in large sample size time series that are difficult to process. For instance, $D = 10^4$s and $\Delta = 10^{-3}$s produces a series with 10^7 points.

7.3 Example LDs from several traffic models

Beyond the simplified yet broadly applicable schematic of 3 *scale ranges* of Internet traffic, a more precise description of LD shape depends on factors such as the precise in-flow packet process. To explore this, we discuss some examples commonly used in Internet modelling.

The simplest packet-level model is to assume a pure Poisson process for the packet arrivals, without any flow structure. This yields aggregated time series with no correlation structure (or *white* spectra) and hence no characteristic scale of time. For that case, the LD has a trivial power-law behavior, valid at all scales, from fine to coarse, characterized by the non tunable scaling parameter $H = 1/2$.

Let us now turn to models introducing hierarchical flow and packet levels. The M/G/∞ or M/G/N models are commonly used for the flow level, and are typically combined with a fluid model for the in-flow packet level (such as the On/Off activities studied here, though more elaborate models have also been proposed). The packet model will dominate structure in the *fine scale range*, e.g., no correlation at fine scales for Poisson packets. The *intermediate scale range* is driven by the flow model. For instance, the M/G/∞ process has an intrinsic flow scale, for example $\mu_F = \mu_{on}$. In the intermediate range of scales, its LD consistently exhibits a satisfactory scaling behavior (see Fig. 3), with a scaling parameter estimated around $\hat{H} \simeq 1.5$, whatever the value of α. This results from a specific feature of the M/G/∞ construction, namely a superimposition of step functions, whose Fourier transform mostly consists of a power law decrease with respect to the frequency, $1/|\nu|^2$, hence yielding $H = 1.5$ (note that using H is an abuse of notation here, given that we have a priori reserved it for LRD which appears at large scales). This feature is not realistic, as it is not found in actual data ([3, 7, 22, 24]). More elaborate in-flow or flow shape models (not using step functions) would result in tunable LD shapes in this intermediate scale range, as studied in [8].

Another approach is to consider the cluster point process (CPP) to model directly the packet/flow hierarchy (see [7] and references therein). It consists first of a renewal process for flow arrivals, with characteristic time scale μ_F. To each flow is associated a *cluster*, corresponding to the packets in each flow, defined via a second renewal process with mean inter-arrival time μ_P and a given flow size distribution. When this distribution is not heavy tailed, the corresponding process is not LRD and the LD is flat (trivial scaling with $H = 1/2$)

in both coarse and fine limits. The levels of the LD (powers of the process) in general differ, giving rise to an intermediate scale range which essentially consists of a smooth transition between them. To introduce LRD, following the spirit of the Taqqu's theorem, the number of packets per flow is modeled as a random variable with a heavy tail distribution (of parameter $\alpha \in (1,2)$). For that case, LRD, with parameter $H = (3-\alpha)/2$ takes on shortly after j_F and the intermediate scaling range smoothly connects the Poisson behavior at small scales to LRD at coarse ones. This has been thoroughly described in [7] and shown to be a physically meaningful as well as accurate model capable of reproducing Internet time series statistics.

7.4 Estimation of the heavy tail parameter

When trying to validate Taqqu's Theorem $H(\alpha)$ relation from real Internet traffic data, there is an additional difficulty that has not yet been addressed: α is unknown and also needs to be estimated from the data. This task is beyond the scope of this present contribution, but we make the following comments.

Estimating the tail index is often based, in the statistical literature, on Hill or Pickand procedures (e.g. [18]) and is non-trivial. It is well known that, counter-intuitively, their performance degrades when α increases from 1 to 2. This is likely to further complicate the validation of the theoretical $H(\alpha)$ for values close to 2 as both H and α are poorly estimated in that zone.

Whatever its value, meaningful estimates require large data sets. Preliminary attempts to estimate α from real data were performed on the publicly available trans-Pacific Internet traffic repository (`mawi.wide.ad.jp`, see [34]). These indicate that traces at least few hours long are required. Finally, we mention that recently a procedure based on the behavior at the origin of the characteristic function of the wavelet coefficients has been proposed which outperforms previous tools [35], and was used in the context of Grid traffic in [33].

8 Conclusion

Taqqu's Theorem is often regarded as a founding result for the modeling of Internet traffic. It relates mathematically the LRD and heavy tails parameters H and α via a stochastic process related to On/Off models. Its popularity stems from the fact that it explains the origin of the (now universally accepted) LRD observed in traffic traces via heavy tails of traffic statistics in a manner that is consistent with network 'physics'. The prominence of the result motivated many researchers to experimentally validate the $H(\alpha)$ relation, both in real traffic, and (more often) in simulations based on popular models. Yet, the literature contains very few successful **quantitative** validations. The goal of this paper was to investigate why this is so.

Our main conclusion is that the crucial issue lies in the need to correctly select the scale range where the asymptotic behavior (implied in Taqqu's Theorem) is actually reached. To quantify this we have proposed a 'reading guide' for wavelet Logscale Diagrams (LDs) centered about 3 scale ranges: fine, intermediate and coarse. We explain that the coarsest scale available for the analysis, $2^{j^*}\Delta$, needs to be at least four orders of magnitude larger than the largest intrinsic scale of time $2^{j_*}\Delta$, characteristic of the data. While j_* is a function on the intrinsic parameters of the data only (especially the in-flow properties), j^* is also a function of the observation duration. The upper limit j_2 of the estimation range of scales can

be set to j^*, but the lower limit j_1 needs to be chosen larger than j_* by some factor, which varies from 2 for α close to 1 to 1000 for α close to 2.

Using M/G/∞ based numerical simulations, extremely long (tens of hours !) synthetic aggregated time series were used which enabled us to observe empirically the $H(\alpha)$ relation in a satisfactory manner. Difficulties for α close (both smaller and larger) to 2 however are stubborn, and we relate these to subtle heavy tail effects. Improving estimation in these cases requires increasing the $j^* - j_*$ gap, which can be achievable either by further increasing the observation duration and/or decreasing μ_P. Furthermore, we point out that improving the agreement between prediction and estimation (bias reduction) for α close to 2 cannot be obtained without paying a price of a huge variance increase, that cannot be accepted practically.

Whatever the chosen model, conclusions would be essentially the same: the details of the in-flow processes, flow arrivals, network and protocol mechanisms impact only the fine and intermediate scale ranges of the LD and hence only control the lower limit of the range of scales where LRD scaling in relation to heavy tail exists. They do not affect the coarse scale range and hence LRD, which, by nature, appears beyond any such characteristic scales.

Given the orders of magnitude taken by traffic parameters and time-scales currently representative of the Internet, the empirical observation of the $H(\alpha)$ relation would require significant effort, notably: i) collecting several hour long traces that are under stable and stationary conditions with (almost) no anomalies, ii) accurately estimating the relevant time scales, iii) making use of the most accurate and recent estimation procedure of α such as that in [35]. Two hour long traces are often the longest observations deemed stationary and used for LRD estimation. Another difficulty is that α cannot be freely varied, so that in practice only few (typically those with H close to 0.8) samples of the $H(\alpha)$ relation can be observed. To overcome this difficulty, real yet controlled traffic, generated on a Grid platform, has been recently used (see [33]).

To summarise, currently the theoretical relation $H(\alpha)$ is mostly beyond practical quantitative observability, and hence cannot be used to predict traffic statistics, in particular the LRD parameter, quantitatively. This however in no ways calls into question the validity of the Theorem itself nor its founding and guiding role in the modern statistical modeling and analysis of Internet traffic.

9 Acknowledgement

References

1. Leland, W.E., Taqqu, M.S., Willinger, W., Wilson, D.V.: On the self-similar nature of ethernet traffic (extended version). ACM/IEEE Transactions on Networking **2**(1), 1–15 (1994)
2. Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V.: Self-similarity through high-variability: statistical analysis of ethernet lan traffic at the source level. IEEE/ACM Trans. Netw. **5**(1), 71–86 (1997)
3. Abry, P., Veitch, D.: Wavelet analysis of long-range dependent traffic. IEEE Trans. on Info. Theory **44**(1), 2–15 (1998)
4. Park, K., Willinger, W.: Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, Inc., New York, NY, USA (2000)

5. Park, K., Kim, G., Crovella, M.: On the relationship between file sizes, transport protocols, and self-similar network traffic. In: Int. Conf. on Network Protocols, p. 171. IEEE Computer Society, Washington, DC, USA (1996)
6. Crovella, M.E., Bestavros, A.: Self-similarity in World Wide Web traffic: Evidence and possible causes. IEEE/ACM Transactions on Networking **5**(6), 835–846 (1997)
7. Hohn, N., Veitch, D., Abry, P.: Cluster processes, a natural language for network traffic. IEEE Trans. Sig. Proc. – Special Issue on Signal Processing in Networking **8**(51), 2229–2244 (2003)
8. Barakat, C., Thiran, P., Iannaccone, G., Diot, C., Owezarski, P.: A flow-based model for internet backbone traffic. In: ACM/SIGCOMM Internet Measurement Workshop, pp. 35–47. ACM Press, New York, NY, USA (2002)
9. Downey, A.B.: Evidence for long-tailed distributions in the internet. In: IMW '01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement, pp. 229–241. ACM Press, New York, NY, USA (2001)
10. Taqqu, M.S., Willinger, W., Sherman, R.: Proof of a fundamental result in self-similar traffic modeling. SIGCOMM Comput. Commun. Rev. **27**(2), 5–23 (1997)
11. Malone, D., Duffy, K., King, C.: Some remarks on LD plots for heavy-tailed traffic. SIGCOMM Comput. Commun. Rev. **37**(1), 41–42 (2007)
12. Ricciato, F.: Some remarks to recent papers on traffic analysis: or the case for public wiki-like platforms for commenting published papers. SIGCOMM Comput. Commun. Rev. **36**(3), 99–102 (2006)
13. Roughan, M., Veitch, D.: Some remarks on unexpected scaling exponents. CCR Online (2007)
14. Adler, R.J., Feldman, R.E., Taqqu, M.S.: A Practical Guide To Heavy Tails. Chapman and Hall, New York (1998)
15. Samorodnitsky, G., Taqqu, M.: Stable Non-Gaussian Random Processes. Chapman&Hall (1994)
16. Lee I.W.C.and Fapojuwo, A.: Estimating heavy-tails in long-range dependent wireless traffic. In: IEEE Vehicular Technology Conference, vol. 4, pp. 2132–2136 (2005)
17. Beran, J.: Statistics for Long-memory processes. Chapman & Hall, New York (1994)
18. Doukhan, P., Oppenheim, G., Taqqu, M.: Long-Range Dependence: Theory and Applications. Birkhäuser, Boston (2003)
19. Embrechts, P., Maejima, M.: Selfsimilar processes. Princeton University Press (2002)
20. Mallat, S.: A Wavelet tour of signal processing. Academic Press (1999)
21. Abry, P., Gonçalves, P., Flandrin, P.: Wavelets, spectrum analysis and $1/f$ processes. In: Wavelets and Statistics, Lecture Notes in Statistics, vol. 103, pp. 15–29 (1995)
22. Veitch, D., Abry, P.: A wavelet based joint estimator of the parameters of long-range dependence. IEEE Trans. on Info. Theory – special issue on "Multiscale Statistical Signal Analysis and its Applications" **45**(3), 878–897 (1999)
23. Abry, P., Flandrin, P., Taqqu, M., Veitch, D.: Wavelets for the analysis, estimation and synthesis of scaling data. In: K. Park, W. Willinger (eds.) Self-Similar Network Traffic and Performance Evaluation. John Wiley & Sons, Inc. (2000)
24. Veitch, D., Abry, P.: A statistical test for the time constancy of scaling exponents. IEEE Trans. Sig. Proc. **49**(10), 2325–2334 (2001)
25. Flandrin, P.: Wavelet analysis and synthesis of fractional brownian motion. IEEE Trans. on Info. Theory **38**(2), 9100–917 (1992)
26. Likhanov, N., Tsybakov, B., Georganas, N.: Analysis of an atm buffer with self-similar ("fractal") input traffic. In: Proceedings of INFOCOM'95, pp. 985–992 (1995)
27. Veitch, D., Abry, P., Taqqu, M.S.: On the automatic selection of the onset of scaling. Fractals **4**(11), 377–390 (2003)
28. Roughan, M., Yates, J., Veitch, D.: The mystery of the missing scales: Pitfalls in the use of fractal renewal processes to simulate lrd processes. In: American Statistical Association and the Institute of Mathematical Statistics (ASA-IMS) conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics. American University, Washington, DC (1999)
29. Lowen, S.B., Teich, M.C.: Fractal-Based Point Processes. Wiley (2005)
30. Bouchaud, J.P., Mezard, M.: Universality classes for extreme value statistics. Journal of Physics A **30**, 7997 (1997)
31. NS-2, The Network Simulator. http://isi.edu/nsnam/ns/
32. Ricciato, F.: On the role of flows and sessions in internet traffic modeling: an explorative toy-model. preprint (2008)
33. Loiseau, P., Gonçalves, P., Dewaele, G., Borgnat, P., Abry, P., Primet Vicat-Blanc, P.: Investigating self-similarity and heavy-tailed distributions on a large scale experimental facility. preprint (2008)
34. Cho, K., Mitsuya, K., Kato, A.: Traffic data repository at the WIDE project. In: USENIX 2000 Annual Technical Conference: FREENIX Track, pp. 263–270 (2000)
35. Gonçalves, P., Riedi, R.: Diverging moments and parameter estimation. J. of American Statistical Association **100**(472), 1382–1393 (2005)