Some Results on Multiscale Queueing Analysis

Trang Dinh Dang, Sándor Molnár, István Maricza
High Speed Networks Laboratory, Dept. of Telecommunications & Telematics
Budapest Univ. of Technology & Economics, H–1117, Magyar tudósok körútja 2, Budapest, Hungary
Tel: (361) 463 3889, Fax: (361) 463 3107, E-mail: {trang, molnar, maricza}@ttt-atm.ttt.bme.hu

***Abstract*-Based on a general multifractal queueing approximation some results on multiscale queueing analysis are presented. The results include the well-known Weibullian tail for the important case of the *monofractal* fractional Brownian motion (fBm) input traffic. It also provides the basics for a new practical method for queueing performance evaluation of multifractal traffic. In addition, several impacts of multifractality on queueing performance are investigated and presented using the estimation method. The results are validated by queueing simulation of measured network traffic.**

I. Introduction

A basic finding of the LAN/WAN traffic analysis is that the packet traffic has the high variability and burstiness nature in a wide range of network environments. One of the remarkable characteristics is the *fractal* nature of the investigated LAN/WAN traffic [1], [2], [3], [4], [5], [6], [7], [8]. This is in close relationship to traffic burstiness because bursts should be defined in terms of time scales over which clustering activities occur [1].

In this framework *long-range dependence* (LRD) and *self-similarity* have been detected and a group of studies is concentrated on how to detect accurately the LRD property and how to estimate the Hurst parameter [9], [10]. A large group of traffic models (fractional Brownian motion (fBm) models, FARIMA models, Cox's M/G/∞ models, on/off models, etc.) to capture LRD and self-similar properties have also been developed [1], [11], [12], [8]. Among these models the fBm [13] was found to be a popular parsimonious and tractable model of traffic aggregation. It was shown that the fBm is an accurate model if the traffic is aggregated from a large number of independent users whose peak rates are small relative to link capacity and the flow control has no significant impact, see [14], [15], [16], [17] for more details.

The performance implications of the fractal property are also addressed in a series of studies [18], [19]. A collection of studies has proven that the fBm based models have a tail queue distribution that decays asymptotically like a Weibullian law [1], i.e., $\mathrm{P}[Q > b] \simeq \exp(-\delta b^{2-2H})$, where δ is a positive constant that depends on the service rate of the queue [13], [20]. This important result shows that queues with fBm input ($H > 1/2$) have a much slower decay than that of the exponential.

Recent measurements and researches of wide-area network traffic, however, also discovered that the LAN/WAN traffic has a more complex scaling behaviour which cannot be explained in a self-similar framework [3], [21], [4]. More precisely, it has been found that aggregate network traffic is asymptotically self-similar over time scales on the order of few hundreds of milliseconds and above but it exhibits *multifractal* scaling below this time scale. It has been also pointed out that the transition from the multifractal to self-similar scaling occurs around time scales of a typical packet round-trip time in the network [22], [4]. However, some studies showed that multifractal scaling can also be present even at large time scales [23]. Therefore multifractal traffic models with a much more flexible rule for the scaling law seem to be needed, especially for some WAN environments. The physical explanations and engineering implications are also addressed in several papers [4], [7].

There is a lack of queueing results available in the cases when the input traffic has a more complex scaling behaviour. Especially, queueing systems with multifractal input are an undiscovered field and only a few results published in the literature. Véhel *et al.* [24] suggested a cascade model for TCP traffic based on the retransmission and congestion avoidance mechanisms with no performance analysis. Riedi *et al.* [6] developed a multiscale queueing analysis in the case of tree-based multiscale input models. Gao *et al.* simulated queues fed by multiplicative multifractal processes in [25] but provided no analytical results. In contrast to these results we consider *general multifractal process* without any restrictions and derive an *analytical approximation for the queue tail asymptotics*.

The aim of this paper is to contribute to the queueing theory of multifractal queues and also to the traffic engineering implications. We present a novel analysis of multifractal queues. First an approximation formula is given for the tail of the queue with multifractal traffic input. The formula results in the asymptotically Weibullian decay for the special monofractal case of fractional Brownian motion input process which is consistent with the previously mentioned results [13], [20]. In this framework we show that any Gaussian process with scaling properties is in the class of monofractal processes and derive the related characterization functions. Practical applications of the results are also included in the paper. This paper also addresses some implications of multifractality on queueing performance. A comparison study between the queueing performance of the monofractal and multifractal processes are presented. We present a detailed queueing analysis for some measured WAN traffic which justifies the practical use of our formula. Our hope is that these results contribute as first steps to get the full understanding of multifractal queue-

0-7803-7661-7/03/$17.00 © 2003 IEEE

ing behaviour.

The rest of the paper is organized as follows. Section II introduces our queueing system under investigation and presents our main result, i.e., the approximation for queue tails with the proofs. Section III discusses the important applications of the result. The brief description of measured data traffic and our queueing analysis is given in Section IV. Finally, Section V concludes the paper and presents our targets for future research directions.

II. Queueing estimation of multifractal traffic

We consider a simple queueing model: a single server queue in continuous time, the serving principle for offered work is defined to be FIFO (First In, First Out), the queue has infinite buffer and constant service rate s. Denote by $X(t)$ the total size of work arriving to the queue from time instant $-t$ in the past up to this moment, time instant 0. The so called *workload process* $W(t)$ is the total amount of work stored in the buffer in time interval $(-t, 0)$, i.e.,

$$W(t) = X(t) - st \tag{1}$$

Our interest, however, is the current buffer length of the queue, denoted by Q. This is the queue length in the equilibrium state of the queue when the system has been running for a long time and the initial queue length has no influence. If this state of the system does exist, i.e., stationarity and ergodicity of the workload process hold, and the stability condition for the system is also satisfied, i.e., $\limsup_t \mathbb{E}[X(t)]/t < s$, then:

$$Q = \sup_{t \geq 0} W(t), \tag{2}$$

where $W(0)$ is assumed to be 0. This equation is also referred to as *Lindley's equation*.

The input process $X(t)$ is considered as a generally defined multifractal process by Mandelbrot *et al.* in [26]. The definition presents multifractal processes in terms of moments which leads to a more intuitive understanding of multifractality.

Definition 1: A stochastic process $X(t)$ is called multifractal if it has stationary increments and satisfies

$$\mathrm{E}\left[|X(t)|^q\right] = c(q) t^{\tau(q)+1} \tag{3}$$

for some positive $q \in \mathrm{Q}$, $[0,1] \subset \mathrm{Q}$, where $\tau(q)$ is called the scaling function and the moment factor $c(q)$ is independent of t.

A simple consequence of the definition it that the scaling function $\tau(q)$ is a concave function. If $\tau(q)$ is a linear function of q the process is called uniscaling or *monofractal*, otherwise it is called multiscaling or multifractal. The definition is very general and it covers a very large class of processes. Multifractal processes are also called processes with *scaling property*.

Note that an alternative approach to multifractal processes, also found in literature, is based on the study of the local erratic behaviour of the process by means of its local Hölder exponents. For details on this approach see [27] and references therein. The most obvious examples of multifractals are self-similar and multiplicative processes.

The following Proposition provides our queue tail approximation for the presented queueing system:

Proposition 1: The probabilities for the queue tail asymptotic of a single queueing model with general multifractal input is accurately approximated by:

$$\log(\mathrm{P}[Q > b]) \approx \min_{q>0} \log \left\{ c(q) \frac{\left[\frac{b\,\tau_0(q)}{s(q-\tau_0(q))}\right]^{\tau_0(q)}}{\left[\frac{b\,q}{q-\tau_0(q)}\right]^{q}} \right\}, \qquad b \text{ large} \tag{4}$$

where $\tau_0(q) := \tau(q) + 1$. The scaling function $\tau(q)$ and $c(q)$ are the functions which define the multifractal input process.

The proof of this Proposition is provided in Appendix VI-A.

For positive multifractal processes, i.e. $X(t) > 0$, Eq. (11) is an equality. In addition, the approximation in Eq. (16) and the inequality in Eq. (12) turn to be more accurate approximations as b tends to infinity. Thus the presented approximation is supposed to be asymptotically tight. The tightness and accuracy of the approximation is also experimentally investigated in Section IV.

Considering the formula in Eq. (4) we see that it has an implicit form and just the given form of the functions $c(q)$ and $\tau(q)$ can provide the final result. The reason behind this is that the definition for the class of multifractal processes gives no restrictions for the functions $c(q)$ and $\tau(q)$ (beyond that $\tau(q)$ is concave). *Our conjecture is that the analysis of queueing systems with general multifractal input may produce some similar general results*. It means that there is no general queueing behaviour for these systems as the Weibullian decay in the case of Gaussian self-similar processes [13]. An actual multifractal model will determine, for example, the queue length probabilities of the system.

III. Applications

A. Fractional Brownian motion

As a simple application first we consider a monofractal Gaussian process, called fractional Brownian motion (fBm). The fBm is self-similar which is a simple case of monofractality and it is also Gaussian. The increment process of fBm is called fractional Gaussian noise (fGn). Queueing analysis of a single queue with fBm input is first presented by Norros [13] which showed the Weibullian decay for the asymptotic tail behaviour, i.e., $\mathrm{P}[X > x] \sim \exp(-\gamma x^{\beta})$ with $\beta \leq 1$. This result is also justified by Large Deviation techniques in [20]. Applying this input process model to our formula should show its use and robustness when comparing to these available results.

First we prove that any Gaussian process with scaling property is in the class of monofractal processes. Furthermore we give the explicit forms for $\tau(q)$ and $c(q)$.

Consider the following lemma:

Lemma 1: A Gaussian process with scaling property is monofractal with parameters

$$\begin{cases} \tau(q) &= \frac{q}{2}\left[\tau(2)+1\right] - 1 \\ c(q) &= \frac{[2c(2)]^{q/2}}{\sqrt{\pi}} \Gamma\left(\frac{q+1}{2}\right), \end{cases}$$

where $\Gamma(\cdot)$ denotes the Gamma function,
$\Gamma(z) = \int_0^{+\infty} x^{z-1} \exp^{-x} \mathrm{d}x, \ z > 0.$

The proof of this Lemma is provided in Appendix VI-B.

Turning back to our case of fBm with $c(2) = 1$ and $\tau(2) = 2H - 1$ where H is referred to as the Hurst parameter, we have

$$\begin{cases} \tau(q) &= qH - 1 \\ c(q) &= \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right). \end{cases}$$

Insert these two functions into our formula in Eq. (4) we get

$$\log(\mathrm{P}[Q > b]) \approx \log\left(\min_{q>0}\left\{\frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right)\frac{\left(\frac{bH}{s(1-H)}\right)^{qH}}{\left(\frac{b}{1-H}\right)^{q}}\right\}\right)$$
$$=: \log(\min_{q>o} g(q)).$$

The minimum value of the $g(q)$ for $q > 0$ function can be easily determined by taking its derivatives. The result is the following:

$$\log(\mathrm{P}[Q > b]) \approx \log(\min_{q>o} g(q)) = \log\left(\frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\Psi^{-1}\left(\log K\right)\right)}{K^{\Psi^{-1}(\log K)-1/2}}\right)$$
$$=: \log(T_{fBm}(H, s, b)),$$

where $K = K(H, s, b) = \frac{1}{2}b^{2(1-H)}s^{2H}(1 - H)^{-2(1-H)}H^{-2H}$, $\Psi(\cdot)$ is the *digamma* function, $\Psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, and $\Psi^{-1}(\cdot)$ denotes the inverse function of $\Psi(\cdot)$.

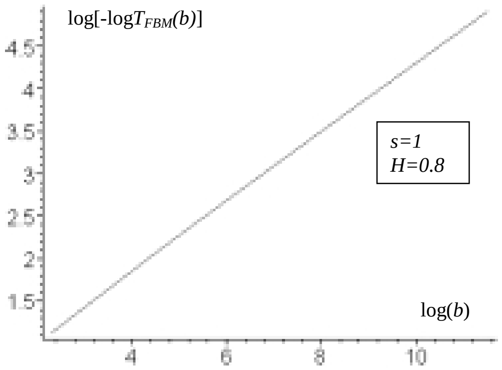


Fig. 1. By setting fixed values for H and s, the line in the log-log plot of $-\log T_{fBm}(b)$ versus b clearly shows the Weibullian decay for $T_{fBm}(H, s, b)$.

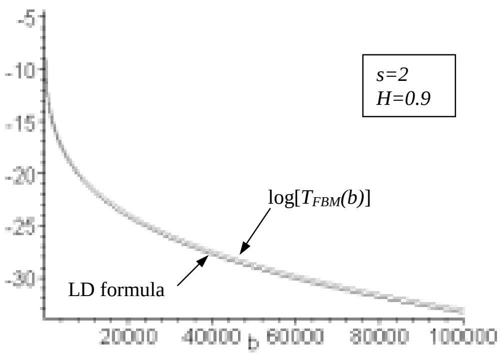


Fig. 2. Our approximation compared to the Large Deviation technique result: the two plots almost coincide for all values of queue size.

The $T_{fBm}(H, s, b)$ function is quite complex with the presence of Gamma, digamma, and its inverse function. However, we have quite a good approximation of $T_{fBm}(H, s, b)$:

Proposition 2: The approximation

$$\frac{1}{\sqrt{\pi}}\frac{\Gamma\left(\Psi^{-1}\left(\log x\right)\right)}{x^{\Psi^{-1}(\log x)-1/2}} \approx \exp(-x) \tag{5}$$

holds for large x, $x > 0$.
The proof and the precise sense of this approximation can be found in [28].

Applying this approximation we find that the queue tail for the fBm case satisfies:

$$\log\left(T_{fBm}(H, s, b)\right) \approx$$
$$\approx -\frac{1}{2}b^{2(1-H)}s^{2H}(1 - H)^{-2(1-H)}H^{-2H}, \quad b \text{ large.} \tag{6}$$

Eq. (6) shows the Weibullian decay of this queue which was first recognized and proven by Norros [13]. Numerical evaluations of the result are presented in Fig. 1 and Fig. 2. In Fig. 1 we fix the values of H and s and then calculate the values of the queue tail approximation $T_{fBm}(H, s, b)$ versus the queue size b and then plot it in the log-log scale. The linearity of the plot also demonstrates the Weibullian decay. In addition, the right hand side of Eq. (6) is exactly the result provided by Duffield and O'Connell in [20]. This asymptotic formula was proven using the Large Deviation technique. The accuracy of the approximation is depicted in Fig. 2. We can see that the plots almost coincide for all calculated values of the queue size.

Our conclusions can be summarized in two main points: (i) the asymptotic tail approximation for the case of fBm has Weibullian decay; (ii) this result is also consistent with the formula presented by Norros [13] and by Duffield *et al.* with Large Deviation technique [20].

In the case of $H = 1/2$ (Brownian motion) the above formula results in $\log \mathbf{P}[Q > b] \approx -2sb/\sigma^2$ where σ^2 denotes the variance of the process, which is in agreement with the queueing formula known from the theory of Gaussian processes [29], [20].

B. Practical solutions

We show here the practical use of the formula. Assume that we are interested in the behaviour of the tail of the steady-state buffer occupancy (queue length) distribution at a specific multiplexer in our network. The first step should be the fine resolution measurements of the input process. We also assume that the input process exhibits multifractal scaling properties. Then the scaling function $\tau(q)$ and the function $c(q)$ can be estimated from the collected data for some available parameters $q > 0$. *We emphasize the importance of the function $c(q)$ as the quantity factor of multifractal processes which is sometimes neglected in a number of studies dealing with multiscaling properties of the high-speed network traffic. The scaling function $\tau(q)$ defines only the quality of multiscaling and it is not enough for the description of a multifractal model and therefore for the analysis of queueing models with multifractal input processes.*

Now we suggest two practical methods for the approximation of the queue tail distribution:

1. Given the service rate s and the two sets $\{c(q)\}$ and $\{\tau(q)\}$, using Eq. (4) the approximation of $\log(\mathrm{P}[Q > b])$ can be computed for each value of b. This method is very

simple but it is the more useful from network planning and capacity dimensioning point of view since we are only interested in some values of the tail probabilities. We mainly focus on the practical use of this method in this study.

2. The input process is fitted to a multifractal model. The two measured sets of $c(q)$ and $\tau(q)$ are fitted by $\tilde{c}(q)$ and $\tilde{\tau}(q)$. Then the analysis of the Eq. (4) with these functions can result in simple closed form of the queue tail probabilities. We use this method when studying the queue tail behaviour of a multifractal model. However, characteristic functions of multiscaling processes are often complex, thus it is difficult to give a closed queueing form for these cases. The more details on this topic is in focus of future work.

C. The impacts of multifractality

As discussed, the tail behaviour of a queueing system depends on both the scaling function and the moment factor of the multifractal traffic input. Applying the estimation method presented above we show in this section a deeper study on these effects through some typical numerical examples.

1) Multifractal versus monofractal: Consider a multiplicative multifractal process with symmetric Beta(α,α) distributed multiplier (see more details in [30]). For this multifractal we can exactly calculate the characteristic functions at a certain cascade level, i.e.,

$$\begin{cases} \tau_0(q) &= \log_2 \frac{\Gamma(\alpha)\Gamma(2\alpha+q)}{\Gamma(\alpha+q)\Gamma(2\alpha)} \\ c(q) &= 2^{N\left(q-\log_2 \frac{\Gamma(\alpha)\Gamma(2\alpha+q)}{\Gamma(\alpha+q)\Gamma(2\alpha)}\right)}. \end{cases}$$

Note that at all stages the generated multiplicative set is normalized to have unit mean. The scaling function of a multiplicative cascade with $\alpha = 15$, level $N = 20$ is presented in Figure 3. Assume that there exists a (mono)scaling process with exactly the same moment factor $c(q)$ as of the mentioned multifractal but it has the uniscaling fractal structure $\tau_0(q) = qH$ (also see in Figure 3 with $H = 0.8$ and $H = 0.9$).

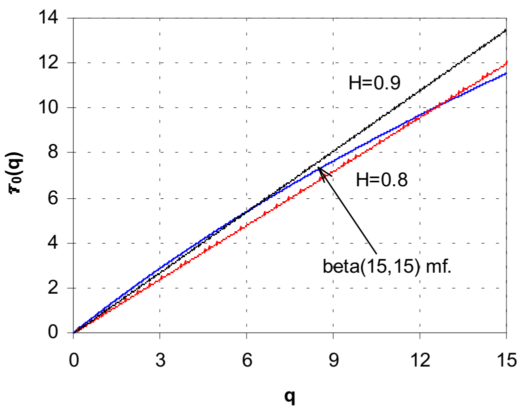


Fig. 3. The scaling functions of the examined fractal processes.

With the knowledge of the characteristic functions of these scaling processes we can calculate the estimation for the tail probabilities of the queueing system for large queue sizes (the service rate is set to be $s = 2.0$) by using our numerical

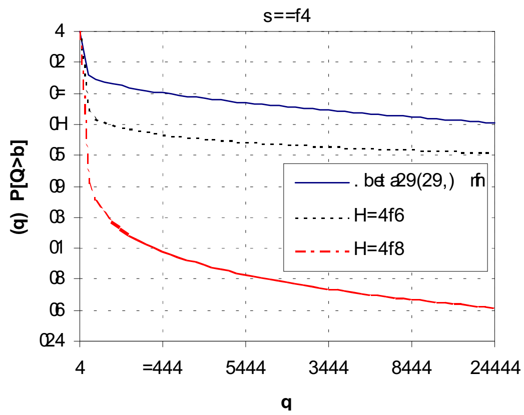


Fig. 4. Queue tail approximation of the examined fractal processes.

method. The results are presented in Figure 4. We observe that the approximated queue tail probabilities of the multifractal case are noticeably greater than the monofractal cases. This result clearly indicates that in the same queueing environment with the same moment function for the input process increments and with scaling functions which are not far from each others in "value" (see Figure 3) the queueing behaviour of the multifractal case is considerably worse than of the monofractal case.

2) The impact of the moment factor: We also examine the impacts of the moment factor of the multifractal input process with the similar consideration. Given a known scaling process fBm with $\tau_0(q) = qH, H = 0.8$ and $c(q) = \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right)$ we make some modification on the moment factor, thus create a new theoretical scaling process with only differences in the absolute moments of the increments. Assume the following fractal process with the characteristic functions:

$$\begin{cases} \tau_0(q) &= qH, \qquad H = 0.8 \\ c(q) &= \begin{cases} \frac{2^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right) & \text{if} \quad 0 \le q \le 2 \\ a(q-2)(q+b) & \text{if} \quad q > 2 \end{cases} \end{cases}$$

We denote the process with the setting $\{a = 0.037, b = 15.51\}$ and $\{a = 0.031, b = 15.51\}$ *up-mod* and *down-mod*, respectively. With these modifications the concerned process increments have the same moments up to the moment order $q = 2$, thus they have the same mean and variance. The difference between the moment factors is seen in Figure 5.

In the performance study of these processes, presented in Figure 6, we can observe the effects of these slight modifications of the moment factor. The change of higher order moments has a clear impact on the queueing behaviour of the process. The *up-mod* scaling process gives rise to the worse queueing performance and the *down-mod* process courses the better behaviour in the same system setting as compared to the original fBm process.

In summary, despite the fact that we should deal carefully with the presented results because they are based on queueing performance approximation we can conclude that the monofractal processes have a better queueing behaviour compared to the multifractal processes and the moment factor also exerts influence on the queueing performance. These important observations should be considered in traffic modeling

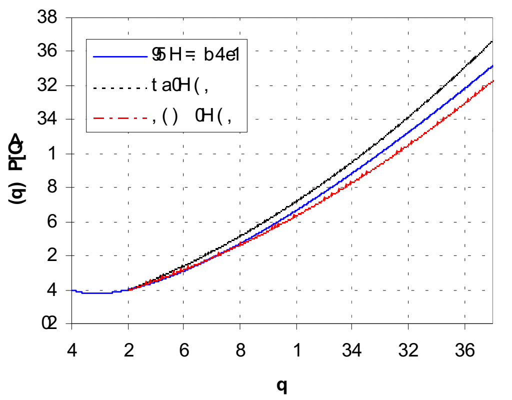


Fig. 5. The moment functions of the examined fractal processes.

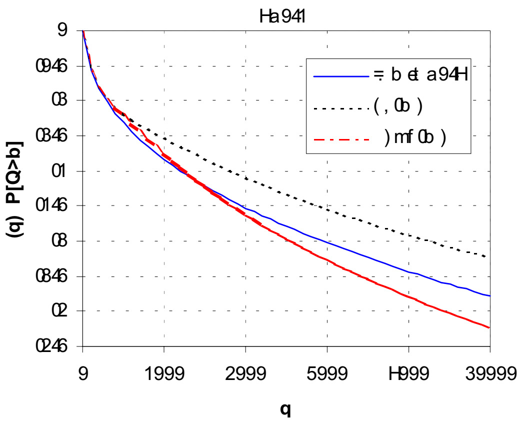


Fig. 6. Queue tail approximation of the examined fractal processes.

and queueing study of the scaling processes.

IV. Queueing Analysis

In this section we show the validation for the mentioned practical solution presented above by the queueing analysis of some real traffic traces. We also provide a simple method for estimation of multiscaling functions $c(q)$ and $\tau(q)$.

A. Data traces

Three data traces were considered in our analysis which are freely available at the Internet Traffic Archive [31]. These traces contain an hour's worth of all wide-area traffic each between Digital Equipment Corporation and the rest of the world. The traces, denoted by DEC-PKT-1, DEC-PKT-2, and DEC-PKT-3, were gathered at Digital's primary Internet access point, which is an Ethernet DMZ network operated by Digital's Palo Alto research groups. The raw traces were made using tcpdump on a DEC Alpha running Digital's OSF/1 operating system, which includes a kernel filter with capabilities comparable to those of BPF. `Tcpdump` captured all IP packet header information with millisecond precision timestamps. Each trace contains more than 3 million packet headers.

We constructed the packet arrival counts traces of time sample of 3 milliseconds from the raw data. Preliminary analysis of these traces exhibits scaling properties, we therefore use them as the inputs of the queue system under investigation.

Data set	Number of arrivals	Mean	Variance
DEC-PKT-1	3 027 907	2.5232	4.4153
DEC-PKT-2	3 987 942	3.3234	5.2416
DEC-PKT-3	4 518 090	3.7652	5.9968

TABLE I
SUMMARY OF THE INVESTIGATED DATA SETS.

B. Simple method for multiscaling functions estimation

The full description of a multifractal model involves both $c(q)$ and the scaling function $\tau(q)$. We present here a simple method for testing of scaling properties and also for the estimation of these functions.

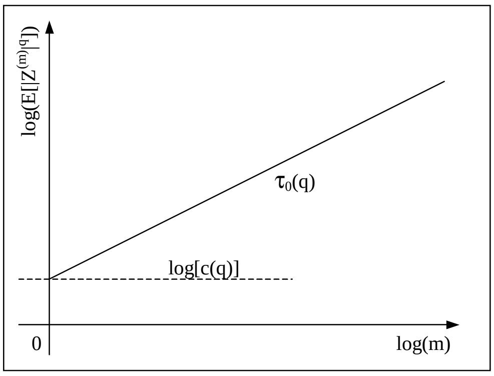


Fig. 7. A simple method for scaling test and the estimation of $c(q)$ and the scaling function $\tau(q)$.

The definition of multifractal processes (Def. 1) claims the stationarity condition for the increments. Therefore it is easy to verify the following relation for the moments of the increments:

$$\mathrm{E}[|Z^{(\Delta t)}|^q] = c(q)(\Delta t)^{\tau(q)+1} = c(q)(\Delta t)^{\tau_0(q)}, \qquad q>0, \quad (7)$$

where $Z^{(\Delta t)}$ denotes the increment process of time sample Δt. Thus this equality also holds for $m = 1, 2, \ldots$

$$\mathrm{E}[|Z^{(m\Delta t)}|^q] = c(q)(m\Delta t)^{\tau_0(q)}, \qquad q>0. \quad (8)$$

Choose Δt as the time unit, then

$$\log \mathrm{E}[|Z^{(m)}|^q] = \tau_0(q)\log m + \log c(q), \qquad q>0. \quad (9)$$

Based on this property, the method is the following: given a data series of a process increments $Z_1, Z_2, \ldots, Z_n$. We denote its corresponding *real* aggregated sequence of the aggregation level m by $\{Z^{(m)}, Z_k^{(m)} = \sum_{i=(k-1)m+1}^{km} Z_i,\ k = 1, 2, \ldots\}$, $m = 1, 2, \ldots$ If the sequence $\{Z_k\}$ has scaling property then the plot of absolute moments $\mathrm{E}[|Z^{(m)}|^q]$ versus m on a log-log plot should be a straight line due to Eq. (9). The slope of the straight line provides the estimate of $\tau_0(q)$ and the intercept is the value for $\log c(q)$. The illustration of the method can be seen in Fig. 7.

Note that we have no need to estimate $c(q)$ and $\tau_0(q)$ for all positive value of q, which is an impossible task. In fact, the largest value of q we should considered depends on the interested finite queue length of the involved queue length probability, see below.

C. *Analysis of theoretical and simulation results*

We present here our analysis results of the mentioned data sets. We validate the use of our approximation in a single queue with constant service rate and general multifractal input. Two typical cases are discussed in this Section.

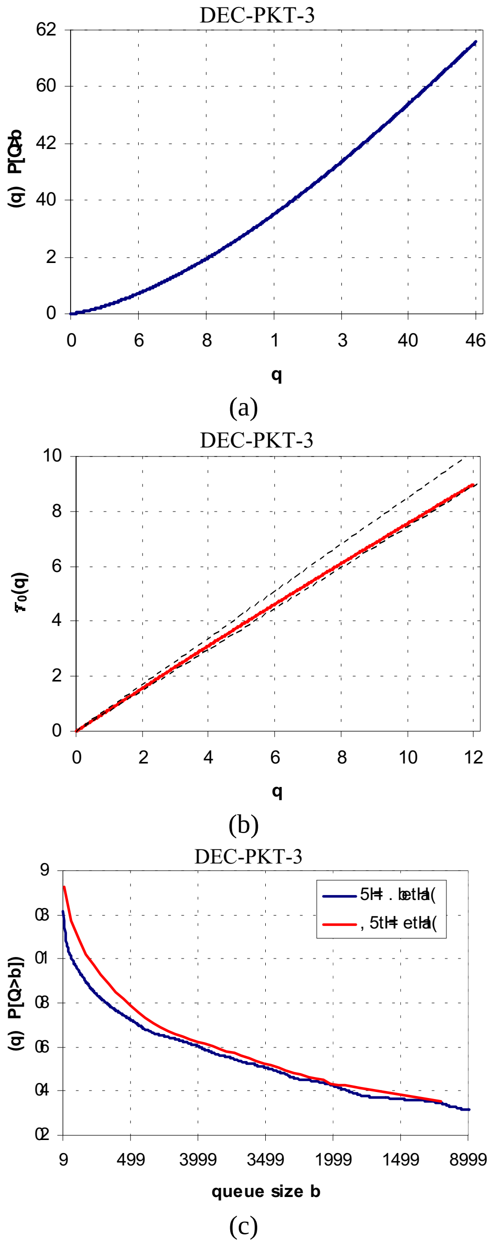


Fig. 8. Analysis results of the DEC-PKT-3 data set: (a) and (b) estimated form of $\tau_0(q)$ and $\log[c(q)]$, respectively; the concavity of $\tau_0(q)$ shows the multiscaling nature of the data set, (d) queueing simulation compared with the theoretical tail probabilities.

After applying the estimation method presented in the previous subsection we get the two sets of estimated characteristic functions $\tau_0(q)$ and $c(q)$ of the data set DEC-PKT-3. The results are given in Fig. 8(b) and Fig. 8(c) (we estimate $\log c(q)$ instead of $c(q)$). As observed in the figure the plot of the function $\tau_0(q) = \tau(q) + 1$ is a concave curve which suggests the multifractal property of DEC-PKT-3. We then make a comparison between our approximation and the queueing simulation of real data traces to validate the use of the formula in practice.

The approximation for probabilities of queue tail presented in Proposition 1 can be rewritten in the form

$$\begin{aligned} &\log \mathrm{P}[Q > b] \approx \\ &\approx \min_{q>0} \left\{ \log c(q) + \tau_0(q) \log \frac{b\tau_0(q)}{s(q-\tau_0(q))} - q \log \frac{bq}{q-\tau_0(q)} \right\} \\ &=: \min_{q>0} \{\log T^*(s,b)\} = T(s,b). \end{aligned} \tag{10}$$

For the sake of calculation simplicity we choose the service rate such that $s = 1$. The lower curve in Fig. 8(d) shows the simulation result of the DEC-PFT-3 data set. Using Eq. (10) the value of the logarithmic tail probability at each concerned value of queue size b is taken by the numerical minimization of $\log T^*(s,b)$ with the estimated sets $\{c(q)\}$ and $\{\tau_0(q)\}$. An example is shown in Fig. 9. In addition, we do not need to

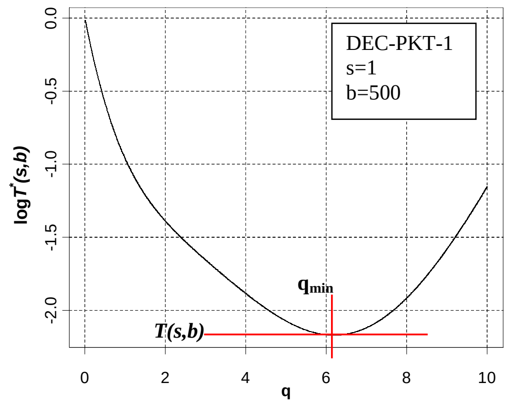


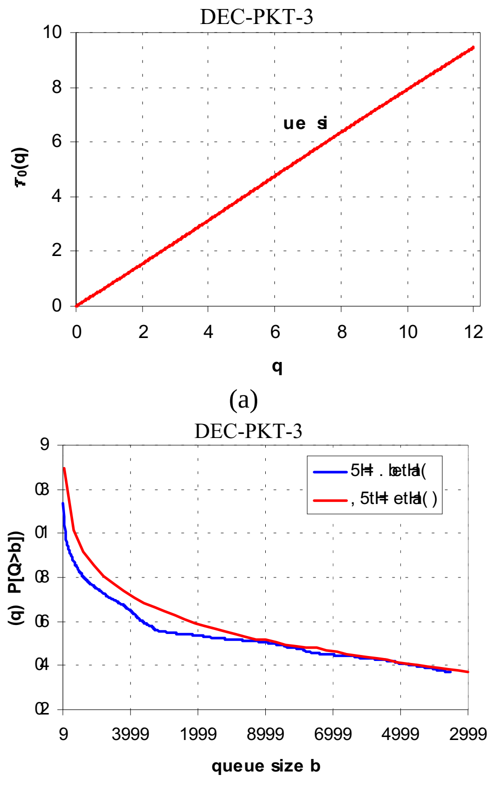
Fig. 9. Theoretical queue tail probability at each value of queue size b is the minimum of $\log T^*(s,b)$, for example DEC-PKT-1, $s = 1, b = 500$.

plot $\log T^*(s,b)$ at each value of q to find its minimum. A simple program routine can do it for all concerned value of b at once. Our theoretical tail probabilities are on the upper curve in Fig. 8. As comparing with the simulation result which is seen in the same figure we found that it has the similar shape and becomes tight as b increases. This validates our result.

Similar analysis has been performed with the DEC-PKT-2 data set. The results are shown in Fig. 10. In this case it is found that the data trace has the exact monofractal structure and can be well modelled by statistical self-similarity with Hurst parameter $H = 0.8$, see Fig. 10(a). Our queueing model deals with general multifractal input so it also involves the case of monofractal processes. Therefore the experimental queueing analysis also provides the correct queueing results which can be seen in Fig. 10(b).

V. Conclusion

In this paper the queueing performance of a single server infinite capacity queue with a constant service rate fed by general multifractal input process was investigated. An asymptotic approximation was derived for the steady-state queue length probabilities. It has been shown that the queueing formula gives the well-known Weibullian queue tail in case of the monofractal fractional Brownian motion input process. It was also proven that the class of Gaussian processes with scaling



(a)

(b)

Fig. 10. Analysis results of the DEC-PKT-2 data set: (a) estimated form of $\tau_0(q)$ which shows the data set can be modelled by statistical self-similarity with $H = 0.8$, (d) queueing simulation compared with the theoretical tail probabilities.

properties is limited to monofractal processes and the characteristic functions were derived. We have introduced a numerical calculation method for estimation of multifractal queueing performance. Some impacts of multifractality were investigated and presented. We have also shown that our formula gives the correct result in analysis of both multifractal and monofractal network traffic cases.

There are several interesting topics for further research. Based on the multifractal process characterization one of our goal is to build a multifractal traffic model parameterized by the multifractal functions. We also intend to carry out more multifractal analyses of measured LAN/WAN traffic with corresponding performance analysis.

VI. Appendices

A. Proof of Proposition 1

Using Lindley's equation the tail probabilities of queue length can be rewritten of the form: $\mathrm{P}[Q > b] = \mathrm{P}[\sup_{t\geq 0} W(t) > b]$. First let consider the quantity $\mathrm{P}[W(t) > b]$:

Replacing $W(t)$ by Eq. (1) we have

$$\begin{aligned} \mathrm{P}[W(t) > b] &= \mathrm{P}[X(t) - st > b] \\ &\leq \mathrm{P}[|X(t)| > b + st] && (11) \\ &= \mathrm{P}[|X(t)|^q > (b+st)^q], && \text{for any } q > 0 \\ &\leq \frac{E[X(t)^q]}{(b+st)^q}, && (12) \end{aligned}$$

The last inequality is applied using Markov's inequality.

Since the input process is multifractal defined by Def. 1 then:

$$\begin{aligned} \mathrm{P}[W(t) > b] &\leq \frac{c(q)t^{\tau_0(q)}}{(b+st)^q} \\ \Rightarrow \sup_{t\geq 0} \mathrm{P}[W(t) > b] &\leq \sup_{t\geq 0} \frac{c(q)t^{\tau_0(q)}}{(b+st)^q} =: \sup_{t\geq 0} f(t). \quad (13) \end{aligned}$$

By straightforward calculation it is easy to see that derivative of $f(t)$ equals zero when

$$t_0 = \frac{b\tau_0(q)}{s[q-\tau_0(q)]} > 0.$$

The second derivative of $f(t)$ at this point is

$$f''(t_0) = c(q)t_0^{\tau_0(q)-1}(b+st_0)^{-q-1}s[\tau_0(q) - q] < 0,$$

thus $f(t)$ has its maximal value at $t = t_0$. Note that we have assumed that $q > \tau_0(q)$, this is justified by the fact that in the monofractal case we have $\tau_0(q) = qH < q$ and by the concavity of $\tau(q)$.

Therefore

$$\sup_{t\geq 0} \mathrm{P}[W(t) > b] \leq \sup_{t\geq 0} f(t) = c(q)\frac{\left[\frac{b\,\tau_0(q)}{s(q-\tau_0(q))}\right]^{\tau_0(q)}}{\left[\frac{b\,q}{q-\tau_0(q)}\right]^q}$$

$$\log\left(\sup_{t\geq 0} \mathrm{P}[W(t) > b]\right) \leq \log\left(c(q)\frac{\left[\frac{b\,\tau_0(q)}{s(q-\tau_0(q))}\right]^{\tau_0(q)}}{\left[\frac{b\,q}{q-\tau_0(q)}\right]^q}\right),$$

$$\log\left(\sup_{t\geq 0} \mathrm{P}[W(t) > b]\right) \leq \min_{q>0}\log\left(c(q)\frac{\left[\frac{b\,\tau_0(q)}{s(q-\tau_0(q))}\right]^{\tau_0(q)}}{\left[\frac{b\,q}{q-\tau_0(q)}\right]^q}\right). \quad (14)$$

For a large class of stochastic processes (including fBm) the following limit holds [32]:

$$\lim_{b\to\infty} \frac{\log(\mathrm{P}[Q > b])}{\log(\sup_{t\geq 0} \mathrm{P}[W(t) > b])} = 1. \quad (15)$$

In addition,

$$\log(\mathrm{P}[Q > b]) \geq \log(\sup_{t\geq 0} \mathrm{P}[W(t) > b]), \quad (16)$$

then the right-hand side of Eq. (14) is a upper bound of a lower bound on $\log(\mathrm{P}[Q > b])$. The used inequalities in Eq. (16) and Eq. (12) become tight for finite large b. Thus our approximation for the queue tail asymptotics is the following:

$$\log(\mathrm{P}[Q > b]) \approx \min_{q>0}\log\left(c(q)\frac{\left[\frac{b\,\tau_0(q)}{s(q-\tau_0(q))}\right]^{\tau_0(q)}}{\left[\frac{b\,q}{q-\tau_0(q)}\right]^q}\right), \qquad b \text{ large.}$$

□

B. Proof of Lemma 1

Denote by $X(t)$ the Gaussian process. Since $X(t)$ has scaling property it satisfies the general definition for multifractal process, i.e., $\mathrm{E}[|X(t)|^q] = c(q)t^{\tau(q)+1}$. Thus the variance of the

$X(t)$ process should be $\sigma_t^2 = c(2)t^{\tau(2)+1}$. The Gaussian process $X(t) \sim N\left(0, c(2)t^{\tau(2)+1}\right)$ has the normal distribution and we have

$$f(x) = \frac{1}{\sqrt{2\pi c(2)t^{\tau(2)+1}}} \exp\left(-\frac{x^2}{2c(2)t^{\tau(2)+1}}\right).$$

The q^{th} moment of $X(t)$ can be calculated by the definition:

$$\begin{aligned} \mathrm{E}[|X(t)|^q] &= \int_{-\infty}^{+\infty} |x|^q f(x)\mathrm{d}x \\ &= 2\int_0^{+\infty} x^q \frac{1}{\sqrt{2\pi c(2)t^{\tau(2)+1}}} \cdot \end{aligned}$$

$$. \exp\left(-\frac{x^2}{2c(2)t^{\tau(2)+1}}\right)\mathrm{d}x.$$

Introduce $y := \frac{x^2}{2c(2)t^{\tau(2)+1}}$. The formula above can be rewritten as follows:

$$\begin{aligned} \mathrm{E}[|X(t)|^q] &= \frac{2^{q/2}}{\sqrt{\pi}}\left[c(2)t^{\tau(2)+1}\right]^{q/2}\int_0^{+\infty} y^{\frac{q-1}{2}}\exp(-y)\mathrm{d}y \\ &= \frac{[2c(2)]^{q/2}}{\sqrt{\pi}}\Gamma\left(\frac{q+1}{2}\right)t^{\frac{q}{2}[\tau(2)+1]}, \qquad (17) \end{aligned}$$

which concludes our proof. □

REFERENCES

[1] W. Willinger, M. S. Taqqu, and A. Erramilli, "A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks," in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, and I. Ziedins, Eds., Oxford, 1996, vol. 4 of *Royal Statistical Society Lecture Notes Series*, pp. 339–366, Oxford University Press.

[2] M. S. Taqqu, V. Teverovsky, and W. Willinger, "Is network traffic self-similar or multifractal?," *Fractals*, vol. 5, pp. 63–73, 1997.

[3] R. H. Riedi and J. Lévy-Véhel, "Multifractal properties of TCP traffic: a numerical study," INRIA research report 3129, Rice University, February 1997.

[4] A. Feldmann, A. C. Gilbert, and W. Willinger, "Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic," *ACM Computer Communication Review*, vol. 28, pp. 42–55, Sept. 1998.

[5] A. C. Gilbert, W. Willinger, and A. Feldmann, "Scaling analysis of conservative cascades, with applications to network traffic," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 971–991, Apr. 1999.

[6] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 992–1018, Apr. 1999.

[7] R. H. Riedi and W. Willinger, "Toward an improved understanding of network traffic dynamics," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds. June 1999, Wiley.

[8] S. Molnár and I. Maricza eds., "Source characterization in broadband networks," Interim report, COST 257, Vilamoura, Portugal, Jan. 1999.

[9] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, One Penn Plaza, New York, NY 10119, 1995.

[10] P. Abry and D. Veitch, "Wavelet analysis of long range dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.

[11] D. L. Jagerman, B. Melamed, and W. Willinger, "Stochastic modeling of traffic processes," in *Frontiers in Queueing: Models, Methods and Problems*, J. Dshalalow, Ed., pp. 271–370. CRC Press, 1997.

[12] S. Molnár and A. Vidács, "On modeling and shaping self-similar ATM traffic," in *15th International Teletraffic Congress*, Washington, DC, June 1997.

[13] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.

[14] D. R. Cox, *Statistics: An Appraisal, Proc. 50th Anniversary Conference*, chapter Long-Range Dependence: A Review, Iowa State University Press, Ames, IA, 1984.

[15] T. G. Kurtz, *Stochastic Networks: Theory and Applications*, chapter Limit Theorems for Workload Input Models, Oxford University Press, 1996.

[16] M. S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," *Computer Communication Review*, vol. 27, pp. 5–23, 1997.

[17] W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," in *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, R. J. Adler, R. E. Feldman, and M. S. Taqqu, Eds., pp. 27–53. Birkhäuser, Boston, 1998.

[18] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," in *IEEE/ACM Trans. on Networking*, New York, NY, Apr. 1996, vol. 4, pp. 209–223, aCM Press.

[19] A. Erramilli, O. Narayan, A. L. Neidhardt, and I. Saniee, "Performance impacts of multi-scaling in wide-area TCP/IP traffic," in *Proc., IEEE INFOCOM 2000*, Tel Aviv, Israel, 2000, vol. 1, pp. 352–359.

[20] N. G. Duffield and N. O'Connell, "Large deviations and overflow probabilities for the general single-server queue, with applications," in *Proc., Cam. Phil. Soc.*, 1994, vol. 118, pp. 363–374.

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

[22] A. Feldmann, A. C. Gilbert, W. Willinger, and T. G. Kurtz, "The changing nature of network traffic: Scaling phenomena," *ACM Computer Communication Review*, vol. 28, pp. 5–29, Apr. 1998.

[23] S. Molnár and T. D. Dang, "Scaling analysis of IP traffic components," in *ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management*, Monterey, CA, 18-20 September 2000.

[24] J. Lévi-Véhel and B. Sikdar, "A multiplicative multifractal model for TCP traffic," in *Proc., IEEE ISCC*, Hammamet, Tunisia, July 2001.

[25] J. Gao and I. Rubin, "Multifractal modeling of counting processes of long-range dependent network traffic," in *Proc., SCS Advanced Simulation Technologies Conf.*, San Diego, CA, Apr. 1999.

[26] B. B. Mandelbrot, A. Fisher, and L. Calvet, *A Multifractal Model of Asset Return*, Yale University, 1997, Working Paper.

[27] R. H. Riedi, "Multifractal processes," in *Theory and Applications of Long Range Dependence*, P. Doukhan, G. Oppenheim, and M. S. Taqqu, Eds. Birkhäuser, Boston, 2002.

[28] T. D. Dang and S. Molnár, "Queue asymptotics with general multifractal input," Tech. Rep., Budapest University of Technology and Economics, July 2001.

[29] M. B. Marcus and L. A. Shepp, "Sample behaviour of Gaussian processes," in *Proc., Sixth Berkeley Symp. Math. Statist. Prob.*, Berkeley, 1972, vol. 2, pp. 423–442, Univ. of Calif. Press.

[30] T. D. Dang, S. Molnár, and I. Maricza, "Capturing the complete characteristics of multifractal network traffic," in *Proc., GLOBECOM 2002*, Taipei, Taiwan, Nov. 2002.

[31] "The internet traffic archive," http://ita.ee.lbl.gov.

[32] J. Hüsler and V. Piterbarg, "Extremes of a certain class of Gaussian processes," *Stochastic Process. Appl.*, vol. 83, pp. 257–271, 1999.