Science in China Series F: Information Sciences

© 2009 SCIENCE IN CHINA PRESS

Springer

www.scichina.com
info.scichina.com
www.springerlink.com

Theoretical mean-variance relationship of IP network traffic based on ON/OFF model

JIN Yi[1†], ZHOU Gang[1,3], JIANG DongChen[1], YUAN Shuai[1], WANG LiLi[2] & CAO JianTing[1]

[1] State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China;

[2] School of Computer Science & Engineering, Beihang University, Beijing 100191, China;

[3] National Digital Switching System Engineering & Technological Research Center, Zhengzhou 450002, China

Mean-variance relationship (MVR), nowadays agreed in power law form, is an important function. It is currently used by traffic matrix estimation as a basic statistical assumption. Because all the existing papers obtain MVR only through empirical ways, they cannot provide theoretical support to power law MVR or the definition of its power exponent. Furthermore, because of the lack of theoretical model, all traffic matrix estimation methods based on MVR have not been theoretically supported yet. By observing both our laboratory and campus network for more than one year, we find that such an empirical MVR is not sufficient to describe actual network traffic. In this paper, we derive a theoretical MVR from ON/OFF model. Then we prove that current empirical power law MVR is generally reasonable by the fact that it is an approximate form of theoretical MVR under specific precondition, which can theoretically support those traffic matrix estimation algorithms of using MVR. Through verifying our MVR by actual observation and public DECPKT traces, we verify that our theoretical MVR is valid and more capable of describing actual network traffic than power law MVR.

traffic matrix, mean-variance relationship, self-similar

1 Introduction

In this paper, we are concerned with theoretical MVR of IP network traffic. MVR is an important function which is currently used as a basic statistical assumption of traffic matrix (TM) estimation[1,2]. TM estimation is a hot research area because of its importance in traffic engineering, reliability analysis, anomalies detection and so on.

A lot of methods of this area hold the same idea that TM is obtained by statistical inference based on indirect measurement (link count measurement for example). These TM estimation methods could be simply described as follows: let Y denote the vector of known incoming and outgoing link counts, X denote the vector of unknown OD counts which are elements of TM. TM estimation is to obtain X on the basis of Y. A linear relation $Y = AX$ is widely agreed, but it is highly undetermined because the amount of unknown OD counts are gener-

Received November 6, 2008; accepted February 11, 2009
doi: 10.1007/s11432-009-0084-y

† Corresponding author (email: jinyi@nlsde.buaa.edu.cn)

Supported by the National Basic Research Program of China (Grant No. G2005CB321901)

ally much more than that of known link counts, i.e. X cannot be directly resolved. Under this circumstance, MVR is adopted as their statistical assumption to infer X, because this function expresses the relation between the first order moment and the second order moment of one traffic rate process, and it can make measurable second order moment data available for estimation. According to MVR, EM[3] and other statistical inference methods are applied to achieve final estimation algorithms.

Because MVR is used as a basis of TM estimation, a proper relationship is crucial to the accuracy of those TM estimation methods. However, this relationship is still under discussion and considered difficult to obtain[2]. To the best of our knowledge, many papers have discussed its form and parameters, while they have all obtained MVR through observing actual network traffic alone and reached the agreement that the relationship is in power law form, i.e. $\sigma^2 \propto \lambda^c$. However, in this form the value of c varies in different network environments making MVR still controversial.

From this perspective, our work is motivated by two points. (i) We consider that power law relationship is reasonable but not sufficient to describe actual network traffic. We have observed network traffic in our laboratory network and campus network for more than one year. We find that though the power law MVR is generally satisfied, there are always some typical traffic traces that cannot be described by it. Therefore, we consider that there should be a more capable MVR. (ii) Although there is no relation between current empirical MVR and the self-similar nature of network traffic, we consider that a proper MVR should be found from some self-similar traffic model. At the very beginning, Vardi[4] assumed MVR as $\sigma^2 = \lambda$. This relationship is theoretically derived from Poisson model which is then considered as the nature of network traffic. When self-similar model is accepted instead of Poisson model, $\sigma^2 = \lambda$ is no longer supported, and till now no new theoretical MVR has been found. Therefore, our work is to find a theoretical MVR derived from self-similar network traffic model.

In this paper, we choose ON/OFF model[5,6] to derive a theoretical MVR. This model is one of the well-known physical explanations of self-similar nature of network traffic. And this approach ensures our relationship to be consistent with self-similarity. Our relationship is more capable than empirical power law MVR to describe actual network traffic and sufficient to support those TM estimation methods of using MVR as their statistical assumption. Our contributions include: (i) We propose an MVR theoretically derived from ON/OFF model. (ii) We explain that the empirical power law MVR is an approximate form of the theoretical MVR and the exponent should be $2H$, where H is Hurst parameter. (iii) We verify the relationship by public DECPKT traces[7] and actual traffic in laboratory and campus network.

2 Related work

2.1 Power law MVR

MVR is now empirically thought to be in power law form, i.e. $\sigma^2 = \varphi\lambda^c$, where φ and c are constants and the value of c is still under discussion[2]. Vardi[4] assumed MVR as $\sigma^2 = \lambda$, which is the property of Poisson model. However, Poisson model is proved to be invalid[8], and as a result, this simple function also fails in theory. Therefore, a lot of papers have tried to find a new MVR to support their methods. Cao et al.[3] follow the power law MVR and choose 2 as approximate integer of c. Subsequent researches validated the relationship as well. They all agreed with the power law form, but they could not reach an agreement on the value of c. Medina et al.[2] evaluated this relationship on 39 POP-pair network, and found that c varies in the range of $[0.5, 4]$. Gunnar et al.[9] observed network traffic on both European and North American core-networks. Their discovery shows that the values of c are 1.5 and 1.6 respectively. Juva et al.[10] proposed a method based on link count covariances, with 1.5 chosen as the best value of c through mathematical optimization.

All the above works agree that MVR is in power law form. However, they failed to provide any theoretical support. And in all these works, the value of c is still considered because there is no exact definition nor explanation on c yet. In contrast, our paper is quite different in that we provide MVR theoretically, validate its power law form, and deduce an exact definition of c.

2.2 Self-similarity and ON/OFF model

Leland et al.[11] did the pioneering work which caused the failure of Poisson model. As a result, previous work derived from Poisson model required verification over again. Although MVR is no longer $\sigma^2 = \lambda$ as derived by the Poisson model, we consider that obtaining MVR from network traffic model is a reasonable approach and there should be a theoretical MVR.

As a plausible physical explanation for network traffic's self-similar nature, ON/OFF model is one of the generally accepted models nowadays. Willinger et al.[6] proposed the definition of ON/OFF model, which provided a good foundation for further discussion. Our work is based on this model so that our conclusion could be consistent with the self-similarity of network traffic that satisfies ON/OFF model.

3 Theoretical MVR derivation

3.1 ON/OFF model definition

Because our derivation is based on ON/OFF model which has been detailed discussed by Willinger et al.[6], we briefly introduce its definition in this subsection.

Suppose that there are M i.i.d. ON/OFF sources. To specify the distributions of both ON- and OFF-periods, let $f_1(x)$, $F_1(x)$, $F_{1c}(x)$, μ_1, σ_1^2 denote the probability density function, cumulative distribution function, complementary distribution, mean length and variance of an ON-period, and let $f_2(x)$, $F_2(x)$, $F_{2c}(x)$, μ_2, σ_2^2 correspond to an OFF-period. Assume $x \to \infty$, either

$$F_{1c}(x) \sim \ell_1 x^{-\alpha_1} L_1(x) \text{ with } 1 < \alpha_1 < 2$$

or $\sigma_1^2 < \infty$, and either

$$F_{2c}(x) \sim \ell_2 x^{-\alpha_2} L_2(x) \text{ with } 1 < \alpha_2 < 2$$

or $\sigma_2^2 < \infty$, where $\ell_j > 0$ is a constant and $L_j > 0$ is a slowly varying function at infinity. Additionally, when $1 < \alpha_j < 2$, set $a_j = \ell_j(\Gamma(2-\alpha_j))/(\alpha_j - 1)$. When $\sigma_j^2 < \infty$, set $\alpha_j = 2$, $L_j \equiv 1$ and $a_j = \sigma_j^2$.

For large T and M, in interval $[0, Tt]$, the aggregate cumulative traffic process behaves statistically like

$$TM\frac{\mu_1}{\mu_1+\mu_2}t + T^H\sqrt{L(T)M}\sigma_{\lim}B_H(t), \qquad (1)$$

where $H = (3 - \alpha_{\min})/2$ is a shape parameter, $L(T)$ is a slowly varying function, and $B_H(t)$ is fractional Brownian motion with zero mean and t^{2H} variance. $\sigma_{\lim}$ has two possible definitions, depending on whether

$$\Lambda = \lim_{t\to\infty} t^{\alpha_2-\alpha_1}\frac{L_1(t)}{L_2(t)} \qquad (2)$$

is finite, 0, or infinite. If $0 < \Lambda < \infty$, set $\alpha_{\min} = \alpha_1 = \alpha_2$, $L = L_2$ and

$$\sigma_{\lim}^2 = \frac{2(\mu_2^2 a_1 \Lambda + \mu_1^2 a_2)}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{\min})}. \qquad (3)$$

If, otherwise, $\Lambda = 0$ or $\Lambda = \infty$, set $L = L_{\min}$ and

$$\sigma_{\lim}^2 = \frac{2\mu_{\max}^2 a_{\min}}{(\mu_1+\mu_2)^3\Gamma(4-\alpha_{\min})}, \qquad (4)$$

where $\min = 1, \max = 2$ if $\Lambda = \infty$, and $\min = 2, \max = 1$ if $\Lambda = 0$.

3.2 Exact MVR derivation

The network traffic process is exactly defined by (1). In this equation, $B_H(t)$ is a fractional Brownian motion which has $E(B_H(t)) = 0$ and $D(B_H(t)) = t^{2H}$. Therefore, we have preliminary mean and variance of the process:

$$\begin{aligned} E &= E\left(TM\frac{\mu_1}{\mu_1+\mu_2}t + T^H\sqrt{L(T)M}\sigma_{\lim}B_H(t)\right) \\ &= E\left(TM\frac{\mu_1}{\mu_1+\mu_2}t\right) = TM\frac{\mu_1}{\mu_1+\mu_2}t, \\ \sigma^2 &= D\left(TM\frac{\mu_1}{\mu_1+\mu_2}t + T^H\sqrt{L(T)M}\sigma_{\lim}B_H(t)\right) \\ &= D(T^H\sqrt{L(T)M}\sigma_{\lim}B_H(t)) \\ &= (Tt)^{2H}L(T)M\sigma_{\lim}^2. \end{aligned}$$

Then, the definitions of $L(T)$ and $\sigma_{\lim}^2$ should be determined. Because $L_1(t)$ and $L_2(t)$ are slowly varying functions, we consider $\lim_{t\to\infty}\frac{L_1(t)}{L_2(t)}$ as constant. Thus the relation between α_1 and α_2 determines the definition of $\sigma_{\lim}^2$. There are three possible cases: $\alpha_1 > \alpha_2$, $\alpha_1 = \alpha_2$, and $\alpha_1 < \alpha_2$. They are discussed as follows.

(i) $\alpha_1 > \alpha_2$. Then we have $\alpha_2 - \alpha_1 < 0$. Thus, we have

$$\Lambda = \lim_{t\to\infty} t^{\alpha_2-\alpha_1} \lim_{t\to\infty} \frac{L_1(t)}{L_2(t)} = 0 \lim_{t\to\infty} \frac{L_1(t)}{L_2(t)} = 0.$$

Therefore, based on definition of ON/OFF model, we have $H = (3-\alpha_2)/2$, $L(T) = L_2$, and $\sigma^2_{\lim}$ follows from (4). By substituting $L(T)$ and $\sigma^2_{\lim}$, we have

$$\begin{aligned}\sigma^2 &= (Tt)^{2H}\ell_2 L_2 M \frac{\mu_1^2}{(\mu_1+\mu_2)^3} \\ &\quad \cdot \frac{2\Gamma(2-\alpha_2)}{\Gamma(4-\alpha_2)(\alpha_2-1)} \\ &= (Tt)^{2H}\ell_2 L_2 M \frac{\mu_1^2}{(\mu_1+\mu_2)^3} \\ &\quad \cdot \frac{2}{(3-\alpha_2)(2-\alpha_2)(\alpha_2-1)}. \end{aligned} \tag{5}$$

In (5), $\ell_2 L_2$ is determined by the distribution of OFF-period length. Nowadays, many papers agree that the packet interval follows a heavy-tailed distribution[12], but actual distribution function is still controversial. According to the definition of ON/OFF, the heavy-tailed distribution of ON- or OFF-period length should be power-law distribution as well, because when $1 < \alpha_j < 2$, $F_{jc}(x) \sim \ell_j x^{-\alpha_j} L_j(x)$ should be satisfied. There are several familiar heavy-tailed distributions[13]: Pareto, Lognormal, inversed Weibull, etc. Only Pareto distribution satisfies this restriction. Thus, we choose Pareto as simplest heavy-tailed distribution.

According to the definition of Pareto distribution, the complementary distribution function of OFF-period length satisfies $F_{2c}(x) = {m_2}^{\alpha_2} x^{-\alpha_2} \sim \ell_2 x^{-\alpha_2} L_2(x)$, where m_2 is the minimum possible value of x. Then we have $\ell_2 L_2 = {m_2}^{\alpha_2}$. According to the property of Pareto distribution, the mean of OFF-period length $\mu_2 = m_2 \frac{\alpha_2}{\alpha_2-1}$, i.e. $m_2 = \mu_2 \frac{\alpha_2-1}{\alpha_2}$. Then the relation between $\ell_2 L_2$ and μ_2 is

$$\ell_2 L_2 = \left(\mu_2 \frac{\alpha_2-1}{\alpha_2}\right)^{\alpha_2} = (\mu_2)^{\alpha_2} \left(\frac{\alpha_2-1}{\alpha_2}\right)^{\alpha_2}. \tag{6}$$

With (5) and (6), for a given interval $[0, Tt]$, we can get refined mean and variance of traffic process:

$$E = TtM \frac{\mu_1}{\mu_1+\mu_2},$$

$$\sigma^2 = (Tt)^{2H} M \frac{\mu_2^{\alpha_2}\mu_1^2}{(\mu_1+\mu_2)^3} f(\alpha_2), \tag{7}$$

where

$$f(\alpha) = \frac{2}{(3-\alpha)(2-\alpha)(\alpha-1)} \left(\frac{\alpha-1}{\alpha}\right)^{\alpha}.$$

Because we are usually interested in the average rate of traffic process, for a given interval $[0, Tt]$, we can get the mean and variance of network traffic's average rate directly from (7),

$$\begin{aligned} E_r &= E/(Tt) = M\frac{\mu_1}{\mu_1+\mu_2}, \\ \sigma_r^2 &= \sigma^2/(Tt)^2 = (Tt)^{2H-2} M \frac{\mu_2^{\alpha_2}\mu_1^2}{(\mu_1+\mu_2)^3} f(\alpha_2). \end{aligned} \tag{8}$$

In (8), T, t, and M are constants. Because α_1 and α_2 are shape parameters that describe the heavy-tailed nature of ON- and OFF-period in a long time range, we consider them constant as well. Therefore, there are two variables μ_1 and μ_2 in (8). To express the relation between μ_1 and μ_2, define

$$\eta = \frac{\mu_1}{\mu_1+\mu_2}. \tag{9}$$

Getting rid of μ_2 by (9) in (8), we have new expressions of E_r and σ_r^2:

$$\begin{aligned} E_r &= M\eta, \\ \sigma_r^2 &= MB_{\alpha_2}(\eta)(Tt/\mu_1)^{1-\alpha_2} f(\alpha_2), \end{aligned} \tag{10}$$

where

$$B_\alpha(\eta) = (1-\eta)^\alpha \eta^{3-\alpha}. \tag{11}$$

(ii) $\alpha_1 = \alpha_2$. In this case, we have $t^{\alpha_2-\alpha_1} = 1$; thus Λ is constant. Then based on ON/OFF definition, we have

$$\sigma_r^2 = M((1-\eta)^2\eta + (1-\eta)^\alpha\eta^{3-\alpha})(Tt/\mu_1)^{1-\alpha} f(\alpha),$$

where $\alpha = \alpha_1 = \alpha_2$. Thus

$$\sigma_r^2 = MB'_\alpha(\eta)(Tt/\mu_1)^{1-\alpha} f(\alpha), \tag{12}$$

where

$$B'_\alpha(\eta) = (1-\eta)^2\eta + (1-\eta)^\alpha\eta^{3-\alpha}. \tag{13}$$

(iii) $\alpha_1 < \alpha_2$. In this case, $\alpha_2 - \alpha_1 > 0$; thus Λ is infinite. According to the definition of ON/OFF model, we have

$$\sigma_r^2 = M(1-\eta)^2\eta(Tt/\mu_1)^{1-\alpha_1}.$$

Thus

$$\sigma_r^2 = MB''_{\alpha_1}(\eta)(Tt/\mu_1)^{1-\alpha_1} f(\alpha_1), \tag{14}$$

where

$$B''_\alpha(\eta) = (1-\eta)^2\eta. \tag{15}$$

According to the discussion on the three cases above, it is easy to find that the mean of traffic average rate is defined as (10), and the variance is strictly decided by $B_\alpha(\eta)$ which has different expressions under different relations between α_1 and α_2.

When ON/OFF model was proposed in 1995, MVR was not mentioned. Thus ON/OFF model only focuses on explaining the self-similar nature of network traffic. Therefore, in ref. [6] the relation between α_1 and α_2 is not restricted, since this relation does not affect the self-similar property of the model. In other words, as for MVR, ON/OFF model is not strict enough. Obviously, this relation is the key to MVR. A proper relation between α_1 and α_2 should be chosen by experimental methods.

In this paper, we choose $\alpha_1 > \alpha_2$ by the following reasons. First, Willinger et al.[6] have made typical experiments, based on which they conclude that $\alpha_1 > \alpha_2$ is satisfied in LAN traffic; and in WAN traffic α_2 is generally very close to 1, even smaller than 1 sometimes, and α_1 is generally close to 2. These experiments indicate that the relation $\alpha_1 > \alpha_2$ is satisfied in both LAN and WAN traffics. Second, we observe actual network traffic to verify this relation as well. We choose t_0=50 ms as a threshold to divide packet intervals into ON- and OFF-periods, and detect their tails by qq-plot[14]. Figure 1 is one typical experiment result, in which we use least squares line to evaluate both tails, and get $\alpha_1 = 1.91$, $\alpha_2 = 1.14$, i.e. $\alpha_1 > \alpha_2$. Finally, in section 5, we provide more discussion to explain that $\alpha_1 \leqslant \alpha_2$ is unsuitable for actual network traffic.

Therefore, according to (10), exact MVR of network traffic average rate in interval $[0, Tt]$ is

$$\sigma_r^2 = MB_{\alpha_2}\left(\frac{E_r}{M}\right)(Tt/\mu_1)^{1-\alpha_2} f(\alpha_2), \tag{16}$$

where

$$B_\alpha(\eta) = (1-\eta)^\alpha \eta^{3-\alpha}.$$

3.3 Properties of exact MVR

μ_1 and η are two variables in exact MVR. In this subsection, we discuss the properties of exact MVR (i)with constant μ_1, (ii)with constant η, and (iii)with varying μ_1 and η by the same trend. From the properties of exact MVR under these three preconditions, we derive sufficient theoretical functions and explanations to describe actual network traffic. Furthermore, we explain the rationality of empirical power law MVR and provide exact definition of the exponent c as well.

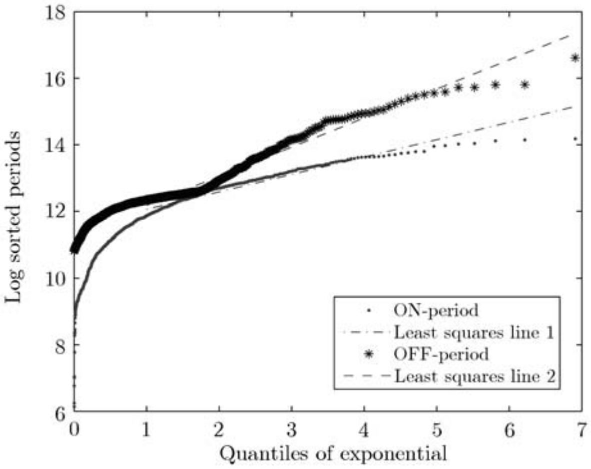

Figure 1 qq-plots for ON- and OFF-periods.

(i) μ_1 is constant. When μ_1 is constant, MVR in interval $[0, Tt]$ can be expressed as

$$\sigma_r^2 = \Phi B_{\alpha_2}\left(\frac{E_r}{M}\right), \tag{17}$$

where $\Phi = M(Tt/\mu_1)^{1-\alpha_2} f(\alpha_2)$ is constant.

It is obvious that the property of MVR under this condition is only determined by function $B_\alpha(\eta)$. Thus, we first discuss the property of $B_\alpha(\eta)$ here.

$B_\alpha(\eta)$ in log-log scale with different parameter α are drawn together in Figure 2. The logarithmic form of $B_\alpha(\eta)$ is $\log(B_\alpha(\eta)) = (3-\alpha)\log(\eta) + \alpha\log(1-\eta)$. Then the derivative of $\log(B_\alpha(\eta))$ is

$$\frac{\partial \log(B_\alpha(\eta))}{\partial \log \eta} = (3-\alpha) - \alpha\frac{\eta}{1-\eta}. \tag{18}$$

It is easy to know that when $\eta = 1 - \alpha/3$, (18) is equal to 0, i.e. when $\eta = 1-\alpha/3$, $B_\alpha(\eta)$ gets its maximum value. When $1-\alpha/3 < \eta < 1$, the derivative of $\log(B_\alpha(\eta)$ is negative, and it is obvious that $\lim_{\eta\to 1}\log B_\alpha(\eta) = -\infty$. In (18), when $0 < \eta \ll 1-\alpha/3$, $\alpha\frac{\eta}{1-\eta}$ is much smaller than $3-\alpha$. Therefore, $B_\alpha(\eta)$ behaves approximately like a power law function

$$B_\alpha(\eta) \approx \eta^{3-\alpha}. \tag{19}$$

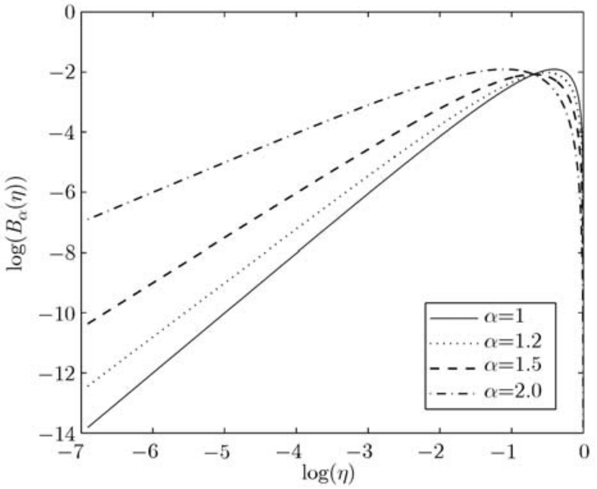

Figure 2 $B_\alpha(\eta)$ with different α in log-log scale.

Because there is a coefficient $\frac{1}{M}$ in (17), we discuss the property of $B_\alpha(p\eta)$ where p is a constant coefficient. We draw function $B_\alpha(p\eta)$ in log-log scale with different p together in Figure 3 which implies that the graph of function $B_\alpha(p\eta)$ is only moved by p with slight distortion. Therefore, the shape of $\sigma_r^2(E_r)$ should be the same as $B_\alpha(\eta)$.

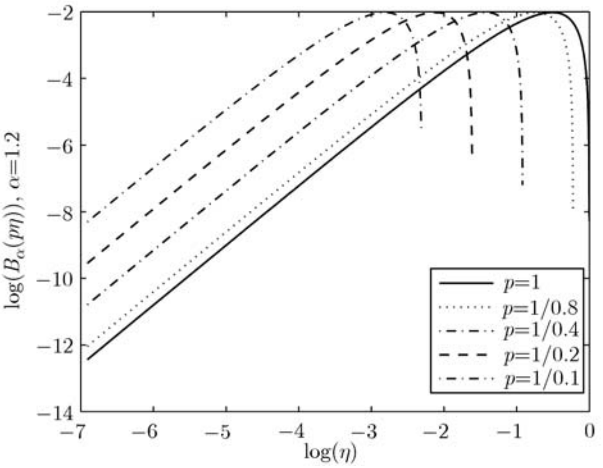

Figure 3 $B_\alpha(p\eta)$ with different p in log-log scale.

According to (19), it is easy to find that when $\frac{E_r}{M} = \eta \ll 1 - \alpha/3$, exact MVR can be simplified into

$$\sigma_r^2 = \Phi B_{\alpha_2}\left(\frac{E_r}{M}\right) \approx \Phi\left(\frac{E_r}{M}\right)^{3-\alpha_2}$$
$$= \varphi E_r^{3-\alpha_2} = \varphi E_r^{2H},$$

where $\varphi = M^{1-2H}(Tt/\mu_1)^{1-\alpha_2} f(\alpha_2)$ is constant.

Thus, when μ_1 is constant and $\frac{E_r}{M} = \eta \ll 1-\alpha/3$, we have approximate power law MVR of network traffic average rate in interval $[0, Tt]$:

$$\sigma_r^2 \approx \varphi E_r^c, \tag{20}$$

where $c = 2H$. If H is close to 1, we can get the simplest form

$$\sigma_r^2 \approx \varphi E_r^2. \tag{21}$$

(20) proves that empirical power law MVR is a reasonable MVR which is an approximate form of our exact MVR.

(ii) η is constant. According to (10), when η is constant, E_r is obviously constant. Thus the derivative of MVR in log-log scale will be infinity in theory, i.e. σ_r^2 is irrelative with E_r.

(iii) μ_1 and η vary by the same trend. This case means that μ_1 and η vary both bigger or smaller. According to (10), $\sigma_r^2(\mu_1)$ is a monotonic increasing function. Assume that $\frac{E_r}{M} \ll 1 - \alpha/3$ is satisfied, as (19) described, η approximately contributes $2H$ to the derivative of MVR in log-log scale, based on which the varying μ_1 adds positive increment to this derivative. Therefore, when $\frac{E_r}{M} \ll 1-\alpha/3$, the derivative of MVR is greater than $2H$.

3.4 Comprehension on μ_1 and η

In this subsection, we provide explanation on μ_1 and η, which makes our discussion on MVR easier to comprehend. Consider a scenario in which there is only one program. There should be an ON-period when this program runs automatically since the intervals between packets are usually small. In ON-period, the interval is mainly determined by application logic, communication protocol, and network environment, i.e. if these factors are stable, the distribution of ON-period length is stable. In contrast, when the program is blocked or terminated, an OFF-period should appear. In OFF-period, the interval is mainly determined by human actions[15] which may vary much. For example, people generally access many Websites by day and go to sleep by night, i.e. people trig HTTP traffic more frequently by day than by night. Therefore, we consider that μ_1 is mainly determined by actual application and communication environment which are generally stable. Plus, η is considered to reflect the busyness of actual application, since it is the ratio of μ_1 to $\mu_1 + \mu_2$.

4 Verification

4.1 Environments and methods

We choose our laboratory network and campus network as environments for observing network traffic. As the first environment, our laboratory network is a typical LAN, in which the traffic is generally at low volume and fluctuates acutely. As the second environment, the campus network transports the whole university's traffic. We consider it a typical fast network. In campus network, the traffics are aggregated by a huge number of sources; thus it is usually at high volume and does not fluctuate very acutely. Besides, the campus network owns an entry to interconnect with China Education and Research Network (CERNET, a 38-PoP backbone). We consider the traffic of this entry as typical backbone traffic.

In order to verify our MVR, two methods are adopted to observe the traffic: (i) through SNMP and (ii) through dumping packets directly. For the former, we observe the incoming and outgoing rates of all the interfaces of typical network devices. The interval is 30 s, i.e. tT=30 s. By this method, we keep observing actual network traffic for more than one year. The measurement data are used for checking Hurst parameter and verifying MVR. For the latter mechanism, we directly dump packets by TCPDUMP whose data are mainly used for verifying ON/OFF model.

4.2 MVR for specific applications

Besides making observation simpler, we consider that the traffics generated by different applications should have different statistical characteristics, and μ_1 is stable for the same application. Therefore, we first verify MVR by the traffics generated by HTTP and P2P which are two typical applications.

As shown in Figure 4, we dump pure HTTP packets from the link interconnecting laboratory network with campus network, and test its Hurst parameter H by wavelet method[16]. We can find that the MVR of HTTP traffic satisfies power law MVR very well, and c is very close to $2H$.

Second, we use the same method to process BitTorent traffic. We can find that the higher the volume of traffic is, the more lightly it fluctuates (Figure 5). When the traffic is at low volume, power law MVR is satisfied, but when the traffic is at high volume, power law MVR is not suitable any more. The failure of power law MVR is obviously reflected by Figure 6 which is drawn by a part of the traffic when the application transports data busily.

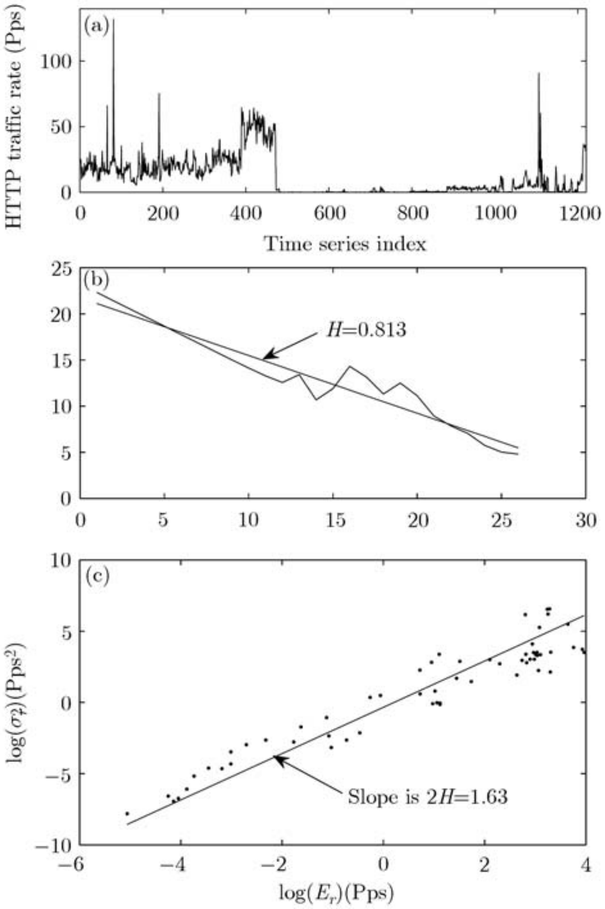

Figure 4 HTTP traffic with 30 s interval. (a) Time series; (b) Hurst estimation; (c) MVR in log-log scale.

The above observations show that our theoretical MVR can describe actual HTTP and P2P traffics well. First, when power law MVR is satisfied, the exponent is close to $2H$ (see (20)). Second, the empirical power law MVR is not suitable all the time. When power law MVR fails, the MVR of P2P traffic matches conclusion (17) very well. Finally, MVRs of HTTP and P2P traffics satisfy (17) well which is derived from exact MVR by constant μ_1. This supports our explanation on μ_1 that for the traffic generated by the same application, μ_1 is stable.

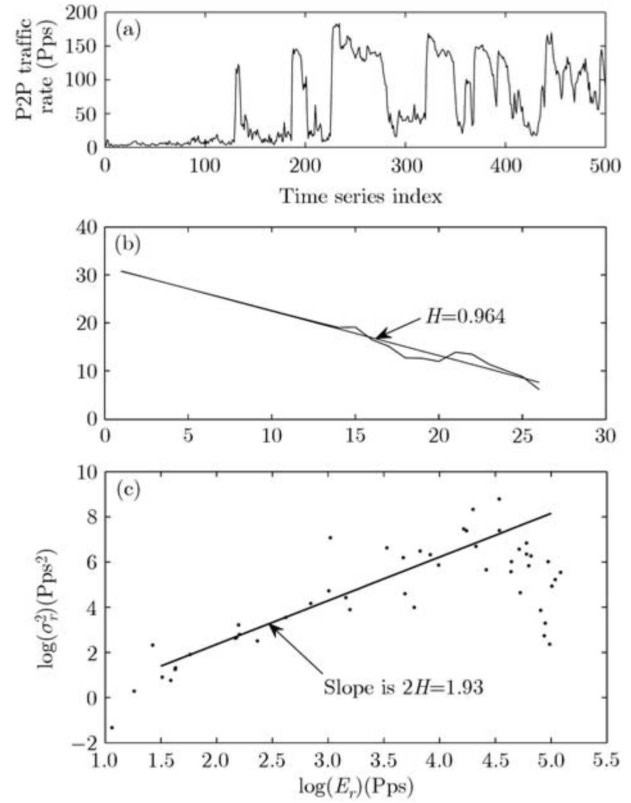

Figure 5 BitTorent traffic with 30 s interval. (a) Time series; (b) Hurst estimation; (c) MVR in log-log scale.

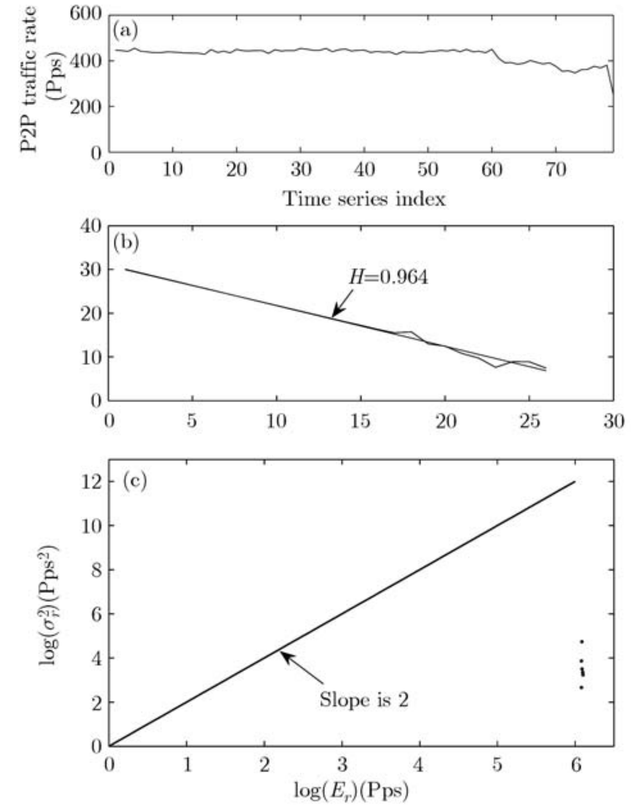

Figure 6 BitTorent traffic (busy) with 30 s interval. (a) Time series; (b) Hurst estimation; (c) MVR in log-log scale.

4.3 MVRs for different interfaces

As for TM estimation, the traffics of different interfaces are statistically inferred together; thus we verify the MVR of the traffics of different interfaces belonging to the same device in this subsection. We choose three devices that represent different typical network traffic cases from the two environments.

The first device is a core switch of campus network. This device has highly aggregated traffics generated by the whole campus. All MVRs of the incoming traffics of this switch are shown together in Figure 7. The second device is the core switch of our laboratory network. It usually has high-volume traffic, while its sources are far less than those of the first device. Its MVRs of incoming and outgoing traffics are shown in Figure 8. The third device is a marginal switch of our laboratory network, with only several hosts connected. The traffic passing through it is usually generated by a small quantity of sources. Its MVRs of incoming and outgoing traffics are shown in Figure 9.

Then we verify our relationship using these devices. First of all, we can find that actual MVR is generally relative to the traffic volume. As shown in Figures 7 and 8, approximate power law MVR is correct as a whole for high-volume traffic, and a line with a slope of $2H$ is those traffics' best trend line, which verifies that (20) is correct. Oppositely, as shown in Figure 9, when the traffics are at low volume, the slope is generally more than $2H$, even close to infinity sometimes. It shows that approximate power law MVR is no longer suitable. This phenomenon could be explained by cases (ii) and (iii) in subsection 3.3. We explain that when the traffic is at low volume, the ratio of different application traffics is easy to change, i.e. μ_1 is no longer stable. Second, the approximate power law MVR satisfies the traffic of campus network better than that of laboratory network, even when the traffic is at low volume in campus network. We consider that because there are much more sources generating traffic in campus network than those in lab-oratory network, the radio of different applica-

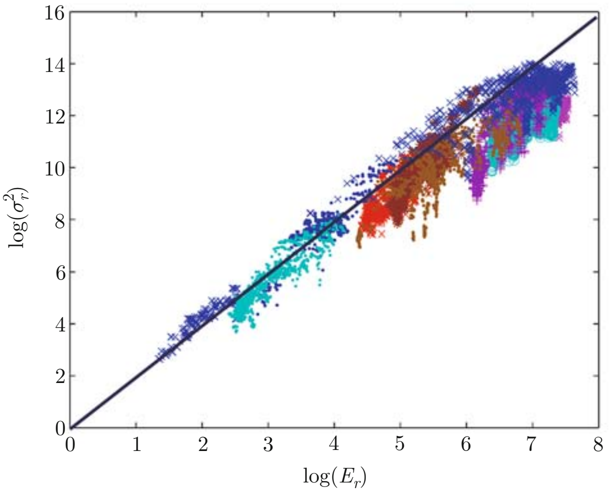

Figure 7 MVR for core switch of campus network.

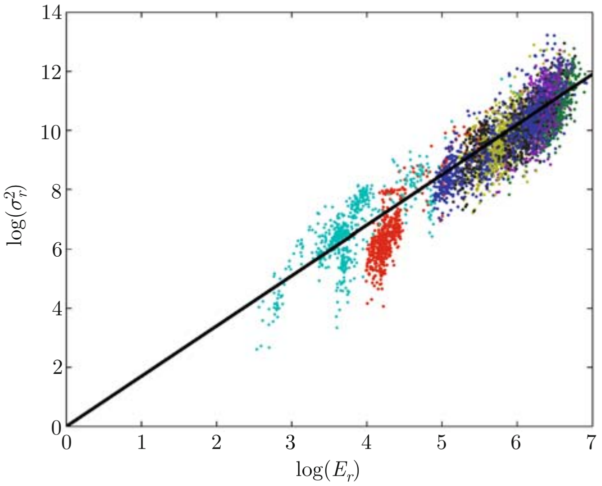

Figure 8 MVR for core switch of laboratory network.

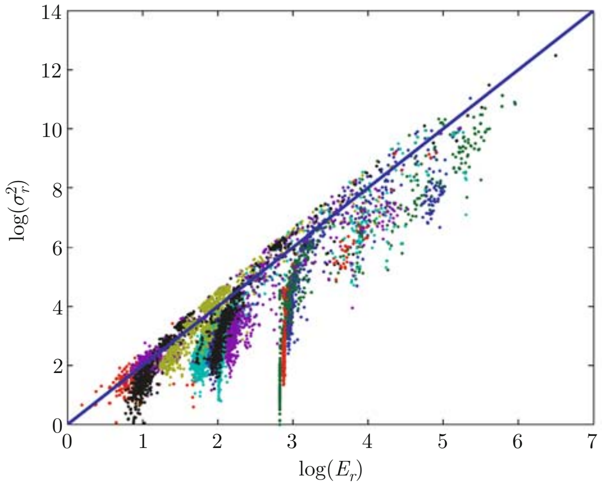

Figure 9 MVR for marginal switch of laboratory network.

tion traffics is stable, i.e. μ_1 is stable, while for LAN traffic aggregated by a small amount of sources, the ratio of different application traffics is easy to change, i.e. μ_1 is unstable.

Therefore, power law MVR is generally reasonable for the traffics which are aggregated by many sources or at proper volume, i.e. empirical power law MVR relies on stable ratio of different application traffics and proper business. This shows that it is not occasional that all previous observations support power law MVR. Our theoretical MVR can describe more cases of actual network traffic than empirical power law MVR.

4.4 Relationship for CERNET entry

CERNET is 38-PoP backbone network for education and research in China. We especially observe the CERNET entry of our campus network to verify our MVR for backbone traffic. The incoming and outgoing traffics of CERNET entry are shown in Figure 10. The estimated Hurst parameters of both traffics are 1.021 and 1.019, which implies that the traffics are all highly self-similar. Then we can find that the MVRs of both traffics follow

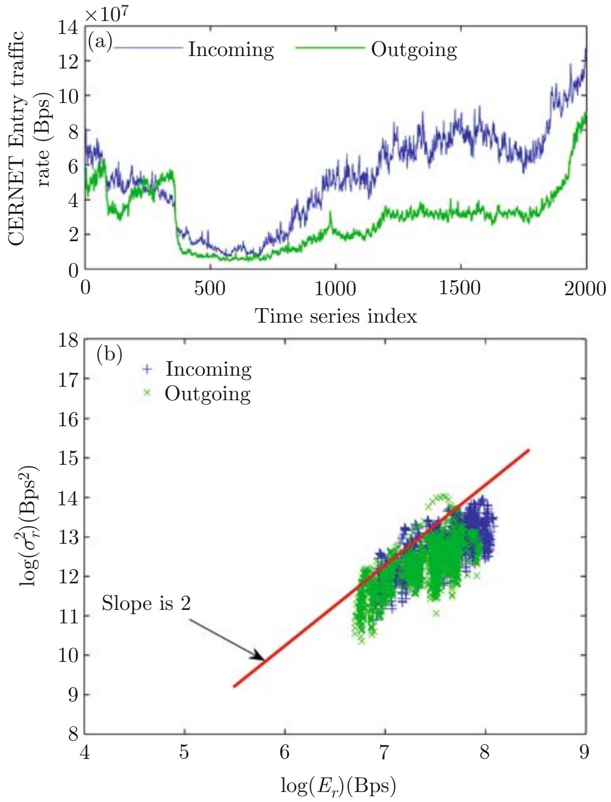

Figure 10 CERNET entry traffics with 30 s interval. (a) Time series; (b) MVR in log-log scale.

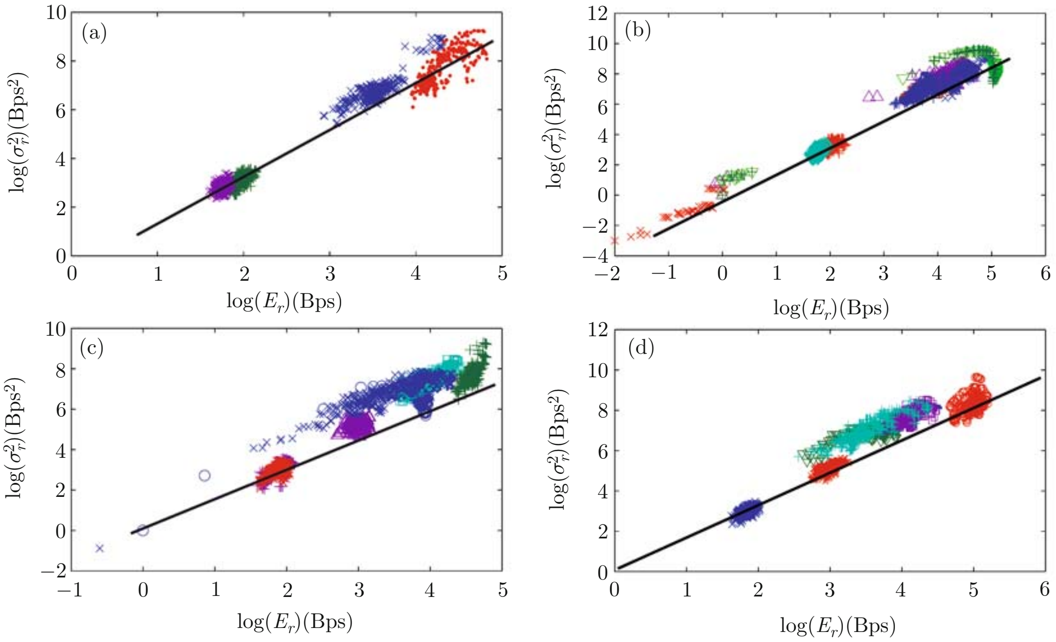

Figure 11 MVRs for DEC-PKT traces with 10 s interval. (a) dec-pkt-1; (b) dec-pkt-2; (c) dec-pkt-3; (d) dec-pkt-4.

power law form, and the slopes of their trend lines are close to 2, i.e. $2H$. For back bone traffic, the ratio of different application traffics is very stable, since the traffic is generated by a huge number of sources. Thus the assumption that μ_1 is constant is reasonable, i.e. (17) and (20) can describe backbone traffic very well.

4.5 MVR for DEC-PKT traces

DEC-PKT traces involve four traces which are used by Paxon et al.[8] to verify the self-similar nature of network traffic. The details of these traces are shown in Table 1. Each trace contains an hour's worth of all wide-area traffic between Digital Equipment Corporation and the rest of the world. In this subsection, we verify MVR by dominant TCP flows in these traces. The Hurst value of these flows are between 0.75 and 0.92. As shown in Figure 11, the power law MVR is generally satisfied, and the exponent is between 1.5 and 1.8 i.e. c is close to $2H$. This shows that for classic traces, (17) and (20) are satisfied very well.

Table 1 DECPKT traces info

Trace	Start time	Packets
dec-pkt-1	22:00, March 8, 1995	3.3 million
dec-pkt-2	02:00, March 9, 1995	3.9 million
dec-pkt-3	10:00, March 9, 1995	4.3 million
dec-pkt-4	14:00, March 9, 1995	5.7 million

5 Discussion on $\alpha_1 \leqslant \alpha_2$

In this section, we show that when $\alpha_1 \leqslant \alpha_2$, MVR should be irrelative with actual value of H, which implies that the condition of $\alpha_1 \leqslant \alpha_2$ is inconsequent with actual observation.

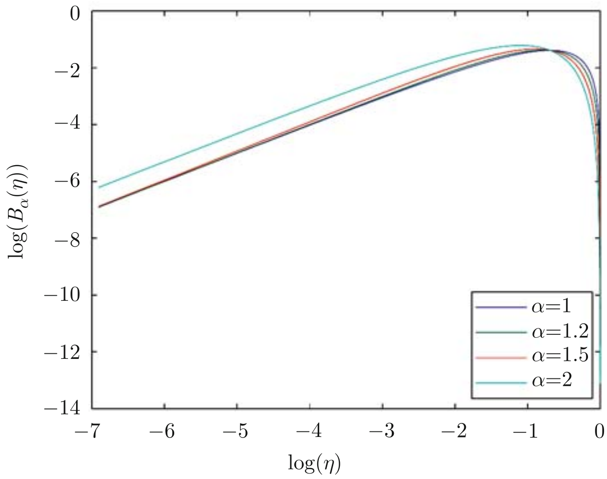

Figure 12 $B'_\alpha(\eta)$.

As shown in Figure 12, when $\eta \ll 1-3/\alpha$, the derivative of $B'_\alpha(\eta)$ in log-log scale is close to 1, which is not consistent with actual observation. Thereby the assumption of $\alpha_1 = \alpha_2$ is unreasonable. In the same way, as shown in Figure 13, when $\eta \ll 1-3/\alpha$, the derivative of $B''_\alpha(\eta)$ in log-log scale is always close to 1, which is not consistent with actual observation either. Therefore, assumption $\alpha_1 < \alpha_2$ is unreasonable.

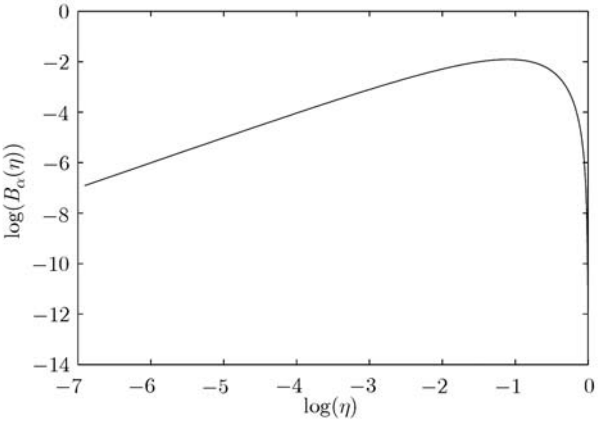

Figure 13 $B''_\alpha(\eta)$.

6 Conclusion

In this paper, we derive an exact mean-variance relationship (MVR) of IP network traffic from ON/OFF model, which is a theoretical approach. Therefore, our relationship is consistent with self-similar nature of network traffic that satisfies ON/OFF model. Then under specific precondition, we derive the power law MVR from our exact MVR, which proves that the empirical power law MVR is an approximate form of our relationship, and the exponent c should be set at $2H$. These conclusions imply that our relationship can theoretically support those proposed TM estimation methods that use power law MVR as their basic statistical assumption. We verify our conclusion by public DECPKT traces and actual network traffic in laboratory network and campus network. And based on actual observation for more than one year, we verify that the power law MVR is generally suitable for the traffic at proper volume or aggregated by a large amount of sources, while our relationship can describe more cases of actual network traffic very well than empirical MVR.

In the further work, we still study several problems. (i) We plan to further study more traffic samples generated by different typical applications to validate our MVR. (ii) The existence of MVR is now widely accepted; thus it should be considered as an important character of network traffic. All proposed traffic models and traffic generation algorithms should be checked whether they can derive proper MVR expression. (iii) New TM estimation methods should be proposed based on our theoretical MVR.

1 Zhou J J, Yang J H, Yang Y, et al. Research on traffic matrix estimation (in Chinese). J Softw, 2007, 18(11): 224–231

2 Medina A, Taft N, Salamatian K, et al. Traffic matrix estimation: existing techniques and new directions. ACM SIGCOMM Comp Com R, 2002, 32: 161–174

3 Cao J, Davis D, Wiel S, et al. Time-varying network tomography: Router link data. J Am Stat A, 2000, 95: 1063–1075

4 Vardi Y. Network tomography: estimating source-destination traffic intensities from link data. J Am Stat A, 1996, 91: 365–377

5 Willinger W, Taqqu M S, Sherman R, et al. Self-similarity through high-variability: Statistical anaysis of ethernet LAN traffic at the source level. ACM SIGCOMM Comp Com R, 1995, 25: 100–113

6 Willinger W, Taqqu M S, Sherman R, et al. Self-similarity through high-variability: Statistical anaysis of ethernet LAN traffic at the source level. IEEE ACM TN, 1997, 5: 71–86

7 LBL, Internet traffc archive, http://ita.ee.lbl.gov/html/contrib/DEC-PKT.html

8 Paxson V, Floyd S. Wide-area traffic: The failure of poisson modeling. IEEE ACM TN, 1995, 3: 226–244

9 Gunnar A, Johansson M, Telkamp T. Traffic matrix estimation on a large ip backbone-a comparison on real data. In: ACM IMC, Taormina, Sicily, Italy, 2004. 149–160

10 Juva I, Vaton S, Virtamo J. Quick traffic matrix estimation based on link count covariances. In: IEEE ICC. Istanbul, 2006. 603–608

11 Leland W E, Taqqu M S, Willinger W, et al. On the self-similar nature of ethernet traffic (extended version). IEEE ACM TN, 1994, 2: 1–15

12 Dainotti A, Pescapé A, Ventre G. A packet-level characterization of network traffic. In: CAMAD, Trento, Italy, 2006. 38–45

13 Sigman K. Appendix: Primer on Heavy-tailed distributions. Queueing Syst, 1999. 261–275

14 Kratz M F, Resnick S I. The qq-estimator and heavy tails. Comm Statist St M, 1996, 12: 699–724

15 Abrahamsson H, Ahlgren B. Using empirical distributions to characterize Web client traffic and to generate synthetic traffic. In: GLOBECOM, San Francisco, CA, USA, 2000. 428–433

16 Giordano S, Miduri S, Pagano M, et al. A wavelet-based approach to the estimation of the hurst parameter for self-similar data. In: DSP, Santorini, Greece, 1997. 479–482

17 Riedi R H, Crouse M S, Ribeiro V J, et al. A multifractal wavelet model with application to network traffic. IEEE Info T, 1999, 45: 992–1018