Accurate Second-Order Moment Multifractal Traffic Modelling

Jeferson Wilian de Godoy Stênico and Lee Luan Ling

School of Electrical and Computer Engineering State University of Campinas - UNICAMP Av. Albert Einstein 400, Campinas, SP, Brazil Emails: *jeferson, lee* @decom.fee.unicamp.br

Abstract. In this paper, we study two second-order moment multifractal traffic models and evaluate which one offers better modern network traffic modeling for a given input network traffic trace. Two multifractal process models have traffic arrival loads with the Lognormal and Pareto distributions. The adopted evaluation procedure is based on two performance measures: empirical traffic arrival load distribution and loss probability at connection. Experimental results carried out on both real network traffic traces and synthetically generated ones have validated our approach.

Resumo. Neste artigo, nós estudamos dois modelos de tráfego multifractais com momento de segunda ordem e avaliamos qual deles apresentam melhor modelagem de tráfego nas redes modernas para um determinado traço de tráfego. Os dois modelos de processos multifractais possuem carga de chegadas de tráfego com distribuições Lognormal e Pareto. O processo de avaliação adotado está baseado em duas medidas de desempenho: distribuição empírica da carga de chegada do tráfego e probabilidade de perda em conexões. Os resultados experimentais realizados em ambos os traços de tráfego reais e aqueles gerados sinteticamente validaram a nossa abordagem.

1. Introduction

Since the publication of the work of Leland et. al. [Leland et. al. 1994], there has been an intensification of research on network traffic involving the theory of fractals. Using Ethernet traffic collected in the network of Bellcore Morristown Research and Center Engineering, Leland at. al. had demonstrated that traffic traces of modern high speed data networks exhibit fractal properties, such as self-similarity and long-range dependence (LRD). It was found that these properties, especially the long-range dependence, have a strong influence on network performance [Norros 1994], however not being adequately modeled by Poisson processes or more generically, Markov models.

In fact, the long-range dependence is an important characteristic of traffic, and has relevant implications in diverse issues such as queuing theory and network design. The heavy tail distribution of the duration or length of sessions or connections that originates from the aggregated traffic is identified as the cause of observed self-similar characteristics [Park and Willinger 2000][Crovella and Bestavros 1997]. Thus, several mathematical models were proposed with the aim of better representing the self-similar characteristic found on network traffic. Particularly the fractional Brownian traffic (fBt) model became widely used due to its mathematical simplicity and the capacity of incorporating the features observed in the self-similar traffic [Park and Willinger 2000].

However, it has been observed that, on the time scales of the order of hundred milliseconds and more, the traffic behavior was well represented by self-similar models, whereas in smaller time-scales, self-similar models cannot effectively match real traffic characteristics. This finding has led the search for more comprehensive traffic models, in order to obtain a more faithful description of the network traffic. The multifractal processes arise as a generalization of self-similar processes, in order to providing better description of varying self-similarity characteristics in different time scales as well as the highly irregular behavior of network traffic. Many different multifractal traffic models have been proposed. Most and widely studied ones include: MWM [Riedi et. al. 1999], AWMM [Vieira and Ling 2009], Multiscaling Models with Lognormal [Stenico and Ling 2010] and Pareto [Stenico and Ling 2011] distributed traffic loads, and VVGM [Krishna et. al. 2003].

Traditionally traffic modeling problems were solved based on best traffic statistical propriety fitting, whether traffic has multiscaling properties or not. However, we strongly believe a robust traffic modeling method should consider both traffic's static and dynamic characteristics, where latter are strongly impacted and driven by network dynamics. This work can be viewed as our very first attempt for robust network traffic modeling in which we evaluated two multiscaling traffic models with Lognormal and Pareto distributed traffic loads and decide which one best models an input real traffic trace. The adopted criterion is based on two measures: empirical traffic arrival load distribution fitting and evaluation of loss probability at connection.

The paper is organized as follows. In Section 2, we present the definition of the multifractal traffic processes, review some their concepts and analyzing the characteristics of the second-order statistical moments. In Section 3, we showed the expressions proposals for the loss probability estimation based on queuing theory. In Section 4 we present our experimental investigation. Finally in Section 5 we conclude.

2. Definitions of Multifractal Processes and Second-order Moments

Let X(t) be the traffic rate at t. Then $W(t) = \int_0^t X(t) dt$ will be the arriving load up to t. Denote by $V(t, \Delta t) = W(t + \Delta t) - W(t)$. The average traffic rate is $\lambda = \lim_{\Delta t \to \infty} V(\Delta t) / \Delta t$.

Let μ and σ^2 represent the mean and variance of $V(\Delta t)$. Given T > 0, the accumulative process W(t) is said to be a multifractal process at time scale T if all of the following condition are satisfied:

i) W(t) has a stationary increment at time scale T, i.e., V(t, T) = V(t).

- ii) μ (mean) and σ^2 (variance) of V(t), satisfy the following conditions:
 - ii.a) $\mu = \lambda T$.

ii.b) There exist an integer M > 0, a set $A = \{\beta_i(T): 0 < \beta_i(T) < 1, i \le M\}$, a set

 $\Phi = \{\phi_i(T): 0 < \phi_i(T) < 1, i \le M, \sum_{i=1:M} \phi_i(T) = 1\}, \text{ and a small constant } \varepsilon > 0 \text{ such that for any } \tau \in \{\tau: T - \varepsilon < \tau < T + \varepsilon, \tau > 0\} \text{ such that}$

$$\sigma^2 \sim \sum_{i=1}^{M} \phi_i(T) \tau^{2\beta(T)} \tag{1}$$

The moments of the first and second orders, given by conditions (ii.a) e (ii.b), respectively, have strong influence on the behavior of multifractal traffic. It is observed that the variance (second order moment) of multifractal processes is related to the Hölder exponent $\beta(t)$. The Hölder exponent of a time process at a particular time instant t_0 is related to the regularity level of the signal at that instant [Daoudi et. al. 1998]. Here we assume that the Hölder exponent of a process has a normal distribution in time scale *T*, i.e., $N(\tilde{\alpha}, \tilde{\sigma}^2)$ with mean $\tilde{\alpha}$ and variance $\tilde{\sigma}^2$ [Liu and Baras 2003]. Therefore, the variance of the distribution σ^2 of the traffic process at time scale *T* can be represented as:

$$\sigma^{2} \sim \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}} \exp\left[-\frac{(\beta-\tilde{\alpha})^{2}}{2\tilde{\sigma}^{2}}\right] T^{2\beta} d\beta$$
(2)

Let $z = T^{2\beta}$, then $\beta = \ln(z)/(2\ln(T))$ and $d\beta/dz = dz/(2\ln(T)z)$. Then Eq (2) becomes

$$\sigma^{2} \sim \int_{0}^{\infty} z \frac{1}{\sqrt{2\pi}(2\ln(T)\tilde{\sigma})z} exp\left[-\frac{\left(\ln(z) - (2\ln(T)\tilde{\alpha})\right)^{2}}{2(2\ln(T)\tilde{\sigma})^{2}}\right] dz$$
(3)

The right hand side of Eq. (3) shows that σ^2 simply has a log-normal distribution $L(\omega, \theta)$ with parameters $\omega = 2\ln(T)\tilde{\alpha}$ and $\theta = (2\ln(T)\tilde{\sigma})^2$.

Under the log-normal distribution of σ^2 , it can be shown immediately that [Liu and Baras 2003]:

$$\sigma^2 \sim exp[2ln(T)\tilde{\alpha} + 2(ln(T)\tilde{\sigma})^2] = T^{2\tilde{\alpha}}T^{2\tilde{\sigma}ln(T)}$$
(4)

3. Loss Probability Estimation for Multifractal Processes

We considered two second-order moment multifractal traffic models in this work. The first model considers that the cumulative traffic load V(t) in the period $[t, t_0]$ has Lognormal distribution [Stenico and Ling 2010], while the second model, Pareto distribution [Stenico and Ling 2011].

A. Multifractal traffic Model 1

Suppose that V(t) has a Lognormal distribution

$$f_{V(t)}(x) = \frac{1}{\sqrt{2\pi}x\theta} e^{-\frac{(\ln(x)-\omega)^2}{2\theta^2}}$$
(5)

The distribution parameters ω and θ can be determined by the knowledge of the mean μ and variance σ^2 of the process V(t). Thus, we can write down μ and σ^2 of the log-normal distribution in function of the distribution parameters ω and θ as:

$$\mu = \exp(\omega + \theta^2/2) \tag{6}$$

and

$$\sigma^{2} = \exp(2\omega + \theta^{2})[\exp(\theta^{2}) + 1]$$
(7)

Conversely, we have

Anais

$$\omega = ln(\mu) - \frac{1}{2}ln\left(\frac{\sigma^2}{\mu^2} + 1\right)$$
(8)

and

$$\theta = \sqrt{\ln\left(\frac{\sigma^2}{\mu^2} + 1\right)} \tag{9}$$

In this model, instead of a normal distribution, it is convenient using an exponential function f(t) = aexp(bt) to describe the behavior of the variance σ^2 under time scale T given by equation (4), thus [Stenico and Ling 2010]:

$$\omega = ln \left[\frac{\lambda T}{(k/\lambda^2)aexp(bt)+1} \right]$$
(10)

$$\theta = \sqrt{\ln((k/\lambda^2)aexp(bt) + 1)}$$
(11)

where *k* is a finite constant.

We assume that the single queue is stable with *buffer* capacity sufficient to accommodate eventual transient bursts. Using the result derived in [Benes 1963], the fully characterized queuing behavior of eventually any traffic type in term of information loss can be calculated by:

$$P_{loss}(t) = \int_{Ct+q}^{\infty} f_{V(t)}(v) dv + \rho \int_{0}^{t} f_{V(u)}(v) |_{v=Cu+q} du$$
(12)

The first term on the right side of (12) for the traffic model 1 can be further expressed in detail as:

$$P_{abs}(t) = \int_{Ct+q}^{\infty} f_{V(t)}(v) dv = \frac{1}{2} - \frac{1}{2} erf\left[\frac{\ln(Ct+q) - \omega(t)}{\sqrt{2}\theta(t)}\right]$$
(13)

Thus, the overall loss probability under the stationary state is:

$$P_{p}(t) = \lim_{t \to \infty} P_{loss}(t) = \rho_{t}^{sup} \left\{ \int_{0}^{t} f_{V(u)}(v) |_{v=Cu+q} du \right\}$$
(14)

or

$$P_p(t) = \left(1 - \frac{\lambda}{c}\right) \int_0^\infty \frac{1}{x\theta\sqrt{2\pi}} e^{-\frac{(\ln(x) - \omega)^2}{2\theta^2}} |_{x = Cu + q} du$$
(15)

Therefore, replacing the equation (10) and (11) in (15) we have the loss probability estimation expression based on the lognormal distribution assumption proposed in [Stenico and Ling 2010]

$$P_{p}(t) = \left(1 - \frac{\lambda}{c}\right) \int_{0}^{\infty} \frac{exp\left[-\frac{\left[\ln\left((Ct+q)\sqrt{(k/\lambda^{2})aexp(bt)+1}\right) - \ln(\lambda t)\right]^{2}\right]}{2\pi \ln\left((k/\lambda^{2})aexp(bt)+1\right)}\right]}{\sqrt{2\pi \ln\left((k/\lambda^{2})aexp(bt)+1\right)(Ct+q)}} dt$$
(16)

B. Multifractal Traffic Model 2

In this modeling V(t) has a Pareto distribution

$$f_{V(t)}(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}} \tag{17}$$

where $\mu = \frac{\alpha k}{\alpha - 1}$ and $\sigma^2 = \left(\frac{k}{\alpha - 1}\right)^2 \left(\frac{\alpha}{\alpha - 2}\right)$ are mean and variance of V(t), respectively.

The distribution parameters α and k can be determined by the knowledge of the μ and σ^2 of the process V(t).

In this model [Stenico and Ling 2011], the function exponential was used for describe the relation between the square mean and the variance under time scale T, i.e., $\mu^2/\sigma^2 = aexp(bt)$, thus we have:

$$\alpha = \frac{\mu^2}{\sigma^2} = aexp(bt) \tag{18}$$

and

$$k = \mu - \frac{\sigma^2 \mu}{\mu^2} = \mu - aexp(bt)^{-1}\mu$$
(19)

or

$$\alpha = \frac{\mu^2}{\sigma^2} + 2 = aexp(bt) + 2$$
(20)

and

$$k = \frac{\mu^3 + \sigma^2 \mu}{\mu^2 + 2\sigma^2} = \mu \left(\frac{aexp(bt) + 1}{aexp(bt) + 2}\right)$$
(21)

Using again the results derived in [Stenico and Ling 2011] and [Benes 1963], we have:

$$P_{abs}(t) = \int_{Ct+q}^{\infty} f_{V(t)}(v) dv = \left(\frac{k}{x}\right)^{\alpha} \text{ for } x \ge k$$
(22)

Thus, the loss probability under the stationary state assumption is:

$$P_{p}(t) = \lim_{t \to \infty} P_{loss}(t) = \rho_{t > 0}^{sup} \left\{ \int_{0}^{t} f_{V(u)}(v) |_{v=Cu+q} du \right\}$$
(23)

or

$$P_p(t) = \left(1 - \frac{\lambda}{c}\right) \int_0^\infty \frac{\alpha k^\alpha}{x^{\alpha+1}} \Big|_{x=Cu+q} du$$
(24)

Note that for multifractal traffic series the variables α and k can be calculated using equations (18) and (19) or (20) and (21), respectively. Substituting the relations given by the equations (18) and (19) into (24), the loss probability can be estimated by [Stenico and Ling 2011]:

$$P_{p}(t) = \left(1 - \frac{\lambda}{c}\right) \int_{0}^{\infty} \frac{\left(\frac{\mu^{2}}{\sigma^{2}}\right) \left(\mu - \frac{\sigma^{2}\mu}{\mu^{2}}\right)^{\frac{\mu^{2}}{\sigma^{2}}}}{(ct+q)^{\frac{\mu^{2}}{\sigma^{2}+1}}} dt$$
(25)

Now substituting the relations given by the equations (20) and (21) into (24), the loss probability can be estimated by:

$$P_p(t) = \left(1 - \frac{\lambda}{c}\right) \int_0^\infty \frac{\left(\frac{\mu^2}{\sigma^2} + 2\right) \left(\frac{\mu^3 + \sigma^2 \mu}{\mu^2 + \sigma^2}\right)^{\frac{\mu^2}{\sigma^2} + 2}}{\frac{\mu^2}{(Ct+q)\sigma^{2^+3}}} dt$$
(26)

4. Experimental Evaluation

Our multifractal traffic modeling methodology is composed of the following steps:

- (1) Data Measurement and Empirical Statistical Parameter Estimation This step consisting in traffic data acquisition and data parameter calculation; more precisely, definition of time resolution, estimation of μ (mean) and σ^2 (variance) of V(t), and V(t) probability function fitting, and calculation of mean $\tilde{\alpha}$ and variance $\tilde{\sigma}^2$.
- (2) Estimation and Parameterization of the Model For each multifractal traffic model, we estimate the distribution parameters ω and θ of V(t) and calculation of mean $\tilde{\alpha}$ and variance $\tilde{\sigma}^2$ (the Normal distribution parameters of the Hölder exponent at time scale *T*).
- (3) Performance Evaluation The accuracy of multifractal traffic modeling is performed based on two performance measures: accuracy of arrival traffic load distribution fitting and loss probability at connection.

For comparison purpose, we also compared with two other widely known multifractal traffic modeling methods. The first model was proposed by Ribeiro et. al. [Ribeiro et. al. 2000] who developed an analysis of queue multi-scale for models based on multifractal cascades in a non-asymptotic approach. For this process modeling the following approximation for loss probability was derived:

$$MSQ(b) = P[Q < b] \coloneqq 1 - \prod_{i=0}^{n} P[K_{2^{n-i}} < b + C2^{n-i}]$$
(27)

where K indicates the series of traffic, n is the number of scales considered and C constant service capacity. This approach is named as the Analysis Multiscale Queuing (MSQ), which incorporates the distribution of data traffic across multiple temporal resolution (not only the second order statistics).

Another approach, similar to the MSQ, is called CDTSQ (Critical Dyadic Time-Scale Queue), where the loss probability can be estimated as:

$$CDTSQ(b) = P[Q < b] \coloneqq P[K_{r_D^*} - Cr_D^* > b]$$

$$(28)$$

Note that the approach CDTSQ comes of the concept Critical Time-Scale – CTS defined as:

$$\mathbf{r}^* = \operatorname{argsup}_{\mathbf{r} \in \mathbb{N}} \mathsf{P}[\mathsf{K}_{\mathbf{r}^*} - \mathsf{c}\mathbf{r}^* > b]$$
⁽²⁹⁾

where $r = 2^m$ for $m \in (0, ..., n)$. More detail see [Ribeiro et. al. 2000].

In our experimental investigation we used the Simpson numerical method for calculating the proposed expression for loss probability estimation. We tested in simulation some real traffic traces TCP/IP: *lbl-tcp-3*, *lbl-pkt-4*, *3-7-I-1* and *4-7-I-9*, available at [http://ita.ee.lbl.gov/html/traces.html]. We also generated several synthetic

series, this series were randomly generated following a specific distribution and for comparison we consider the Lognormal and Pareto distributions. For each series used, we get its process statistics such as mean, variance, under server capacity $C = 5 \times 10^5$ bytes/s, and varying buffer size q.

Table I shows some statistical information (means and variances) of traffic traces used.

Traffic Trace	Mean	Variance
lbl-tcp-3	$1.74 \text{ x } 10^4$	3.39 x 10 ⁸
lbl-pkt-4	3.64×10^3	$1.70 \ge 10^7$
3_7_I_1	4.69×10^3	$5.40 \ge 10^7$
4_7_I_9	$1.22 \text{ x } 10^4$	5.39 x 10 ⁸
Synthetic Lognormal	2.22×10^4	2.44×10^8
Synthetic Pareto	$1.11 \ge 10^3$	$1.47 \text{ x } 10^7$

Table I Mean, Variance

Figure 1 compares the loss probability estimates resulted from applying Eqs. (25) and (26) (under the Pareto distribution modeling) for traffic trace *lbl-tcp-3*. Clearly the two performance curves are very close. Thus, for this work we adopt Eq. (25) thereafter.



Figure1. Differences between Equations 25 and 26.

Figure 2 compares the performance curves for traffic trace *lbl-pkt-4* estimated under the Lognormal and Pareto arrival modeling approaches. Clearly the Lognormal arrival method offers more traffic modeling accuracy (close to the connection simulation result).



Figure 2. Loss Probability versus Size of Buffer for the traffic Trace Ibl-pkt-4.

Figure 3 compares the performance curves for traffic trace 3-7-I-1 estimated under the Lognormal and Pareto arrival modeling approaches. In this case the Pareto arrival method offers better traffic modeling accuracy (close to the connection simulation result).



Figure 3. Loss Probability versus Size of Buffer for the Traffic Trace 3-7-I-1.

For the purposes of showing the advantage of using Multifractal traffic modeling, we additionally tested two other real traffic traces and 2 synthetically generated traces and add two new performance curves provided by the previously mentioned MSQ and CDTSQ methods. Figures 4, 5, 6 and 7 compare these four calculated performance curves with respect to the corresponding simulated one and confirm the superiority of multifractal modeling approaches.



Figure 4. Loss Probability versus Size of *Buffer* for the Traffic Trace 4-7-I-9.



Figure 5. Loss Probability versus Size of Buffer for the Traffic Trace Ibl-tcp-3.



Figure 6. Loss Probability versus Size of *Buffer* for the Synthetic Lognormal arrival Traffic Trace.



Figure 7. Loss Probability versus Size of *Buffer* for the Synthetic Pareto Arrival Traffic Trace.

The investigation results shown by Figures 2, 3, 4 and 5 suggests that traffic traces *lbl-pkt-4* and *lbl-tcp-3* can be best modeled by a Multifractal Lognomal arrival traffic model while traffic traces $4_7_{-1}9$, and $3_7_{-1}1$ by a Multifractal Pareto arrival traffic model. Figures 8, 9, 10 and 11 plot the empirical distributions of the corresponding arrival traffic load processes what conform the accurate decision of traffic models based on the loss probability estimates. Similar results also hold for the synthetically generated traffic traces, as illustrated by the Figures 12 and 13.



Figure 8. PDF Approximations for the Traffic Trace Ibl-pkt-4.



Figure11. PDF Approximations for the Traffic Trace lbl-tcp-3.



Figure 12. PDF Approximations for the Synthetic Lognormal Traffic Trace



Figure 13. PDF Approximations for the Synthetic Pareto Traffic Trace.

5. Conclusion

In this paper, we use two performance measures, namely empirical traffic arrival load distribution and loss probability at connection, as the criterion for accurate multifractal traffic modeling. Two proposals for calculating the loss probability were studied, considering the multifractal characteristics of traffic on a single server with finite *buffer*.

We strongly believe that a robust traffic modeling method should consider both traffic static and dynamic characteristics, where latter are strongly impacted and driven by network dynamics. This work can be viewed as our very first attempt for robust and complete network traffic modeling.

Experimentally we evaluated 4 different real network traffic traces and 2 synthetically generated traces. Both performance measure criterions provide accurate decision on the modeling fitting. Mostly important the proposed multifractal traffic models outperform some other well-known multifractal based approaches suggested in literature.

For future work, we should improve further our traffic models by considering other discriminating statistical information, for instance, higher order moments of traffic processes and/or other heavy tail probability distributions.

References

- Benes V.(1963) "General Stochastic Processes in Theory of Queues", Reading, MA: Addison Wesley.
- Crovella, M. E. and Bestavros, A. (1997) "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes". In: IEEE/ACM Transactions on Networking, 5(6):835—846.
- Daoudi, K., Lévy-Véhel, J. and Meyer, Y. (1998) "Construction of Continuous Functions with Prescribed Local Regularity". Journal of Constructive Approximation, vol 14 no 3, pp 349–385.
- http://ita.ee.lbl.gov/html/traces.html.
- Krishna, P.M. Gadre, V.M. and Desai, U.B.(2003) "Multifractal Based Network Traffic Modeling", Kluwer Academic Publishers.
- Leland W., Taqqu M., Willinger W. and Wilson D.(1994) "On The Self-Similar Nature of Ethernet Traffic" (extended version), IEEE/ACM Transactions on Networking, v.2, n.1, pp 1-15.
- Liu N. X. and Baras J. S. (2003) "Statistical Modeling and Performance Analysis of Multi-Scale Traffic", Proceedings of Twenty – Second Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2003, pp.1837-1847, San Francisco, CA, USA.
- Norros, I. (1994) "A Storage Model with Self-Similar Input, Queueing Systems",16, pp.387-396.
- Park, K. and Willinger, W. (2000) "Self-Similar Network Traffic and Performance Evaluation". John Wiley and Sons New York.
- Ribeiro V. J., Riedi R. H., Crouse M. S. and Baraniuk R. G. (2000) "Multiscale Queueing Analysis of Long-Range-Dependent Network Traffic", IEEE INFOCOM, pp. 1026-1035, Tel Aviv, Israel.
- Stenico, J.W.G. and Ling, L.L. (2010) "A Multifractal Based Dynamic Bandwidth Allocation Approach for Network Traffic Flows", IEEE International Conference on Communications (ICC), 23-27, pages 1 – 6.
- Stenico, J.W.G and Ling, L.L. (2011) "A Control Admission Sheme for Pareto Arrivals with Multi-Scale Characteristics", In: Proceedings of The International Workshop on Telecommunications - IWT 2011, pp. 220-224, Rio de Janeiro – Brazil.
- Riedi, R.H., Crouse, M.S., Ribeiro, V.J. and Baraniuk, R.G. (1999) "A Multifractal Wavelet Model with Application to Network Traffic". IEEE Transactions on Information Theory, vol. 45, pp. 992-1018.
- Vieira F.H.T. and Ling, L.L.(2009) "Adaptive Wavelet Based Multifractal Model Applied to the Effective Bandwidth Estimation of Network Traffic Flows". IET Communications, vol.3, pp. 906-919.