FRACTALS IN NETWORKING: MODELING AND INFERENCE

R. RIEDI [‡†], A. KESHAVARZ-HADDAD[†], S. SARVOTHAM[†] AND R. BARANIUK[†]

Departments of Statistics[‡] and of Electrical and Computer Engineering[†], Rice University MS-138, 6100 S. Main Street, Houston, Texas 77005. E-mail: riedi@rice.edu

This paper provides some insight into the causes and implications of the complex small scale dynamics of network traffic loads which are still not fully understood.

Introduction The use of fractal models in computer networking has a strong tradition (see references in [1]). Most prominently, fractional Brownian motion provides a parsimonious abstraction of aggregate traffic at time scales of a second and beyond, which is found useful for design and management and which explains the occurrence of detrimental traffic bursts in terms of *user behavior.* At time scales small enough to be relevant for control, queueing and multiplexing, on the other side, multifractal cascades appear to be more accurate than self-similar models.

Modeling First reports of multifractal scaling in network traffic traces [2] were quickly followed by models based cascading multiplication such as the Binomial cascade and more recently more general iterative products[1]. Being parsimonious and computationally inexpensive these multifractal models proved accessible to analysis and a welcome alternative to the forbiddingly expensive network simulations.

Inference An application as important as simulation is rooted in the fact that networks are stateless and not aware of individual connections. For load balancing and certain rate sensitive applications such as broadcasting, however, a reliable estimate of the available bandwidth is most useful. Inspired by the task of estimating parameters on multiple scales, e.g., the tool *pathChirp* uses efficient trains of exponentially spaced probe packets. When queued, probes are spaced according to traffic load arriving between probes, thus allowing to infer the free capacity.

Alpha-beta decomposition Searching for the causes of multifractal bursts (instances of extreme workload) in traces of aggregate traffic, individual connections with exceptionally large sending rate were found to be the typical culprit, as opposed to a "conspiracy" of an exceptional number of connections as predicted by the classical On-Off model[4]. Only few in number, the high rate *alpha* connections contrast strongly against the large crowd of average *beta* connections, similar to the alpha and beta males in the animal kingdom. Notably, there is strong statistical evidence that alpha connections tend to occur over paths of short response time[5]. As a major conclusion, the alpha-beta decomposition of traffic points to the heterogeneity in the *network topology* as the main cause for the multifractal bursts. Moreover, with only a few connections being potentially harmful and important to monitor, operating a network with relevant state information becomes feasible.

Alpha-beta traffic modeling First, let us revisit the celebrated On-Off model for network traffic [4], a process introduced very early by B. Mandelbrot. In this framework, a traffic source is modelled as being sending traffic at a constant rate (the On state), or as being silent (the Off state) with heavy tailed durations of the On and Off states. It is well known that in the limit of an infinite number of sources the aggregate (sum) of such sources becomes a Gaussian process with LRD, converging to fractional Brownian motion at large time scales. On the other hand, the aggregate of a fixed number of On-Off sources converges to Levy stable motion in the limit of infinitely fast clock time.

Queuing of alpha-beta traffic With response time being the "clock" of the transport protocol, Levy motion seems thus one appropriate choice of a model for the alpha component of the traffic. The large crowd of beta connection with roughly equal sending rates and "clocks" aggregate convincingly to fractional Brownian motion. Assuming the scaling parameter H is equal for the Levy and Brownian traffic components, this sum is again self-similar and allows to apply scaling techniques from queuing theory [3] as well as the concept of critical time scales which is borrowed from large deviation theory. Both agree in predicting Pareto queue tail probability, which is considerably worse than today's Weibull tails of self-similar queues[3].

Alternatively, we may pose the contributions of alpha connections not as a Levy stable motion but simply as one On-Off source with particularly large rate. A natural approach to queueing is then to consider the alpha source as reducing the link capacity. The somewhat involved analysis of the resulting variable-service-rate queue predicts that the presence of alpha connections has little influence except when their rate is larger than the bandwidth available with the beta background traffic in which case the queue tail probability will be Pareto.

Conclusion In measured traffic traces, alpha connection can easily subsume half of the bandwidth; nevertheless, they appear to operate at a constant yet high rate. Thus, the On-Off burst model is a closer approximation to the current state than the self-similar burst model. The above analysis, however, points out the catastrophic consequences of the "what-if-scenario" where alpha sources become sufficiently powerful and aggressive to subsume more than the available bandwidth on a link leading to Pareto laws with extremely large waiting times and surprisingly high packet drop probabilities.

Acknowledgments Support came in part from DARPA NMS F30602-00-2-0557 and through NSF ANI-00099148.

References

1. P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. The multiscale nature of network traffic: Discovery, analysis, and modelling. *IEEE Signal Processing Magazine*, 19(3):28–46, May 2002.
2. J. Lévy Véhel and R. Riedi. *Fractional Brownian Motion and Data Traffic Modeling: The other End of the Spectrum*, chapter 15, pp 185–202. Fractals in Engineering. Springer, 1997, eds. Lévy Véhel, Lutton, Tricot.
3. I. Norros. *Four approaches to the fractional Brownian storage.* pp 154–169, Fractals in Engineering 1997, eds. Lévy Véhel, Lutton, Tricot.
4. M. Taqqu and J. Levy. *Using renewal processes to generate LRD and high variability.* In: Progress in probability and statistics, E. Eberlein and M. Taqqu eds., volume 11. Birkhäuser, Boston, 1986. pp 73–89.
5. X. Wang, S. Sarvotham, R. Riedi, and R. Baraniuk. Network traffic modeling using connection-level information. *Proc. SPIE2002 ITCom, Boston, MA.*