Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computer Networks 48 (2005) 293–313

Computer Networks

www.elsevier.com/locate/comnet

Multifractality in TCP/IP traffic: the case against

Darryl Veitch [a,*], Nicolas Hohn [a], Patrice Abry [b]

[a] ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN), Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Victoria 3010, Australia
[b] CNRS, UMR 5672, Laboratoire de Physique, Ecole Normale Supérieure de Lyon, F-69364 Lyon, France

Available online 18 January 2005

Abstract

The discovery of long-range dependence (a kind of asymptotic fractal scaling) in packet data from LANs and WANs, was followed by further work detailing evidence for multifractal behaviour in TCP/IP traffic in WANs. In terms of networking however, physical mechanisms for such behaviour have never been convincingly demonstrated, leaving open the question of whether multifractal traffic models are of black box type, or alternatively if there is anything 'real' behind them. In this paper we review the evidence for multifractal behaviour of aggregate TCP traffic, and show that in many ways it is weak. Our study includes classic traces and very recent ones. We point out misunderstandings in the literature concerning the scales over which multifractality has been claimed. We explain other pitfalls which have led to the multifractal case being overstated, in particular the possibility of 'pseudo scaling' being confused with true scaling, due to shortcomings in the statistical tools. We argue for an alternative point process model with strong physical meaning. It reproduces the higher order statistics of the data well, despite not being calibrated for them, yet is not multifractal. From its standpoint, the empirical multifractal behaviour is seen as a misinterpretation due to a lack of power in the statistical methodology.

Keywords: Multifractal; Estimation; Internet; Traffic; Cluster

1. Introduction

1.1. Motivation

Teletraffic analysis and practice was transformed by the discovery of scale invariance properties in packet traffic [1]. The presence of large-scale asymptotic scale invariance, or long-range dependence (LRD), is remarkably universal, and has

* Corresponding author.
E-mail addresses: dveitch@unimelb.edu.au (D. Veitch), n.hohn@ee.mu.oz.au (N. Hohn), pabry@ens-lyon.fr (P. Abry).

1389-1286/$ - see front matter
doi:10.1016/j.comnet.2004.11.011

become an indispensable part of traffic modelling, in particular for TCP/IP traffic in the Internet. This role is destined to continue, as the phenomena has a physical underpinning which is both generic and readily understandable in networking terms, namely the heavy tailed nature of file sizes [2], which, through a well-known mechanism [3], results in heavy tailed flows and thereby LRD.

The discovery of evidence for *multifractal* behaviour, a richer form of scaling behaviour associated with non-uniform local variability, raised hopes that another 'traffic invariant' had been found which could lead to a complete, robust model of aggregate wide area network (WAN) traffic over all time scales. There is now a literature which accepts the existence of multifractal traffic, exploring alternative multifractal models [4], traffic generators [5], and related performance studies [6]. More broadly, it has become somewhat accepted that traffic has multifractal characteristics, despite the fact that physical mechanisms, and network meaning, has never been established in the way it has for LRD.

In this paper we review the evidence underlying the adoption of multifractal traffic models. We are motivated primarily by two factors arising from our own work in the modelling of TCP/IP packet traffic: (i) the weakness of the evidence seen when using the available statistical tools in a careful way and (ii) a realisation of the lack of statistical *power* of those same tools, leading to the possibility of erroneous interpretation. The question we wish to answer is whether the original enthusiasm for multifractal models was warranted, or is warranted today, when using the default statistical tools (arguably the best available) in a consistent and thorough way. We conclude that the evidence is not only weak but misleading, and that (in most senses) there has been up to now no compelling reason to conclude that a MF model is indicated, or is particularly natural, to describe traffic.

It is not possible for us here to definitively rule on the deeper question of whether traffic *is* multifractal or not, for three reasons. First, the set of available statistical tools are not powerful enough to clarify all the related issues. Improvements are needed in their performance, the knowledge of their performance under different conditions, and important capabilities such as hypothesis tests are absent. Second, ultimately there is no '*is*', modelling data by mathematical processes with multifractal properties reduces to a philosophical issue of model choice, there may always be *some* sense in which a MF model is correct, or rather, useful and/or appropriate (over some scale range). Finally, traffic is an evolving phenomenon, and so conclusions clearly cannot be final in a temporal sense.

After describing necessary background on multifractals and the statistical tools in the remainder of this introduction, Section 2 provides a succinct overview of the key parts of the literature. In Section 3 we compare and contrast the claims of this prior work, and attempt to clarify the causes of the sometimes contradictory claims, particularly with regard to the scale-range over which evidence of multifractality is found. We then offer our own reexamination of the question for several traces, including two of historical importance. Section 4 discusses drawbacks in the existing statistical procedures and tools, and illustrates circumstances where they can be misleading. Through a non-multifractal point process cluster model we recently proposed in [7], Section 5 completes our discussion by combining issues of physical meaning with estimator limitations to decide against multifractal models. This model is greatly preferable to multifractal alternatives on physical and networking grounds. Although not being fitted for the purpose or designed to do so, it can produce multifractal-like statistical signatures which can be *as convincing* as those for the data, although it is not multifractal. We summarise our findings in Section 6.

1.2. Wavelets, scaling and multifractals

It is not possible here to give a detailed introduction to the field of statistical estimation, or the realm of multifractal processes. We provide a concise practically oriented background sufficient to support our presentation. We follow the wavelet viewpoint, first introduced to traffic analysis in [8], and since become the defacto standard, due to its advantageous statistical and computational properties. We use software we developed ourselves (freely available at [9]) to perform the statistical analysis both at second order and at higher order,