A Binomial Measure Method for Traffic Modeling

Lingbo Pei
Institute of Command Automation
PLA University of Science and Technology
Nanjing China
gemini_nj@126.com

Ming Chen
Institute of Command Automation
PLA University of Science and Technology
Nanjing China
mingchen@public1.ptt.js.cn

Jun Zhou
Institute of Command Automation
PLA University of Science and Technology
Nanjing China
Zhouj_nj@126.com

ABSTRACT

According to some discoveries and analysis results of network measurement in modern communication networks in recent years, much literature has proven that traffic present the nonstationary Poisson characteristic at sub-second time scales in IP backbone, which not accord with self-similarity characteristic for last decade. This paper presents the binomial measure method for traffic characteristic analysis. Some novel properties about the method are also discussed. The method can produces a time series with self-similarity characteristic, and more, realizes the transition from self-similarity distribution to Poisson one. The validity of the method is verified by simulation experiments on NS2. The purpose of the paper is to provide a comprehensive presentation of the binomial measure method base multifractal theory in traffic engineering field.

Categories and Subject Descriptors

I.6.5 [**Simulation and Modeling**]: Model Validation Analysis, C.4 [**Performance of Systems**]: modeling techniques.

General Terms

Experimentation, Performance, Theory.

Keywords

Nonstationary Poisson characteristic, self-similarity, multifractal binomial measure method.

1. INTRODUCTION

Understanding network traffic characteristics and distribution are generally viewed as important steps toward the network architecture analysis and control of network performance. With the urgent demand of quality of network service, more and more attention are paid on the analysis and modeling of Internet traffic.

About ten years ago, numerous network measurement studies showed that Internet traffic exhibited the self-similarity phenomena, which be viewed as a pervasive characteristics in modern high-speed communication networks [1,2,3]. But the self-similarity of network traffic could not be described by the tradition short-dependence model-Poisson model. Many traffic models were put forward to explain the traffic characteristics [4,5,6,7].because the traffic presents multifractal characteristics, wavelet analysis is introduces as a powerful techniques for both modeling and estimation in self-similarity traffic. Multifractal wavelet models have been developed to generate traffic with heavy-tailed distribution and long-range dependence at smaller scales [8].

Starting in 2000, both network structures and applications have developed dramatically, current network traffic can be well represented by the Poisson model for sub-second time scales based on the analysis of the WIDE backbone traces [9]. Paper [10] argues that towards Poisson distribution of traffic has close relations with the link load, and the multiplex gain has impact on it. We think that from the initial stage of Internet to 1984, called the Poisson traffic model dominant stage; from 1984 to early 2000's, called the self-similarity traffic model dominant stage; from early 2000's to now, called the multi-mode stage, i.e., the self-similarity distribution and Poisson distribution coexist extensively in backbone.

The scaling behaviors at small time scales can be described by the multifractal theory, which detailed the self-similarity behaviors. In this paper, we present the binomial measure method for traffic characteristic analysis. Some novel properties about the method are also discussed. The method can produces a time series with self-similarity characteristic, and more, realizes the transition from self-similarity distribution to Poisson one. We study the impact of the binomial measure method on traffic characteristics.

The remainder of this article is structured as follows. Defines about self-similarity process and multifractal measures are discussed in Section 2. An overview of some novel properties about the binomial measure method is given in Section 3. The results of applying the methodology to the case study and some simulation experiments on NS2 are described in Section 4, which shows the applicability and the effectiveness of the methodology related to network traffic. Finally, Section 5 concludes the article and discusses some of the future research works.

2. MULTIFRACTAL MEASURE

Willinger and Wilson in 1993 discovered that self-similarity is an ubiquitous feature of empirically observed network traffic. This is an important notion in the understanding of the traffic's dynamic nature for modeling analysis and control of network performance

Valuetools'07, October 23–25, 2007, Nantes, France.

Self-similarity process is the simple long-range dependence (LRD) traffic model. Its definition of continuous process and discrete time process was presented in [3].

Firstly, we give the notion of self-similar processes.

Definition 1[3]: let X (t) is stationary stochastic process, if X (t) is a self-similar continuous-time stochastic process with stationary increments, namely, for all $a \succ 0$, have

$$\{X(t_1), X(t_2), \cdots, X(t_n)\} \underset{=}{d} \{a^{-H} X(at_1), X(at_2), \cdots, X(a(t_n))\}$$

Where " = " implies the equality of the finite-dimensional distributions, and exponent H is the self-similarity parameter. a is a scaling factor. Intuitively, self-similarity describes the phenomenon in which certain process properties are preserved irrespective of scaling in space or time.

Through the data analysis of Ethernet and WAN traffic measurements, at small-time scales, it has been observed that Internet traffic presents complex scaling and multifractal characteristics. The extensively used multifractal definition is as follow:

Definition 2[11]: In given unit interval I= [0, 1], which has unit quality, let us start by splitting I into n subintervals, δ is defined as the length of subinterval. Let μ_i denote the quality of i-th subinterval ($0 \prec \mu_i \prec 1$), then, $\alpha(t_0)$ denotes singularity exponent of μ_i at t_0 time point. Namely,

$$\alpha(t_0) \equiv \lim_{\delta \to 0} \frac{\lg(\mu_i)}{\lg(\delta)} \qquad (1)$$

According to this definition, we will denote a local parameter H_{local} to mark the traffic variability.

Definition 3: consider a discrete time series; we define the unit with energy E_0, by splitting unit into two subintervals of equal length, with the two subintervals proceed in the same manner. At stage j, have 2^j subintervals exist. By the Equation (1), let $\delta = 2^{-j}$, we have

$$E_i = 2^{-jH_{local}} E_0 \qquad (1 \le i \le j)$$

Usually, supposes E_0=1, then,

$$H_{local} = \frac{\lg E_i}{\lg(2^{-j})}$$

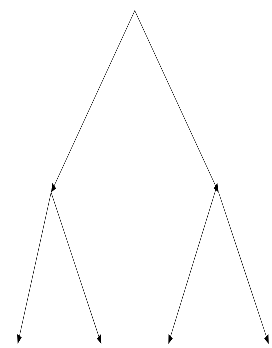

Figure 1. Illustration about binomial measure method.

The binomial measure method is based on multifractal notion, in given unit interval I= [0, 1], we split unit into two subintervals of equal length. And partition the unit quality into r and 1- r . Parameter r is random variable. We called it multiplier. We find that if we select different probability distribution function about r .we will get different binomial measure method. In Figure 1, we give the illustration about binomial measure method. In short, we called the binomial measure method as BM method.

3. BINOMIAL MEASURE METHOD

3.1 Self-similarity Series based on BM Method

Self-similarity describes the phenomenon in which the behavior of a process is preserved irrespective of scaling in space or time. According to hierarchical ideas, we class the Internet into access network and backbone network, just as shown in Figure 4.in Section 4. Traffic on access links is aggregated on backbone irrespective of space or time. We think that multiplier r has been adjusted easily, we suppose that r is independence random variable in I= [0, 1] and obeyed uniform distribution. As follow, we give the strict analysis in theory.

In given unit interval I= [0, 1], we define a random variable U. and split unit interval I into 2^m subinterval. The stage is m. The length of each subinterval is equal. Then k-th subinterval was described as:

$$I_m^k = [k2^{-m}, (k+1)2^{-m}] \quad k = 0,1,\cdots,2^m - 1$$

Then $U(I_m^k)$ is a random variable in subinterval I_m^k

r_{mk} is a random variable, $m = 0,1,2,\cdots,$ $k = 0,1,\cdots,2^m - 1$

r and r_{mk} have the some distribution. Because r is independence random variable in I= [0, 1] and obeyed uniform distribution. Namely, r and 1- r have the some distribution.

Then, we have Recurrence formula:

$$\begin{cases} U(I_{m+1}^{2k}) = r_{mk}U(I_m^k) \\ U(I_{m+1}^{2k+1}) = (1-r_{mk})U(I_m^k) \\ U(I_0^0) = r_{00} \end{cases} \quad (2)$$

$m = 0,1,2,\cdots,\ k = 0,1,\cdots,2^m - 1$

Obviously, r_{mk} is i.i.d random variable, then $U_m = \prod_{i=1}^{m} r_i$

Consider the relation in logarithm,

$$\log U_m = \sum_{i=1}^{m} \log r_i$$

According to the centre limit theorem, when $m \to \infty$, $\log U_m$ obeyed normal distribution. And

Mean is $E\{\log U_m\} = mE\{\log r\}$

Variance is $\text{var}\{\log U_m\} m \,\text{var}\{\log r\}$

So to U_m, its mean is $E\{U_m\} = \{E(r)\}^m$

Variance is $\text{var}\{U_m\} = \{Er^2\}^m - 2^{-2m}$

Let t= 2^{-m}, have $\text{var}\{U_t\} = t^{-\log_2 Er^2} - t^2$

When $m \to \infty$, namely, have $t \to 0$,

$$\text{var}\{U_t\} \sim t^{-\log_2 Er^2}$$

Namely, considering it is function with slowly decline, we think that the time series is self-similar process, and have

$$H = -\frac{1}{2}\log_2 Er^2 \quad (3)$$

3.2 Traffic Characteristic Transition based on BM Method

Many researchers considered traffic modeling from Poisson to self-similarity for ten years. The puzzling question is that how to explain the new phenomenon of nonstationary Poisson characteristic in backbone traffic. We discovered that BM method can gives a reasonable explanation for the new traffic characteristic. We select right multiplier to adjust the network links traffic, then, may transit traffic characteristic from self-similarity to Poisson distribution. As follows, we give two properties about the BM method. We use ρ to replace above r.

Property 1: the BM method has the character of adaptation on invariable traffic. Different values of ρ will cause two states of Network Traffic: 1. $\rho = 0.5$, the characteristic of traffic is sustained; 2. $\rho \succ 0.5$, the velocity on equal area become unequally and burst.

This property is obvious, when ρ is assigned for 0.5 every time, the characteristic of traffic will not change actually, when ρ is assigned for more than 0.5, the unequal of the velocity causes the burst traffic. It can be a way to form self-similarity traffic.

Property 2: The shape function of the BM method on burst traffic represents that different value of ρ will cause three states of the network traffic's characteristics. Suppose the velocity of traffic in continued intervals are m_1 and m_2, and $m_1 \succ m_2$ .then, 1. $\rho = m_1 / (m_1 + m_2)$, the velocity is sustained; 2. $\rho \succ m_1/(m_1+m_2)$, and it is assigned to the bigger velocity of two areas, the velocity will be strengthened; 3. $\rho \succ m_1/(m_1+m_2)$ and it is assigned to the smaller velocity of two areas, the velocity will be smoothly. The follow is the provability of state 3.

Prove:

Suppose the velocities of traffic in continued intervals are m1 and m2, and $m_1 \succ m_2$, set $\Delta_1 = m_1 - m_2$ represents burst of two intervals. According to the binomial m, $\rho(\mathrm{m}_1 + \mathrm{m}_2)$ represents the variety of the first interval, $(1-\rho)(\mathrm{m}_1 + \mathrm{m}_2)$ represents the variety of the second interval, when $\rho \succ m_1/(m_1+m_2)$, and have:

$$\Delta_2 = \rho(m_1 + m_2) - (1-\rho)(m_1 + m_2)$$

Then, $\Delta_2 - \Delta_1 = 2[\rho m_2 - (1-\rho)m_1]$ .so we can see that the burst velocity of two continues intervals become weakly when we used the BM method.

Secondly, we prove the validity about the method in theory, and then in Section 4, we give simulation experiment based on NS2.

Let $\{\mathrm{X(t)}\}_{\mathrm{t}=0,1,2\cdots}$ denote the self-similarity traffic created by the ON/OFF model in LAN, which is a stationary stochastic process. We assume there is M flows, the packet series of m-th flow is $\{\mathrm{X}^{(\mathrm{m})}(\mathrm{t})\}_{\mathrm{t}=0,1,2\cdots}$ so, $\{\mathrm{X(t)}\}_{\mathrm{t}=0,1,2\cdots} = \sum_{m=1}^{M}\{X^{(m)}(t)\}$,

Output flows is $\{\mathrm{Y(t)}\}_{\mathrm{t}=0,1,2\cdots}$, When aggregated flows is inputted into the backbone links, we let $\{\mathrm{Y(t)}\}_{\mathrm{t}=0,1,2\cdots}$ denote the shaped flows by the BM method.

In traffic series, we let packets observed at small-time scales as unit quality, we use the MB method to split the unit and assign the value ρ and 1－ρ to them. We use such process to carry on the division and the adjustment traffic in LAN, and suppose the division process is T, its recurrence formula is:

$$S_{j,k}(t) = \prod_{1 \le j \le T} \rho_{j,k} X(t)$$

$$\begin{cases} X_{j+1}^{2k} = \rho_{j+1} X_j^k \\ X_{j+1}^{2k} = (1-\rho_{j+1}) X_j^k \end{cases} \quad j = 0,1,2\cdots,T \quad k = 0,1,2,\cdots 2^T - 1$$

Moreover, we have

$$S_{j,k} = \rho_{j,k_i} \rho_{j-1,k_{i-1}} \cdots \rho_{1,k_1} X_{0,k_0}$$

$$\begin{cases} \rho_{i,k_i} = \rho_i & if \ k_i = 0; \\ \rho_{i,k_i} = 1-\rho_i & if \ k_i = 1; \end{cases}, \quad 1 \le i \le j$$

Time scale $T_i = 2^{-j} T_0$, where, we think observed small-time scale as unit time, then, T_i may be viewed as piecewise interval, ρ_{i,k_i} is quality factor. So,

$$S(t) = \sum_{k=0}^{k=2^T-1} S_{i,k}$$

Due to ρ_i is a series of independent random variables with $E(\rho_i)$ =1/2, ρ_i and 1- ρ_i have same distribution. Energy function is:

$$E_i = \frac{1}{N_i} \sum_k |d_{i,k}|^2 = E[d_{i,k}{}^2] = E[S(t)^2]$$

Where ρ_i are i,i,d, so E(ρ_i)=E(ρ), therefore,

$$\begin{aligned} E_i &= 2^j E[\rho_1{}^2] E[\rho_2{}^2] \cdots E[\rho_j{}^2] E[X_0{}^2] \\ &= 2^j E[\rho^2]^j E[X_0{}^2] \end{aligned}$$

So, $\log(E_i) = j(1 + \log E[\rho^2]) + \log(E[X_0{}^2])$ (4)

In wavelet transition, consider the relation in logarithm of energy function E_i and scale i, [12] have

$$\log E_i = j(2H_{local} - 1) + C \qquad (5)$$

C is a positive real number, Compare Equation. (4) to Equation(5), then have

$$1 + \log E[\rho^2] = 2H_{local} - 1$$

Above mentioned $E(\rho)$ =1/2, and var(ρ)=1/4,

Have $E(\rho^2)$ = var(ρ)+ $E(\rho)^2$ =0.5

In fact, ρ is obeyed to normal distribution approximately, so,

$$H_{local} \approx 1 + 1/2 \log E[\rho^2] \approx 0.5$$

This result stated that if the traffic series are shaped by the Binomial Measure method, then their local H parameter at small-time scale equal to 0.5, network traffic can be well represented by the Poisson model.

4. CASE STUDY AND SIMULATION

An important issue connected with modeling and performance studies concerns the generation of time series for use in simulations. Firstly, we use the binomial measure method to produces a time series with self-similarity characteristic.

We first think that ρ is an independent random variables in intervals [0, 1], and it obeyed uniform distribution. According to above formula (1), we let j is equal to 18, and produce a time series with length is equal to 2^{18}. The series is shown in Figure 2.

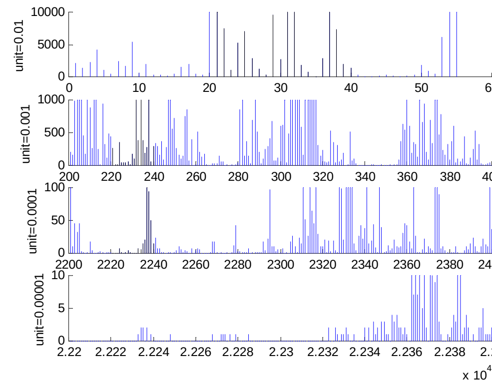

Figure 2. Time series with different interval.

Figure 2.shows a traffic series based on binomial measure method, Where we find that this series present the scale-invariance property. The Figure 2. is similar to the Internet trace in [1]. In Section 3, we proved that the series produced by the binomial measure method has self-similarity characteristics in the theory. Now, we use Variance-Time method to fit H value.

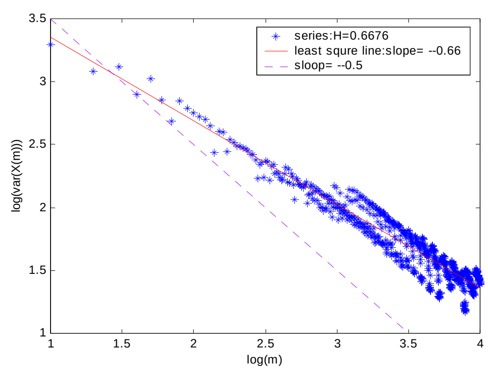

Figure 3. V-T method for fitting H value.

According equation (3) and ρ obeyed uniform distribution, we have $H = 0.692$ in theory, in Figure 2, H parameter is 0.6676. So this algorithm about self-similarity series produce is effective.

Secondly, in order to verify the validity of the binomial measure method to transition from self-similarity distribution to Poisson one, we simulate the model in NS-2[13]. We design a hierarchical network structure with 3 layers: local network, access network and backbone network (see Figure.4). In left-hand local networks, we aggregate lots on/off sources to generate self-similarity traffic which is used as output of local networks, in simulation; we use 10 LAN sources as local networks. The distribution between on and off transitions uses the Pareto distribution, the parameter used in Pareto distribution is α =1.2. The formula of H= (3- α)/2 gives us the relationships between H and α , H is 0.9 theoretically.

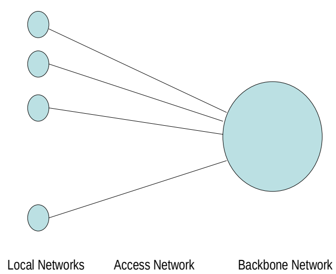

Figure 4. The simulation environment

In our experiments, the bandwidth of local network is 100 Mbps , the bandwidth of access network is 2 Mbps, and the bandwidth of backbone network is 1000Mbps, and links delay is 20 ms, buffer manage adopt tail-drop technology.

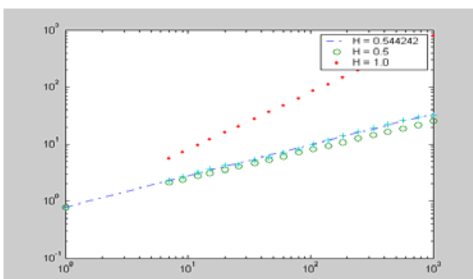

Figure 5. α =1.2, H value with Binomial measure method.

In this experiment, just as shown in Figure 5, we have impacted the traffic in access link with the binomial measure method, and use R/S method to fit the H parameter. It is surprised that above method has great impact on the self-similarity traffic. The impact factor ρ shape the traffic behaviors indeed, the progress toward Poisson can be observed, the value of H is 0.5 approximately.

5. CONCLUSION

Network measurement has proven that traffic present the non-stationary Poisson characteristic at sub-second time scales in IP backbone. Poisson and self-similarity possibly coexist in Internet's different regions, also possibly mutually transforms under the certain conditions. We present the binomial measure method for traffic characteristic analysis. The method can produces a time series with self-similarity characteristic, and more, realizes the transition from self-similarity distribution to Poisson one. The binomial measure method may have important implications for traffic modeling.

6. REFERENCES

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the Self-similar Nature of Ethernet Traffic. In *IEEE/ACM Transactions on Networking*," Vol 2, pp. 1-15, 1994.

[2] Park K., Willinger, W. Self-similar Network Traffic and Performance Evaluation. Hoboken , NJ : Wiley-Interscience , 2000.

[3] T. Karagiannis, M. Molle, M. Faloutsos. Long-Range Dependence –Ten Years of Internet Traffic Modeling. *IEEE Internet Computer*, pp,2-9, 2004.

[4] M.E.Crovella and A. Bestavros, Self-similarity in World Wide Web Traffic: Evidence and Possible Causes. *ACM SIGMETRICS*, pp. 160-169, 1996.

[5] Sally Floyd and Vern Paxson, Why We Don't Know How to Simulate the Internet. Tech, Rep., LBL Network Research Group, 1999.

[6] K. Park, G, Kim, and M, Crovella, On the Relationship Between File Sizes, Transport Protocols, and Self-similar Network Traffic. in proceedings of the *IEEE International Conference on Network Protocols*, 1996.

[7] W. Willinger, M. S. Taqqu, R, Sherman, and D. V. Wilson, Self-similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level. *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71-86, 1997.

[8] V. J. Ribeiro, R.H. Riedio, M .S. Crouse, and R. G. Baraniuk. Simulation of NonGaussian Long-Range-Dependent Traffic Using Wavelets. In *Proc. ACM SIGMETRICS*, pages 1-12, 1999.

[9] Thomas Karagiannis., Mart Molle., Michalis Faloutsos, Andre Broido.A Nonstationary Poisson View of Internet Traffic. *Proc. of the IEEE INFOCOM* 2004. Hong kong: IEEE, 2004.

[10] J.Cao, W. Cleveland, D. Lin, and D. Sun. Internet Traffic Tends to Poisson and Independent as the Load Increases. Bell Labs Technical Report, 2001.

[11] Suo, Cong, e.g., The multifractral model based on the discrete wavelet transform. Journal of china institute communications. 2003, 24(5).

[12] Liang-xiu Han, e. g, characterizing network traffic based on the wavelet technique. Mini-micro system. 2001, vol22. No.9.

[13] http://www.isi.edu/nsnsam/ns