Multi-Scale Internet Traffic Analysis Using Piecewise Self-Similar Processes

Yusheng Ji(a), Member

SUMMARY Numerous studies have shown that scaling exponents of internet traffic change over time or scaling ranges. In order to analyze long-range dependent traffic with changing scaling exponents over time scales, we propose a multi-scale traffic model that incorporates the notion of a piecewise self-similar process, a process with spectral changes on its scaling behavior. We can obtain a performance curve smoothened over the range of queue length corresponding to time scales with different scaling exponents by adopting multiple self-similar processes piecewise into different spectra of time scale. The analytical method for the multiscale fractional Brownian motion is discussed as a model for this approach. A comparison of the analytical and simulation results, using traffic data obtained from backbone networks, shows that our model provides a good approximation for Gaussian traffic.

key words: piecewise self-similar, long-range dependence, multi-scale, fractional Brownian motion, traffic modeling

1. Introduction

As numerous research results have demonstrated, internet traffic has long-range dependence that shows a slowly decreased correlation structure against its correlation lags [1]–[4]. In order to model this kind of traffic, self-similar processes are advocated instead of traditional Poisson-based models. More recent studies suggest that this kind of long-memory property in current internet traffic shows a more complicated scaling behavior: unlike the unchanged Hurst parameter of a self-similar process, scaling exponents of such processes change over time or the order of process moments [5], [6]. In order to capture such traffic features more precisely, multifractal stochastic models are proposed as one of the promising models for the time dependent scaling nature of network traffic [7]–[10]. By considering the higher-order statistics of network traffic, multifractal models are able to capture more irregularities in the scaling behavior.

On the other hand, examining on the statistical feature of traffic traces monitored from real networks, it is also found that the scaling exponents of network traffic changes over time scales [11]–[14]. In [11], an analytical method for modeling TCP/IP traffic with such scaling behavior is proposed. However, since the method in [11] is based on a microscopic view of TCP/IP connections, it may be restricted by assumptions on the characteristics of each composing TCP connection. In this paper we propose a novel approach using only the properties obtained from the aggregated traffic for modeling long-range dependent traffic with changing scaling exponents over time scales. We can obtain a performance trajectory smoothened over a range of queue length or system load corresponding to time scales with different scaling exponents, by adopting different self-similar processes into different time scales. We introduce the notion of piecewise self-similar processes, which is a superordinate concept of the multiscale fractional Brownian motion, a stochastic model to mimic some long-memory behaviors in biomechanical and financial research fields [15], [16]. An analytical method for the multiscale fractional Brownian motion is discussed as the model case for our approach. A comparison of the analytical and simulation results, using traffic data obtained from backbone networks, shows that in spite of the small computational complexity, our method provides a good approximation for Gaussian traffic.

2. Multi-Scale Behavior of Internet Traffic

We use data traces monitored from the Science Information Network (SINET) [17] in this paper, as well as the Abilene Network of the Internet2 project. SINET is a nation-wide internet backbone run by the National Institute of Informatics (NII), and connects more than 700 universities and institutions in Japan. The SINET data trace was obtained from the backbone traffic on an OC-3 ATM link, during the daily peak hours from about 15:00 on May 7, 2001, lasting for two hours. The Abilene data is from the westbound traffic on an OC-48 packet-over-SONET link [18], from about 9:40 in the morning of August 14, 2002 (local time), lasting for 40 minutes. The periods of the data traces used were chosen so that the long-term average rates and the second-order property were almost kept unchanged with time, ensuring the stationarity of the traffic. A quantitative analysis of the stationarity of the data traces used are shown in Appendix. In the following statistical analysis, the unit time scales are chosen to be $10^{-3}$ seconds for the SINET OC-3 data and $10^{-4}$ seconds for the Abilene OC-48 data, to ensure that the time stamps are accurate, and that the characteristics during the short time intervals are not lost.

2.1 Wavelet-Based Statistical Analysis

In order to clarify the statistical properties of the traffic data, we first used the wavelet-based estimation tool developed by D. Veitch and P. Abry [19] for the multi-scale traffic analy-
sis. Let $k$ denote the index of time in unit of $T_j=2^jT_0$ (also called scale $j$), where $T_0$ is the unit time scale, then statistical process $X(t)$ can be built up from a sum of weighted scaling functions and wavelets, such as

$$X(t) = \sum_k c_X(j_0, k)\phi_{j_0,k}(t) + \sum_{j=j_0}^{j_f} \sum_k d_X(j,k)\psi_{j,k}(t),$$

(1)

where $\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t-k)$, and $\phi(t)$ is a low-pass filter called scaling function, $c_X(j_0, k)$ is the $k$-th scaling coefficient at scale $j_0$, some coarse time scale; Similarly, $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t-k)$, and $\psi(t)$ is a bandpass filter called Haar wavelet, $d_X(j,k)$ is the $k$-th wavelet coefficient at scale $j$.

Scaling processes with self-similarity, long-range dependence, monofractal or multifractal properties all exhibit power-law spectra in some time scale ranges. The variance of wavelet coefficient $d_X(j, k)$ of scaling process $X(t)$ in scale $j$ can be expressed as

$$E[|d_X(j,k)|^2] \sim \alpha^q,$$

(2)

where $C_2$ is a constant, $\alpha = 2H+1$, and $H$ is the second-order scaling exponent. Therefore, a linear relationship between $\log_2 E[|d_X(j,k)|^2]$ and $\alpha_j + \log_2 C_2$ can be observed in a log-log scaled diagram.

The graphs in Fig. 1 show the second-order logscale diagram of the sample traces. The statistics are based on the number of bits that arrived during each time slot at the unit time scale. In the log-log scaled diagram, the estimated slope $\alpha$ of the logarithmic variance of wavelet coefficient $y_j$ against scale $j$ are 0.726 for SINET data and 1.19 for Abilene data. As their corresponding second-order scaling exponents are much larger than 0.5 in large time scale, we can conclude that both data traces have long range dependence. But however, the second-order properties in the figure are not linear against the scale $j$, with roughly two regions that can be differentiated at the scale of about 8–9 octaves for SINET, and 10–12 octaves for Abilene data, corresponding to time scales of 0.2–0.5 seconds for SINET, and 0.1–0.4 seconds for Abilene data. Therefore, the scaling exponents of both traces changed over time scales (or frequencies).

If a process has scaling property in some second-order statistics as in Eq. (2), then it will probably also have scaling property for the $q$-th moment as

$$E[|d_X(j,k)|^q] \sim C(q)2^{q\alpha_j},$$

(3)

where $C(q)$ is a function of order $q$, $\alpha(q) = q\alpha_j + q/2$, and $\alpha_j$ is the $q$-th order scaling exponent. A Linear Multiscale Diagram (LMD) [20] provides the relationship between the estimated scaling exponent $\alpha_j$ against order $q$. The multifractality can be observed in an LMD by the non-trivial change of the estimated scaling exponent against order $q$.

Figure 2 shows the Linear Multiscale Diagrams (LMD) of these two traces. The estimated exponents $h_q$ for small scales (Figs. 2(a) and (c)) do not change much over order $q$. In fact, the $q$-th order ($q \neq 2$) logscale diagrams (which are not shown in this paper) almost have the same shape as those of the second order. The differences in $h_q$ at the order range of $q \in [0, 8]$ for both traces are less than 0.02. Therefore, these two traces do not show significant multifractality in small time scale. This is also compatible with the results reported by Zhang et al., who argue that traffic in backbone networks is mostly monofractal rather than multifractal in small time scales [21].

On the other hand, in spite of the fact that the $h_q$ for the SINET data does not show significant change even in large scales (Fig. 2(b)), the $h_q$ for the Abilene data has obvious changes over order $q$ (Fig. 2(d)), indicating a multifractal-like property in large time scale. However, Figs. 1 and 2 suggest that these traffic data do not show significant mul-

![Fig. 1](image1.png)

**Fig. 1** Second-order logscale diagram of sample traces ($j=0$ corresponds to the time scale of 1 ms for SINET data, and 0.1 ms for Abilene data, and so as in Fig. 2).  

![Fig. 2](image2.png)

**Fig. 2** Linear Multiscale Diagrams of sample traces (Graphs in the top level are for SINET data, and bottom for Abilene data; the left column for small scales, and the right for large scales).
tifractal property in small time scales, and are long-range
dependent.

2.2 Variance-Time Property

For the discrete time process \( X(t) \), if the average of \( X(t) \) for consecutive numbers of \( m \)

\[
X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(t), \quad k = 1, 2, \ldots,
\]

then the second-order self-similarity can be expressed as

\[
\text{Var}[X^{(m)}(k)] = \sigma^2 m^{-2(1-H)}, \tag{4}
\]

where \( \sigma^2 \) is the variance of \( X(t) \). The second-order self-
similarity captures the property that the correlation structure has
preserved under time aggregation.

Variance-time plots can be used to verify the second-
order self-similarity or long-range dependence by observing
the variance function of \( X^{(m)}(k) \) versus the time scale \( m \). Let
\( X^{(m)}(k) \) be the number of bits that arrived during each con-
secutive time period with a length of \( m \Delta t \), where \( \Delta t \) is
the length of time slot at the unit time scale of the trace, then
each curve in Fig. 3 shows the variance of \( X^{(m)}(k) \) against
the time scale \( m \). Here the time scale \( m \) is the same no-
tion as scale \( j \) in wavelet-based analysis, and has the same
amount with \( 2^j \) in Fig. 1.

For both traces, the decaying rates of variance are not
constant along time scale \( m \). They show a kinked variance-
time property, like that reported by V. J. Ribeiro [22], with
roughly two scaling regimes: When \( m \) is small, the rate of
decay for the variance is high, and is rather close to the ex-
ponential distribution; as \( m \) increases, the rate of decay for
the variance reduces, showing a stronger dependence. Com-
paring two traces, we can find that in large time scale the
Abilene data shows a stronger dependence, which makes
the kink sharper. Figure 3 shows that a long-range depen-
dence exists in the traffic. Furthermore, the dependent fea-
tures are more complex than the exactly second-order self-
similar ones.

3. Piecewise Approximation of Multi-Scale Property

As a generalization of the notion of self-similarity, we pro-
pose here the notion of piecewise self-similar process to deal
with a process with time-scale dependent scaling properties.
A piecewise self-similar process, denoted by the \((M_K)\)-SS
process, is a continuous process with a number of \( K \) spectral
changes: it has the same behavior as a self-similar process
with a Hurst parameter \( H_i (0 \leq i \leq K) \) in the \( i \)-th spec-
trum of time scale. Note that \( H_0 \) represents the asymptotic
long-range behavior (when the frequency equals to 0). We
define piecewise self-similar process as a superordinate con-
cept of multiscale fractional Brownian motion (MBM)
introduced by J.-M. Bardet et al. [13], [14], denoted as \((M_K)\)-
FBM, whose composing sub-processes are FBM processes,
that is, an \((M_K)\)-FBM process behaves the same as an FBM
process with a Hurst parameter \( H_i (0 \leq i \leq K) \) in the \( i \)-
th spectrum of time scale. Hence, a multiscale fractional
Brownian motion process is also a piecewise self-similar
process, but the reverse is not always true, i.e., a piece-
wise self-similar process may not be a multiscale fractional
Brownian motion, since its composing sub-processes can be
any kind of self-similar processes other than FBM.

The time-scale dependent behavior of an \((M_K)\)-SS
process can be found in real traffic as in Figs. 1 and 3. In most
cases, one significant spectral change can be observed, and
others may not be so notable. Here we study the perform-
ance approximation of a piecewise self-similar process by
elaborately composing the performance of each self-similar
sub-process in the corresponding spectrum of time scale
piecewise.

Meanwhile, fractional Brownian motion (FBM) or its
incremental process fractional Gaussian noise (FGN) has
been used to describe the long-range dependence in network
traffic. An FBM process is a Gaussian centered process with
stationary increments. Differing from a regular Brownian
motion, the increments of an FBM are dependent, and an
FBM process has the exact self-similarity. We say a pro-
cess has Gaussian property if the marginal distribution of
the process is a Gaussian distribution. FBM process is the
only self-similar process with Gaussian property [23]. Since
a Gaussian distribution can be described by its mean \( \lambda \)
and standard deviation \( \sigma \), and the second-order scaling pro-
certainty can be described by the exponent \( H \) as in Eq. (2), an
FBM process is completely described by the triple \( \{\lambda, a, H\} \),
where \( a = \sigma^2 / \lambda \) (both \( \sigma \) and \( \lambda \) are at the unit time scale) is
called the variance coefficient, and \( H \in (0.5, 1) \) is the Hurst
parameter. When \( H=0.5 \), an FBM process becomes a regu-
lar Brownian motion process.

The complementary distribution of the queue length for
a queue with a service rate \( C \) fed by an FBM process with
parameters \( \{\lambda, a, H\} \) is given by [24]

\[
\Pr(Q > x) = e^{-yx^{2(1-H)}} \quad \text{as } x \to \infty, \tag{5}
\]

where \( y \equiv \frac{1}{\lambda \Delta t (1-H)} \left( \frac{(C-a)(1-H)}{H} \right)^{2H} \).
where \( \Phi(\cdot) \) is the residual distribution function of the standard Gaussian distribution. The lower bound approximation in (6) may underestimate the loss probability due to the fact that it only takes into account the event which happens with largest possibility. But as mentioned in [24], another approximation, which leads the RHS of (6) to the RHS of (5), may have an effect which happens to cancel out the effect of lower bound approximation in (6). However, since approximation (5) holds asymptotically when queue length approaches to infinity, attentions on the queue length or traffic load applied to the system should be paid carefully when we use (5) for the performance approximation.

From the central limit theorem, the aggregation of many i.i.d. processes with self-similarity is an FBM. Hence, we use (5) for the performance approximation. For describing an MFBM process, we need first to estimate the time-scale dependent Hurst parameter \( H_i \) from the variance-time plots or other estimation methods. Comprehensive surveys and comparisons of estimating methodologies for parameters in long-range dependent processes can be found in other literature [25]–[27]. Here we use the variance-time plots for the Hurst parameter estimation since they are expressed by the second-order properties, and hence can also be used for estimating the other parameter \( a_i \) in the triples of the MFBM process.

From the property of the second-order self-similarity in Eq. (4), if \( s_i \) and \( l_i \) are the smallest (highest) and the largest (lowest) time scales (frequencies) of the \( i \)-th spectrum, and \( v(m) \) is the variance value at time scale \( m \), then \( H_i \) can be expressed as a function of \( m \) and \( v(m) \), i.e.,

\[
H_i = f(m, v(m)), \quad (s_i \leq m \leq l_i).
\] (7)

Several kinds of methods can be considered for calculating \( H_i \). With a comparatively smooth variance-time relation, the simplest method may be to use only the coordinates (expressed in linear scale) of the points at the two ends of the spectrum: \( (s_i, v(s_i)) \) and \( (l_i, v(l_i)) \). Then \( H_i \) can be expressed as

\[
H_i = 1 - \frac{\log(v(s_i)/v(l_i))}{2\log(l_i/s_i)}.
\] (8)

Another method is to find a straight line using the coordinates of all points within the domain of the spectrum by the least square method. If \( m_1, m_2, \ldots, m_K \in [s_i, l_i] \) \((N \geq 2)\) are time scales of intermediate points between and including the smallest and the largest time scales of the \( i \)-th spectrum, then \( H_i \) can be expressed as

\[
H_i = 1 - \frac{N \sum_{n=1}^{N} \log(m_n^i) \log(v(m_n^i)) - \left( \sum_{n=1}^{N} \log(m_n^i) \right) \left( \sum_{n=1}^{N} \log(v(m_n^i)) \right)}{2 \left( \sum_{n=1}^{N} \log^2(m_n^i) \right) - N \sum_{n=1}^{N} \log^2(v(m_n^i))}.
\] (9)

By using the least square method, the influence of unexpected fluctuations in the variance values can be avoided.

On the other hand, the variance coefficient of a FBM process \( \{\lambda, a, H_i\} \) can be expressed as

\[
a_i = \frac{v_i(m_K)}{\lambda m_K^{2H_i}}.
\] (10)

where \( m_K \) denotes the unit time scale (in units of second), and \( v_i(m) \) is the variance value at this time scale. Hence, the time-scale dependent variance coefficient \( a_i \) in the \( i \)-th spectrum can be calculated by using \( H_i, m_K \), and any pair of abscissa and ordinate values (expressed in linear scale) of a point on the line from where \( H_i \) is calculated. If \( \{m, v(m)\} \) is a pair of such values, then in the log-log scaled coordinates we have

\[
2(1 - H_i) = \frac{\log v_i(m_K) - \log v_i(m)}{\log m - \log m_K}.
\] (11)

Reorganizing Eq. (11) and substituting \( v_i(m_K) \) in Eq. (10) we get

\[
a_i = \frac{10^{\log(v_i(m)+2(1-H_i)\log(m/m_K))}}{\lambda m_K^{2H_i}}.
\] (12)

With known triples \( \{\lambda, a_i, H_i\} \) for the MFBM process, we also need to know the relationship between time scale and queue length in order to use Eq. (5) to obtain the performance.
by $\{\lambda, a, H\}$, the critical time scale for the queueing performance of the FBM process can be expressed as

$$m = \frac{x}{C - \lambda} \frac{H}{1 - H} = \frac{x}{C(1 - \rho)} \frac{H}{1 - H},$$

(13)

where $\rho$ indicates the average load. This relationship was also used in our previous studies to compare the performance of two FBM processes in different time scales [12], [14].

Note that the probability in Relation (5) is approximated by minimizing the value inside the parenthesis of function $\Phi(\cdot)$ in the RHS of (6), and the critical time scale $m$ expressed by Eq. (13) is exactly the length of time $t$ when the minimum is achieved. Hence, for a certain queue length $x$, the traffic property at critical time scale $m$ contributes most significantly to the formation of queues larger than $x$, while contributions from properties in other time scales can be neglected. Moreover, since a piecewise self-similar process behaves the same as its composing sub-process at the correspondent spectrum of time scale, we can use the property of the piecewise self-similar process at the critical time scale of a certain queue length, i.e. the property of a “critical” or “relevant” self-similar process, to calculate the queueing performance of the piecewise self-similar process at this queue length. With a certain pair of $m \in [s, l]$ and $H_s$, and the known excess bandwidth $C - \lambda$, the corresponding queue length $x$ can be calculated by Eq. (13). Consequently, the queueing performance can be obtained through Eq. (5).

However, as Eq. (5) is used for the performance approximation in each spectrum of time scale, our method has the same limitation as Eq. (5), which holds asymptotically with increased queue length $x$ or offered system load $\rho$.

4. Performance Evaluation Algorithm for Multi-Scale Analysis

The pseudo code of the performance evaluation algorithm is shown in Fig. 4. The two-point method (i.e., Eq. (8)) for calculating Hurst parameters is used. Although the spectrums of time scale of a piecewise self-similar process we defined in last section are non-overlapping on the axis of time scale, each result of the pseudo code is calculated using the medium value of overlapped piece of time scale in order to avoid the “reversed-order” problem as described below, and obtain a smoother performance curve.

For a kinked variance-time property, the Hurst parameter significantly changes around the kink. Suppose that $H_s$ and $H_l$ are two different Hurst parameters at time scales $m_s$ and $m_l$ of a kinked variance-time property. If the difference between these two Hurst parameters is too large, the two values of $x$ calculated using Eq. (13) will have reversed order against the order of corresponding time scales. Note that in log-log scaled coordinates, the horizontal (or vertical) distance between two abscissa (or ordinate) values is the proportion of these two values. For a convexly kinked variance-time property (a similar situation also occurs with concavely kinked one), this “reversed-order” problem can be formulated as: For any two pairs of time scale and corresponding Hurst parameter $(m_s, H_s)$ and $(m_l, H_l)$, where $m_s < m_l$ and $H_s < H_l$, the values of $x$ obtained through Eq. (13) will be $x(m_s, H_s) > x(m_l, H_l)$, if the following relationship holds:

$$\frac{m_l}{m_s} < \frac{H_l(1 - H_s)}{H_s(1 - H_l)}.$$  

(14)

This “reversed-order” problem will prevent the performance curve from being smooth, i.e., with the same time scale $m$, we get different queue length $x$ by Eq. (13). In order to avoid the “reversed-order” situation, the difference of Hurst parameters of two neighboring time scales should not be too large. For example, if $m_s/m_l = 1.1$ and $H_s = 0.60$, then $H_l$ should be no more than 0.62.

Since the distance between two neighboring pieces of time scale is fixed, i.e., $m_s/m_l = m$ in the algorithm in Fig. 4, we can limit the difference of Hurst parameters of two neighboring time scales by adjusting the piece length of time scale (plen in Fig. 4). As it will be shown later, the piece length will significantly affect the calculated value of the Hurst parameters around the kink, while being almost unrelated in the regions where the variance-time property is closed to a straight line in the log-log scaled diagram. To obtain the performance curve smoothened over the whole range of queue length, the piece length of time scale should be selected according to the magnitude of the kink.

In addition to the above mentioned phenomenon, if we have different Hurst parameters $H$’s and variance coefficients $a$’s for the same time scale $m$ (or its corresponding queue length $x$), Eq. (5) will produce different proba-

```
// data used:
// m[i]: time scale, i=0, ..., I, and m[i+1]/m[i] = nn
// v[i]: variance at time scale m[i]
// using two-point method for calculating h[i]

input C, \lambda;  // service and average rates
input mk;  // unit time scale in sec
input plen;  // piece length of time scale

n = plen / nn / 2;

For each i from n to I-n {
    read m[i-n];  // smallest time scale of piece i
    read m[i];  // medium time scale of piece i
    read m[i+n];  // largest time scale of piece i
    read v[i-n], v[i], v[i+n];
    \beta[i] = \log(v[i-n]) / v[i+n] / \log(plen);
    h[i] = 1 - \beta[i] / 2;  // see Eq.(8)
    a[i] = power(10, \log(v[i]) + \beta[i] * \log(m[i]) - \log(mk));  // see Eq.(12)
    x[i] = m[i] * (C-\lambda) * (1-h[i]) / h[i];  // see Eq.(13)
    \gamma[i] = ComputeGamma(h[i], a[i], C, \lambda);  // using Eq.(5)
    p[i] = ComputeProb(x[i], \beta[i], \gamma[i]);  // using Eq.(5)
    write x[i], p[i];
}
```

Fig. 4 Pseudo code of the example algorithm for multi-scale analysis.
In order to verify the efficiency of our proposed method, we compare the analytical results with simulation results using traffic data monitored on SINET and Abilene networks in 2001 and 2002 respectively. The same data traces are used as those in Figs. 1, 2 and 3. SINET data is monitored during the peak hours of the afternoon, and chosen among other data, because they show better Gaussian properties. The Abilene data is obtained from a two-hour data trace starting from 9:00 in the morning, and the starting time and duration of the data used here is chosen so that the long term average rate is kept almost unchanged.

Table 1 gives the statistical parameters at the minimum time scales of these two traces. Simulations were executed using the data traces as inputs to the queue.

### 5.1 Queue Length Distribution

Figures 5 and 6 show the complementary distribution of the queue lengths of the traces for SINET and Abilene data respectively. In each figure, two different values of the load are provided to the queue.

Algorithm in Fig. 4 is used for the analysis. Every point of “Multiscale Analysis” in Figs. 5 and 6 indicates a result obtained through the medium value of each overlapped piece of time scale by Eq. (5) using the parameters of an FBM sub-process. Unless it is designated elsewhere, the two-point method as shown in Eq. (8) is used for calculating the time-scale dependent Hurst parameters.

In the analysis using the proposed method, the piece length of time scale are set to be about 40 and 80 in Figs. 5 and 6 respectively. The two values are chosen to be different because of the fact that a sharper change of Hurst parameter requires a larger piece length, and Abilene data shows a sharper angle in the variance-time plots in Fig. 3.

Since the Gaussian property is well suited for the SINET data, the analysis fits the simulation well, except in the area where simulation results fall suddenly due to the limited length of data traces. For the Abilene data under the load of 80%, the performance obtained from simulation shows somehow an irregular behavior comparing with the SINET traffic. This may result from the deviation from Gaussian property as it is suggested by the comparative large value of high order moments of Abilene data in Table 1, as well as the multifractal-like property indicated by the changing exponents over order $q$ at large time scales in Fig. 2(d).

Due to the limited length of traffic trace used, the variance-time property (Fig. 3) for Abilene data shows fluctuation at large time scales. This also results in a large fluctuation in queue length distribution when the two-point method is used, as shown by the dots denoted by “Multiscale Analysis” in Fig. 6(b). For example, when the offered load is 80%, 10 seconds (corresponding to the abscissa value of 100,000 for Abilene data in Fig. 3) is the critical time scale for the queue length of about $10^8$ bits. This means that the fluctuation of variance-time plots from the abscissa value of 100,000 in Fig. 3 causes the fluctuation of queue length distribution from the queue length of $10^8$ bits in Fig. 6(b). On the other hand, the least square method provides a steady result for queue length distribution as shown by the dots denoted by “Multiscale Analysis (LSM)” in the same figure.

### Table 1

<table>
<thead>
<tr>
<th>Data origin</th>
<th>SINET</th>
<th>Abilene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum time scale (MTS)</td>
<td>1 ms</td>
<td>0.1 ms</td>
</tr>
<tr>
<td>Average rate (bps)</td>
<td>$5.87 \times 10^7$</td>
<td>$4.36 \times 10^8$</td>
</tr>
<tr>
<td>Variance at MTS</td>
<td>$6.06 \times 10^8$</td>
<td>$5.75 \times 10^8$</td>
</tr>
<tr>
<td>Skewness at MTS</td>
<td>-0.16</td>
<td>0.68</td>
</tr>
<tr>
<td>Kurtosis at MTS</td>
<td>-0.49</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Fig. 5 Complementary distribution of queue length for SINET data.
5.2 Impact of Piece Length of Time Scale on Performance

In order to investigate the impact of piece length of time scale on performance, we show here the results when the piece length is changed in the algorithm shown in Fig. 4. Figure 7 captures the differences of the queue lengths obtained by different piece lengths, using Abilene data with an offered load equal to 70%. The distance between two neighboring pieces of time scale, i.e., $m_i/m_{i-1}$, is fixed at 1.1. We can see that the differences in the performance produced by different piece lengths mainly occur at the queue length corresponding to time scale around the kink point.

If the piece length is set to be too small (e.g., 6 or 37 in the figure), as discussed as the “reversed-order” problem in Chapter 4, it will result in a zigzagged performance curve due to the sudden change of Hurst parameter. On the contrary, a piece length that is too large (e.g., 158 in the figure), produces a rather pessimistic result (for convexly kinked variance-time property) while making the performance curve smooth. An appropriate piece length of time scale can be seen as the smallest one that does not result in a reversed order of queue lengths, while producing a smoothened performance curve along a large range of queue lengths.

6. Conclusions

Most internet traffic does not display exact self-similarity, but has more general scaling properties. Methods using a self-similar process to approximate the performance of long-range dependent processes are only efficient for self-similar processes with single Hurst parameters. We proposed an analytical method for solving the problem with the performance evaluations of processes with multi-scaled long-range dependence, by paying attention to the changing scaling properties along time scale. The key point behind our idea is to use the piecewise self-similar process for the performance evaluation.

Our approach is generally focused on a large class of internet traffic. If the Gaussian property can be assumed in the traffic, the analysis can be performed based on only the second-order property along the time scale. Through the comparison of our analytical results with results produced by simulations, we have proved that our method is efficient.

We also provided a systematic algorithm for the performance evaluation using our approach. However, the algorithm is not strong enough against the change of parameter “piece length” around the area where scaling property changes significantly, and we have only provided an empirical method for fixing the parameter value. In order to make our approach more practical, one of our future work will be to find an automatic and efficient way for resolving this problem.

Due to the fact that a large amount of the network traffic do not fit with the Gaussian properties well, our future work will be to extend our method to the direction of applying the notion of the piecewise self-similar process to non-Gaussian traffic. This can be carried out by using other self-similar processes, such as self-similar processes with $\alpha$-stable marginal distributions [23], as the sub-processes of the piecewise self-similar process. The $\alpha$-stable self-similar processes are able to describe the heavy-tailed properties of aggregated traffic. One example of such processes is the linear fractional stable motion (LFSM) or its incremental process linear fractional stable noise (LFSN), an extension of the fractional Brownian motion or its counterpart fractional Gaussian noise. An algorithmic procedure for estimating the
parameters of LFSN processes was given by A. Karasaridis and D. Hatzinakos [30].

Other researches suggest a Poisson assumption in short time scale and a non-stationary view along with long-range dependence in long time scale behavior of network traffic [31], [32]. Through observing the long-term behavior of measured traffic traces, nonstationarity seems to be another remarkable feature in those data. Finding an efficient way for characterizing the nonstationarity may be another direction of our future work.

Acknowledgments

The author would like to thank Professor Shoichiro Asano, Associate Professor Jun Matsukata, Associate Professor Shunji Abe and Dr. Takayuki Fujino for their helpful discussions and invaluable advices. The author also wishes to thank the anonymous reviewers for their comments which helped to improve the quality of this paper. This study is supported in part by JSPS through the Grants-in-Aid for Scientific Research with project numbers 13680502 and 16500046.

References


Appendix

In this paper, we have used two data traces for our analysis, and regarded them as stationary processes. Here we give quantitative analysis to show the feasibility for our assumption of stationarity. Since the data are assumed to be
Gaussian, we only need to show that they are weekly stationary, i.e., their first and second moments keep statistically unchanged along time axis.

Figure A·1 shows the arrival rate per second of the data streams. In each graph, a linear relationship between the arrival rate and time is obtained by using the least square method on all plotting points. For SINET data, the total arrival rate increased during two hours is 0.10% of the average arrival rate; while for Abilene data, the total arrival rate increased during 40 minutes is 1.70% of the average arrival rate. These two lines show that the long-term average arrival rates for both traces are almost kept unchanged with time.

Figure A·2 shows the variance-time plots of partial streams with time lags. In Fig. A·2(a), the two-hour SINET trace is divided into seven one-hour streams, each with 10 minutes time lag against its previous one. In Fig. A·2(b), the 40-minute Abilene trace is divided into five 20-minute data streams, each with five minutes time lag against its previous one. Figure A·2 shows that the second moment of SINET trace exhibit very little change with time, while the second moment of Abilene trace varies in large time scale. This results is compatible with the results in Fig. 2, which shown that Abilene data exhibit more irregularity (deviation from self-similarity) in large time scale.

Yusheng Ji received B.E., M.E. and D.E. in electrical engineering from the University of Tokyo in 1984, 1986 and 1989 respectively. She jointed the National Center for Science Information Systems in 1990. Currently she is an associate professor at the National Institute of Informatics, and the Graduate University for Advanced Studies. Her research interests include network architecture, traffic control and performance analysis for quality of service provisioning in high speed networks. She is also a member of IPSJ and IEEE.