Revisiting the multiscaling hypothesis at medium timescales

C. NARDUZZI[1], P.A. PEGORARO[1], S. UHLIG[2*]

[1]Department of Information Engineering, Università di Padova
via Gradenigo 6/b, 35131 Padova, Italy

[2]Department of Computing Sciences and Engineering, Université Catholique de Louvain
2, Place Sainte-Barbe, B-1348 Louvain-la-neuve, Belgium

claudio.narduzzi@unipd.it, paoloattilio.pegoraro@unipd.it, suh@info.ucl.ac.be

Résumé – Les lois d'invariance d'échelle décrivent l'absence d'une échelle de temps caractéristique particulière contrôlant un processus stochastique. Un consensus existe dans la littérature concernant un modèle du trafic qui serait monofractal Gaussien auto-similaire aux échelles de temps plus grandes qu'un délai aller-retour des segments du protocole de transport TCP. Dans cet article, nous donnons des indices qui laissent penser qu'un comportement multifractal pourrait être présent à ces échelles de temps. Nous présentons des résultats qui diffèrent de la littérature, et prétendons qu'un modèle multifractal stationnaire doit être envisagé pour modéliser le trafic aux échelles de temps typiquement plus grandes que quelques centaines de millisecondes.

Abstract – Scaling refers to the absence of a particular characteristic time scale controlling a stochastic process. The literature on network traffic widely agrees on a monoscaling Gaussian self-similar traffic model at timescales larger than the round-trip time (RTT) of TCP segments. In this paper we show evidence that multiscaling may be present at these timescales. We discuss some scaling results that differ from traditional work and, as a consequence, claim that stationary multiscaling needs to be taken into account when modelling network traffic at timescales typically larger than a few hundred milliseconds.

1 Introduction

In the past decade much research has been carried out on the analysis and modelling of network traffic. Following the discovery of the self-similar structure of network traffic [1], statistical analysis was applied to traffic measurements to obtain meaningful, parsimonious models and identify robust invariants even in a complex, variegated and rapidly changing environment such as the Internet.

Scaling refers to the absence of a particular characteristic time scale controlling a stochastic process. At timescales larger than the round-trip time (RTT) of TCP segments, the literature widely agrees on a Gaussian self-similar traffic model, which has monoscaling behaviour. In this paper we show evidence that multiscaling can be present at these timescales. We shall first review multiscaling analysis of network traffic, then discuss some scaling results that differ from traditional work. To support our claims, we rely on the analysis of well-known traces used in the literature. We first show that all studied traces are compatible with multiscaling behaviour at timescales of a few RTTs. Second, we show that the wavelet coefficients at these timescales are not Gaussian. Third, we show that neither the violation of Gaussianity of the wavelet coefficients, nor the multiscaling features come from biases due to non-stationarity. As a consequence, we claim that stationary multiscaling needs to be taken into account when modelling network traffic at timescales larger than a few RTTs, i.e. typically larger than a few hundred milliseconds.

*S. Uhlig is funded by the Belgian National Fund for Scientific Research (FNRS).

2 Wavelet-based scaling analysis

Multiscaling analysis of network traffic was performed by the application of the classic Abry-Veitch Multiscale Diagram wavelet tool (A-V tool, in short) to the time series of traffic counts $Y_T(n)$. This series contains raw measurement information, obtained by observing packets through a link and aggregating them with time step T.

The A-V tool relies on properties of the wavelet detail coefficients of $Y_T(n)$ to reveal the presence of scaling characteristics. Denoting detail coefficients by $d_Y(j,k)$, where j is the scale index and k the time translation index, the behaviour of their moments is expressed by the following general scaling form:

$$E[|d_Y(j,k)|^q] = C(q)2^{j\alpha(q)} \quad \text{for } j = j_1 \dots j_2. \qquad (1)$$

The indices j_1 and j_2 identify the boundaries of the region considered for scaling analysis.

Through equation (1), the two functions $\alpha(q)$ and $C(q)$ are defined. In this study, the interest is centered on $\alpha(q)$: for a simple mono-scaling process $\alpha(q) = q(H - \frac{1}{2})$ holds over all scales, H being the so-called Hurst parameter[1]; for multiscaling processes the behaviour is more complex as $\alpha(q)$ is no longer linear in the moment order q.

The A-V tool evaluates $\alpha(q)$ by weighted linear regression of the logarithm of moment estimates against log-scale, for each moment order q. For the results presented in this paper, the required confidence intervals of log-moment estimates were always computed from data, as described in [2] and in the code

[1]In this context variations may arise from different definitions of α, different normalization of wavelets and different starting processes, that is, cumulative or increment traffic processes.

[3], so that no a priori assumptions are required on the probability distribution of wavelet coefficients. Along with $\alpha(q)$, the tool also computes the associated confidence intervals.

The regression quality factor Q, calculated using the confidence intervals of the Logscale Diagrams [4], provides information on the quality of alignment for the chosen time scale range (j_1, j_2). Quality of regression seems to be a rather neglected issue, but it should be emphasized [5] that analysis and evaluation of scaling behaviour are meaningless without the evidence for "good" alignment.

Finally, it is important to remind that moment estimation becomes critical when the order q is large, because of the higher sensitivity to outliers and of theoretical limits [6], [7]. In the following analysis, we always consider $q < 10$.

From $\alpha(q)$, the so-called Linear Multiscale Diagram (LMD) $h(q)$ is computed as:

$$h(q) = \frac{\alpha(q)}{q} - \frac{1}{2} \tag{2}$$

LMD presentation of scaling parameters is better suited for a decision between the monoscaling and multiscaling regimes.

3 Multiscaling: a brief assessment

This section is dedicated to a brief discussion of relevant issues in multiscaling analysis of WAN traffic traces. In the literature, traces of varying lengths and from different types of link (in terms of bandwidth and traffic load) have been considered, although these peculiar features have not always been made clear. Tools have also differed, as both time-domain and wavelet-based moment analysis have been used. Perhaps not surprisingly, this resulted in partially heterogeneous or apparently contradictory results. For these reasons, we wish first to comment on basic assumptions and on different choices of timescale ranges, with the aim of avoiding the difficulties mentioned above. A very good summary with similar purposes can be found in [5].

In the first place, multiscaling, which is related to the approach chosen for traffic analysis, needs to be separated from the concept of multifractality. The latter specifically refers to the local behaviour of fractal processes showing a number of different scaling exponents with respect to time. Multiscaling, on the other hand, concerns a range of timescales, not only to the regularity behaviour as scale tends to zero. Furthermore, it is correctly defined also for discrete processes like traffic counts, without approximation. This paper is concerned with multiscaling.

The notion of time scale size should be considered with some care. It is generally accepted that the order of magnitude of TCP round-trip times defines the boundary between small (or fine) scales and large (or coarse) scales, with the former usually referring to durations up to hundreds of ms and time scales of seconds or more being termed large. At large scales self-similarity (or, in this context, monoscaling) is usually taken for granted and seems well rooted into the physical behaviour of the traffic [8, 9]. Multiscaling is more controversial and has been found in WAN network traffic analysis at timescales both larger and smaller than a few hundred milliseconds [10, 11, 12, 13]. However, it may be difficult to see a straightforward relationship to traffic behaviour, because of different packet capturing resolutions and different aggregation levels used when results are reported.

Aggregated traffic counts $Y_T(n)$ are actually the result of two different kinds of aggregation, as noted in [14]: a vertical aggregation originated by the superposition of flows and a horizontal aggregation due to the finite length of the time step. To some extent, the two effects could be considered interchangeable: while flows over high-capacity backbones can easily reach a high degree of vertical aggregation over small timescales, the same can happen for slower flows with longer aggregation time steps. However, some of the mechanisms governing a network (e.g., the retransmission timeout in the TCP protocol) are dependent on time rather than on link bandwidth. This implies that, at least for certain timescales, the suitability of a traffic model may depend on the link features.

From the viewpoint of experimental analysis, as well, the impact of bandwidth and link loading on the estimation of multiscale diagrams needs to be carefully considered. Assuming a generic average packet size of 250 bytes, a fully loaded 100 Mbps link can generate a potential flow of 50,000 pkts/s, which drops to 10,000 pkts/s for a 20% load. However, this yields an average of just 10 packets per time step if T=1 ms. At this aggregation level, wild fluctuations of experimental data are likely to limit the usefulness of a stochastic process model, particularly where small-timescale behaviour is concerned. On the other hand, traffic analysis on a 10 Gb/s backbone, under the same conditions, would yield 1000 packets per time step. This not only means that a continuous flow model is applicable, but also that vertical aggregation is already large. Statistical properties of the underlying process may then be quite different and approach large-scale behaviour even at small timescales [15].

One final problem refers to the issue of non-stationarity. For timescales above RTT a more accurate assessment should further distinguish between medium values, ranging from hundreds milliseconds to minutes, and large values, from minutes to hours and more. In fact, it is well-known that Internet traffic exhibits daily and seasonal non-stationarity, that can affect scaling behavior at very large scales. In recent works attention has shifted to possible effects of non-stationarity at smaller timescales; it has been shown in [16, 17] that even at medium timescales there is a need to check for stationarity in data.

4 Experimental analysis of WAN traffic traces

The approach we adopted for this work is based on the considerations presented in the previous section. We carried out a very careful assessment of all factors that might affect our judgement on experimental results, trying to avoid any preconstructed belief on our data. In this paper we concentrate on the scaling behaviour at medium timescales of links with traffic speeds of the order of a few Mb/s (i.e., not large backbone traffic) and look at durations of about one hour. This allows us to focus on the controversial aspects of multiscaling, under conditions where all the factors discussed above may influence the analysis of traffic processes.

We have considered several publicly available traces from the Internet Traffic Archive [18] and the Auckland IV archive [19]:

- pAug89, containing a million packet arrivals seen on an Ethernet link at the Bellcore Morristown Research and Engineering facility. This trace is LAN traffic (with a small portion of transit WAN traffic).
- dec-pkt (1995): traces containing an hour's worth of wide-area traffic between Digital Equipment Corporation and the rest of the world.
- lbl-tcp-3 (1994), a trace containing two hours' worth of wide-area TCP traffic between the Lawrence Berkeley Laboratory and the rest of the world.
- Auck (2001), a set of GPRS-synchronized long IP-traces captured by the WAND group using the DAG system (only a less than one-hour portion).

Reported timing accuracies are 10 μs for pAug89, 1 ms for dec-pkt, 1 μs for lbl-tcp-3 and better than μs for the Auck trace.

To assess multiscaling we compute a linear regression on the LMD, keeping into account the variances resulting from application of the A-V tool; this yields intercept and slope parameters along with their uncertainties. If the confidence interval of the slope parameter is compatible with a horizontal line (zero slope), we decide for mono-scaling, otherwise for multiscaling.

This analysis revealed the presence of multiscaling in the region that spans timescales from hundreds of milliseconds to minutes for all of the traces listed above, with the exception of the LAN pAug trace used for comparison. For reasons of space limitations, only the Auck trace will be discussed, with traffic data aggregated over a time step $T = 1ms$. In doing so, we shall briefly comment on our criteria to ensure a satisfactory degree of confidence in multiscaling results. Further details can be found in [20].

Although the basic heuristic is reasonable and simple enough, it is by no means definitive. Therefore we investigated several other aspects related to multiscaling to ensure, as far as possible, a coherent picture with no contradictory results.

The first test involved the probability distribution of wavelet detail coefficients at the analysed timescales. Fig. 1 shows the cumulative distribution function (CDF) of wavelet coefficients at scale index $j = 10$ (a timescale of about 1 s), overlayed on a Gaussian CDF with equal mean and variance[2]. Logarithmic scale is used for the vertical axis, so that the Gaussian CDF appears as a straight line. It is apparent that wavelet coefficients are not normally distributed, thus ruling out monoscaling. Distribution tails are heavier than in the Gaussian case because nearly 20% of data, on the two sides, depart from the reference straight line. Then, it can be safely considered that deviation from Gaussianity is not the result of a small number of outliers but an intrinsic characteristic of the distribution.

Doubts might still remain concerning the possibility that correct scaling estimation is affected by non-stationary behaviour in the traffic trace (though observed mean and variance appear to be stationary). We addressed this point by repeating the analysis of the trace using mother wavelets with an increasingly large number of vanishing moments. The rationale is that the order of the polynomial trend the wavelet transform is blind to, increases accordingly. If scaling estimation is affected by polynomial trends, we expect that, progressively, the point will be reached where non-stationarities are mostly filtered out. Then,

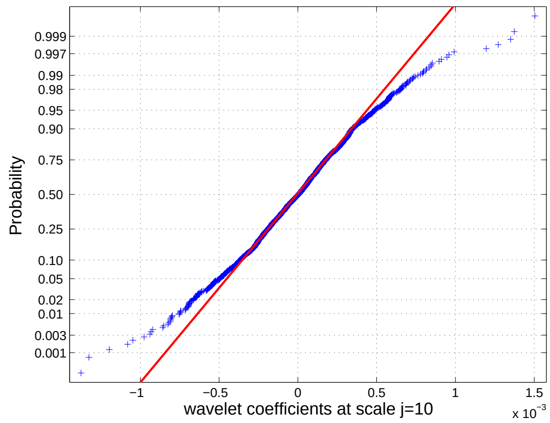

FIG. 1: CDF plot of the Auckland trace wavelet coefficients at scale $j = 10$ compared to a Gaussian CDF (straight line).

any further change of the mother wavelet will have no effect on the results. The outcome of this approach is presented in Fig. 2, which shows that no significant changes occur, in this case, when the mother wavelet has at least 5 vanishing moments. Multiscaling results obtained with this mother wavelet are, arguably, immune from non-stationary trends.

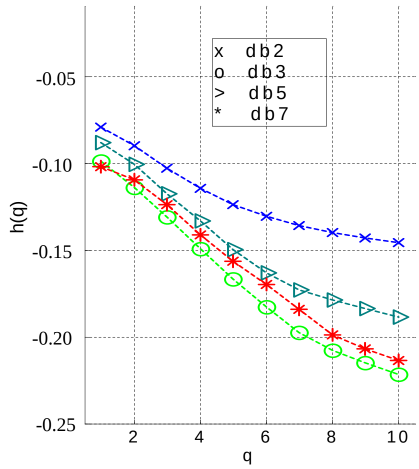

FIG. 2: LMD of Auck trace calculated with different numbers of vanishing moments for the mother wavelet. Confidence intervals are not shown to avoid confusion.

Non-Gaussianity of detail coefficients at medium timescales over the whole observation interval is confirmed by comparisons among distributions of detail coefficients belonging to different sub-intervals. We divided the wavelet coefficient time series into a number of non-overlapping blocks and applied the Kolmogorov-Smirnov (K-S) test to pairs of blocks[3]. This is a statistical test which assumes constant probability distribution as the null-hypothesis and in no case this was rejected on the basis of available data. Finally, we applied statistical tests for normality to each block of wavelet coefficients and obtained the rejection of the Gaussian hypothesis; therefore, we concluded for a non-Gaussian distribution.

[2] Comparison of the two CDFs is akin to the analysis of a quantile-quantile plot.

[3] A significance level $\alpha = 5\%$ was used to be conservative.

We initially devised these tests as yet another way to check for stationarity: in fact, if non-stationarity was at the origin of non-Gaussianity, the K-S test would fail somewhere along the trace. However, this also allowed to address another kind of problem. Although the A-V tool is blind to polynomial trends, some different non-stationarities can affect the results of the estimation [4, 21]. As the estimation of detail coefficient moments involves averaging along time, it might be possible that some localised phenomena are difficult to highlight at small moment orders, but become important at larger q, affecting scaling analysis. The fact that non-Gaussian behaviour is uniformly found all over the observation interval suggests that no such situation occurred. To further analyse this possibility, we plotted detail coefficient behaviour over time for each scale and found out that, for the Auck trace, no such outliers were present.

Putting together the analysis by CDF-plot, statistical test and time diagrams, it follows that non-Gaussianity and multiscaling seem intrinsically linked to the behaviour at medium timescales of traffic traces related to medium capacity network links.

5 Conclusions

Several studies concerned with the traffic count process have argued in favour or against the presence of multiscaling; studies of the process of TCP flow arrivals have also uncovered complex scaling behaviours [17]. While multiscaling at small timescales in the byte count process has been associated with the behaviour of TCP, neither the process of TCP flow arrivals nor the process of byte counts at medium timescales present a reasonable connection to it. The rationale for this statement is that TCP adapts its behaviour at timescales of a few RTTs, typically a few hundreds of milliseconds. This makes unlikely that TCP alone be responsible for the presence of multiscaling at timescales of seconds or more, as we showed in this paper. Multiscaling at medium timescales must then reveal stochastic properties of the signal, not network-protocol related ones. If this is actually the case, then it is on the side of signal processing tools that an answer must be found to discover the cause of this behavior.

References

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, “On the self-similar nature of ethernet traffic (extended version),” *IEEE/ACM Trans. on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[2] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, “Wavelets for the analysis, estimation and synthesis of scaling data,” in *Self-Similar Network Traffic and Performance Evaluation* (K. Park and W. Willinger, eds.), pp. 39–87, Wiley, New York, 2000.

[3] http://www.cubinlab.ee.mu.oz.au/ darryl/.

[4] P. Abry and D. Veitch, “Wavelet analysis of long-range-dependent traffic,” *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[5] N. Hohn, D. Veitch, and P. Abry, “Multifractality in tcp/ip traffic: the case against,” *Computer Networks*, vol. 48, no. 3, pp. 293–313, 2005.

[6] B. Lashermes, P. Abry, and P. Chainais, “New insights of the estimation of scaling exponents,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 497–523, 2004.

[7] P. C. B. Lashermes, P. Abry, “Scaling exponents estimation for multiscaling processes,” in *Proc. ICASSP*, vol. 2, pp. 509–512, May 2004.

[8] M. Taqqu, W. Willinger, and R. Sherman, “Proof of a fundamental result in self-similar traffic modeling,” *Computer Communication Review*, vol. 27, no. 2, 1997.

[9] M. Crovella and A. Bestavros, “Self-similarity in world wide web traffic:evidence and possible causes.,” in *Proceedings of ACM SIGMETRICS International Conference on Measuring and Modeling of Computer Systems*, May 1996.

[10] R. H. Riedi and J. L. Vehel, “Multifractal properties of tcp traffic: a numerical study,” Tech. Rep. 3129, INRIA Roquencourt, March 1997.

[11] P. Mannersalo and I. Norros, “Multifractal analysis of real atm traffic: A first look,” Tech. Rep. COST257TD(97)19, VTT Information Technology, Jan. 1997.

[12] A. Feldmann, A. Gilbert, W. Willinger, and T. Kurtz, “The changing nature of network traffic: Scaling phenomena,” *Computer Communications Review*, vol. 28, no. 2, pp. 5–29, 1998.

[13] A. Feldmann, A. C. Gilbert, and W. Willinger, “Data network as cascades: Investigating the multifractal nature of internet wan traffic,” in *Proceedings of ACM SIGCOMM’98*, pp. 42–55, 1998.

[14] J. Kilpi and I. Norros, “Testing the gaussian approximation of aggregate traffic,” in *Proceedings of 2nd Internet Measurement Workshop*, pp. 49–61, ACM SIGCOMM, Nov. 2002.

[15] Z. Zhang, V. Ribeiro, S. Moon, and C. Diot, “Small-time scaling behaviors of internet backbone traffic: An empirical study,” in *Proc. IEEE INFOCOM 2003*, vol. 3, pp. 1826–1836, Apr. 2003.

[16] S. Uhlig, O. Bonaventure, and C. Rapier, “3d-ld: a graphical wavelet-based method for analyzing scaling processes,” *Proceedings of 15th ITC Specialist Seminar*, Jul. 2002.

[17] S. Uhlig, “Non-stationarity and high-order scaling in tcp flow arrivals: a methodological analysis,” *ACM Computer Communication Review*, vol. 34, no. 2, pp. 9–24, 2004.

[18] Internet Traffic Archive. http://ita.ee.lbl.gov/.

[19] NLANR. http://pma.nlanr.net/Traces/long/auck4.html.

[20] L. Benetazzo, C. Narduzzi, and P. A. Pegoraro, “Internet traffic measurement: a critical study of wavelet analysis,” in *Proc. IEEE Instrum. Meas. Technol. Conf., IMTC/2005*, vol. 3, pp. 2294–2299, May 2005.

[21] S. Stoev, M. Taqqu, C. Park, and J. Marron, “Strenghts and limitations of the wavelet spectrum method in the analysis of internet traffic,” Tech. Rep. 2004-8, SAMSI, Mar. 2004. Research Triangle Park, NC, USA.