Journal of Computational Information Systems 7:5 (2011) 1488-1499
Available at http://www.Jofcis.com

On the Conflict of Truncated Random Variable vs. Heavy-Tail and Long Range Dependence in Computer and Network Simulation

Wanyang DAI†

Department of Mathematics and State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing 210093, China

Abstract

Heavy-tailed distribution (HTD) and long-range dependence (LRD) have appeared in the studies of telecommunication, finance, and etc. However, the truncation of a heavy-tailed random variable is inevitable in a computer and network simulation due to the hardware and software limitations such as numbering format of binary digits and pseudo random number generators. Thus a natural issue arises: how well the approximation of a concerned objective in a simulation will be if a truncated random variable is used to replace the original heavy-tailed one? As pointed out in this paper, the difference of the means (or variances) between the truncated version and the original one will still be infinite if the original random variable is heavily tailed with infinite mean and/or variance, which indicates that the truncated approximation may be not consistent with what should be if the HTD/LRD assumptions are imposed. So, based on this observation and our previously established approximation theory via reflecting Gaussian process (RGP) with or without LRD, we can provide more reasonable interpretations to some well-known large-scale computer and statistical experiments in characterizing the superposition of internet traffic sources, which clarify that the findings in these simulations should be more close to the reality but not to the mathematical assumptions about HTD/LRD, which are imposed in their models and analysis. Hence our findings smooth, to some extent, the gaps between HTD/LRD theory and real-world simulations. More importantly, our findings also indicate that it is more suitable to use the truncated version of a heavy-tailed random variable in a simulation and related analysis in order to be more consistent with reality and to avoid more complicated discussions via HTD/LRD theory.

Keywords: High-speed Network; Source Traffic; Characterization; Queueing System; Computer and Network Simulation; Pseudo Random Number Generator; Numbering Format of Binary Digits; Truncated Random Variable; Heavy-tailed Distribution; Long-range Dependence; Reflecting Gaussian Process

1. Introduction

HTD/LRD have been used in many areas to characterize source traffics in telecommunication (see, e.g., Refs. [5~6], [11], [13], [19], and etc.), price jumps and correlations in finance (see, e.g., Refs. [1], [2]), and etc. Particularly, in the early 1990s, the authors such as Refs. [11] and [19] disclosed one of the most interesting findings in high-speed networks, i.e., source traffic data were shown to possess LRD. Since then, stochastic modeling associated with LRD data has become an active area of research. For examples, in the early 2000s, a number of scientists in Bell Labs focused on traffic data and related queueing modeling based on statistical multiplexing. Their achievements (see, e.g., Refs. [5~6], and etc.) through large-scale computer and statistical experiments indicate that the packet interarrival times and packet sizes tend to i.i.d.

† Corresponding author.

Email addresses: nan5lu8@netra.nju.edu.cn (Wanyang DAI)

(independent and identically distributed) when the number of the input connections increases. Their findings are of great importance in designing and implementing telecommunication software systems and hardware devices since HTD/LRD sources will significantly increase system design complexity and the cost of high-speed memory. Moreover, for some particular cases and under certain time scaling, Refs. [6] developed some theory to make their simulation findings justified.

However, since the above well-known simulations are computer based ones, the truncation of a heavy-tailed random variable used in the simulations is inevitable due to the hardware and software limitations of a computer, such as, the numbering format of binary digits and the pseudo uniform random number generators (see, e.g., Ref. [14]). Then a natural issue arises concerning the truncation: how well the approximation of a concerned objective designed in a simulation will be if a truncated random variable is used to replace the original heavy-tailed one? As pointed out in this paper, the difference of the means (or variances) between the truncated version and the original one will still be infinite if the original random variable is heavily tailed with infinite mean and/or variance, which indicates that the truncated approximation may be not consistent with the HTD/LRD based theory. In fact, some of the simulations conducted by authors in Refs. [5~6], and etc. have revealed some gaps between the simulation findings and the related HTD/LRD theory.

For example, the theory developed in Ref. [6] to justify their simulation findings is under the so-called critical time scale introduced in Ref. [15], i.e., the characteristics of an individual source becomes less important when the number of superposed sources increases. Nevertheless, in general, the characteristics of the superposed limiting process heavily depends on the ways of the time/space scales (see, e.g., Refs. [6], [10], [12]). If the time scale of interest does not change with superposition, things will be different since individual source behavior still plays an important role.

So, by the above observation and our previously established approximation theory in Ref. [10] concerning RGP with or without LRD, we can provide more reasonable interpretations to the simulation findings in Refs. [5~6], and etc., which are based on the same queueing model and the same time/space scaling with both heavy-tailed random variable and its truncated counterpart as the driving random environmental factors. From these interpretations, we can see that the simulation findings in Refs. [5~6], and etc. should be more close to the reality but not to the mathematical assumptions about HTD/LRD, which are imposed in their models and stemmed from Ref. [19] where only bounded range of time scales are concerned. Hence our findings smooth, to some extent, the gaps between HTD/LRD theory and real-world simulations. More importantly, our findings also indicate that it is more suitable to use the truncated version to replace a heavy-tailed random variable in a simulation and related analysis in order to be more consistent with reality and to avoid more complicated discussions via HTD/LRD theory.

The rest of this paper is organized as follows. In Section 2, we make a comparison between a heavy-tailed random variable and its associated truncated version, and state the reason why the hardware and software systems in a computer can cause truncation to a heavy-tailed random variable. Moreover, in the section, we also describe the possible impact of the truncation of a random variable on some well-known simulation findings in high-speed networks. In Section 3, we use our previously established approximation theory via RGP to justify that some well-known simulation findings should be more close to the reality but not to the imposed mathematical assumptions about HTD/LRD. Finally, in Section 4, we present the conclusion of the paper and make some remarks about the impact of a truncated heavy-tailed random variable on financial engineering, insurance and time series analysis.

2. Truncated Random Variable vs. HTD/LRD in a Computer Simulation

2.1. Truncation Against HTD/LRD Related Simulation

Consider a random variable X with distribution $F(x)=P\{X \le x\}$ over $x\in[p, q]$, where $p,q\in(-\infty, +\infty)$ with $p<q$, and moreover, suppose $[p, q]=[p, +\infty)$ if $q=+\infty$ and $[p, q]=(-\infty, q]$ if $p=-\infty$. Then we can define the truncated random variable Y over $[p, q]$ as follows,

$$Y=\begin{cases} X & if\ X\ takes\ values\ in[p,q], \\ ignored & if\ X\ does\ not\ take\ values\ in[p,q]. \end{cases} \tag{1}$$

In other words, the distribution of Y can be represented in a conditional probability form, i.e., for any $y\in[p,q]$,

$$P\{Y\le y\}=P\{X\le y \mid X\in[p,q]\}=\frac{P\{X\le y, X\in[p,q]\}}{P\{X\in[p,q]\}}. \tag{2}$$

Thus we know that the truncated mean corresponding to X is given as follows,

$$E\left[XI_{\{X\in[p,q]\}}\right]=E\left[E\left[XI_{\{X\in[p,q]\}}\middle|X\in[p,q]=E[Y].\right]\right]=E[Y]. \tag{3}$$

Similarly, we know that the truncated variance corresponding to X is given by

$$Var\left(XI_{\{X\in[p,q]\}}\right)=Var(Y). \tag{4}$$

Therefore, when the random variable X is lightly tailed, we know that the following claims are true as $p\to-\infty$ and $q\to+\infty$,

$$\begin{aligned} &P\{X\in(-\infty,p)\cup(q,+\infty)\}\to 0, \\ &E[X]-E[Y]\to 0, \\ &Var(X)-Var(Y)\to 0, \end{aligned}$$

which imply that Y can be used as an approximating random variable for X in a simulation.

However, when the random variable X is heavily tailed, the mean and/or variance of X may be infinite. For example, if X is a Pareto distributed random variable, then the distribution of X has the following expression with parameters $\beta\ge 0$ and $x_m\ge 0$,

$$F(x)=\begin{cases} 1-\left(\dfrac{x_m}{x}\right)^{\beta} & for\ x\ge x_m, \\ 0 & for\ x<x_m, \end{cases} \tag{5}$$

which implies that both the mean and variance of *X* will be infinite if $\beta \le 1$ and that the mean is finite and the variance is infinite if $1 < \beta \le 2$. Moreover, the distribution for the corresponding random variable *Y* that is truncated by a given constant $b \ge x_m$ can be expressed as follows,

$$F^t(x) = \begin{cases} 1 & \text{for} \quad x > q; \\ \dfrac{1-\left(\dfrac{x_m}{x}\right)^\beta}{1-\left(\dfrac{x_m}{q}\right)^\beta} & \text{for} \quad x_m \le x \le q, \\ 0 & \text{for} \quad x < x_m. \end{cases} \tag{6}$$

Therefore it follows from (2.5)-(2.6) that we may have the following claims for any given $q > x_m$ when *X* is Pareto distributed,

$$E[X]-E[Y]=\infty \text{ and/or } Var(X)-Var(Y)=\infty, \tag{7}$$

i.e., the differences of the mean and/or variance between the heavy-tailed random variable and its truncated counterpart may be infinite. Therefore, if $E[X]=\infty$ and/or $Var(X)= \infty$, it is not suitable to use the truncated random variable *Y* as an approximating one for *X* in a simulation no matter how large *E*[*Y*] and/or *Var*(*Y*) are. In addition, we have the observation that the truncated random variable may take a large (maybe huge but bounded) value with relatively large probability but itself is essentially not a heavy-tailed random variable. So, to be concise in applications and to avoid more complicated discussions through applying HTD/LRD theory when a random variable appeared in a real-world application (e.g., in finance) is obviously bounded by some large (maybe huge) constant (e.g., the total wealth in the world or in a local environment), a good choice is to use the truncated version of a heavy-tailed random variable in a simulation.

2.2. The Reason Causing Truncation in a Computer Simulation

In a today's computer, numbering format is based on a fixed number of binary digits, which is denoted by *L*, e.g., *L*=32. So the integer $M=2^L-1$ is the largest integer accepted by the computer. Due to this fact, some random numbers generated from pseudo random number generators employed in a simulation will be truncated if the associated random variable is heavily tailed. For example, consider the Pareto distributed random variable *X* and a uniformly distributed random variable *U* over [0, 1], it follows from the discussion in page 32 of Ref. [18] or page 36 of Ref. [14] that

$$X = F^{\leftarrow}(U) \tag{8}$$

where $F^{\leftarrow}$ denotes the left-continuous inverse of *F* and has the following expression,

$$F^{\leftarrow}(u) = \frac{x_m}{(1-u)^{1/\beta}}, u \in (0,1). \tag{9}$$

Therefore, once we generate the uniformly distributed random numbers from *U* over (0, 1), we get the Pareto distributed random numbers from *X* over $[0, \infty)$. Currently, there exist some pseudo uniform random number generators such as the Kiss generator whose period is of order 2^{95} (see, e.g., pages 39-43 in Ref. [14]). Nevertheless, all the generated random numbers $F^{\leftarrow}(u)$ Corresponding to u near the unity will be truncated in a simulation if $F^{\leftarrow}(u) > M$.

To be clear, we also remark that there exists a class of uniformly distributed random number generators with the form of $X_{n+1} = f(X_n)$ such as linear congruence generator, whose period is bounded by M(see, e.g., pages 40-43 in Ref. [14]), then the largest pseudo uniform random number except the unity, which generated from these generators, are bounded by a constant $c = (2^L - 1)/2^L$. Therefore, if $1 < \beta \leq 2$ in (9), we can see that the $F^{\leftarrow}(u)$ corresponding to the generated random number u is always less than *M* for relatively small x_m, which indicates that these generators themselves also have the truncation ability to a heavy-tailed random variable since all the values of $(F^{\leftarrow}(c), \infty)$ to the Pareto distributed random variable are truncated.

2.3. Impacts on Some Well-known Simulations Related to HTD/LRD

The above observations indicate that many existing simulations and related analysis based on HTD/LRD appeared in communication systems, financial engineering, and etc. (see, e.g., Refs. [5~6], [19]) may need some new interpretations. In fact, some of these simulations are well-known and have shown some gaps between the simulation findings and the related HTD/LRD theory.

Concretely, in the early 1990s, one of the most interesting findings in high-speed networks is the one that traffic data were shown to possess LRD (see, e.g., Refs. [11], [13], [19]). In particular, The authors in Refs. [19] showed that ON/OFF processes with heavy-tailed ON-period distributions exhibit LRD, which is based on detailed statistical analysis of real-time traffic measurements on large time scales but bounded by $M=2^L-1$ in seconds from Ethernet LAN's at the level of individual sources. Since then, stochastic modeling associated with LRD data has become an active area of research. For examples, in the early 2000s, a number of scientists in Bell Labs focused on traffic data and related queueing modeling based on statistical multiplexing. Their achievements (see, Refs. [5~6], and etc.) through large-scale computer and statistical experiments indicate that the packet interarrival times and packet sizes tend to i.i.d. when the number of the input connections increases. Their findings are of great importance in designing and implementing telecommunication software systems and hardware devices since HTD/LRD sources will significantly increase system design complexity and the cost of high-speed memory. Moreover, for some particular cases and under certain time scaling, the authors of Ref. [6] developed some theory to make their simulation findings justified.

For example, a particular queueing model used in their theory is based on the following assumptions: the distributions of ON- and OFF-periods are Pareto and exponential respectively and packet sizes are constant. They showed that the sequence of probabilities that steady state unfinished works exceed a threshold tend to the corresponding probability assuming Poisson input process when the number of input sources tends to infinity, which coincides with their simulation studies.

Nevertheless, as pointed out in Ref. [6], their theory heavily depends on the so-called critical time scale introduced in Ref. [15], under which the characteristics of an individual source become less important. If the time scale of interest does not change with superposition, then things will be different since individual source behavior still plays an important role. Interested readers are also directed to Ref. [12] for more details about the influence on LRD caused by different time scales.

Now, since the simulations conducted in Refs. [5~6] and etc. are computer based ones, the truncation of a heavy-tailed random variable as discussed in Section 2 is inevitable, we can provide some different interpretations about the findings in their simulations. Further, our interpretations seem to be more consistent between their simulation findings and our approximation theory developed in Ref. [10]. For example, under the same queueing model and the same time/space scaling, if the ON-period random variables with Pareto distributions are truncated, then the queueing length process corresponding to large superposition number of input sources is roughly equal to the one with input traffic whose interarrival times are i.i.d., which coincides with their simulation findings in Refs. [5~6], and etc. Nevertheless, if the ON-period random variables are not truncated, then the queueing length process corresponding to large superposition number of input sources is roughly equal to the one whose input traffic is of LRD. From this example, we can see that whether or not a heavy-tailed random variable is truncated will have a significant impact on a simulation. In the following section, we will use a generalized queueing model to concretely justify our findings based on the theory developed in Refs. [10] and the observation found in the current paper concerning the truncation of a heavy-tailed random variable in a simulation.

3. Demonstration Through a Generalized Queueing System

In the study of Ref. [10], the author established some approximation theory to justify several heavy traffic limit theorems under different time/space scaling for a queue with superposed LRD ON/OFF input sources and general service time distribution. In this section, we employ one of the limit theorems, which is under a time/space scaling similar to the diffusive scaling as used in the conventional central limit theorem, to demonstrate our findings as stated at the end of Section 2.3. The rest of this section is organized as follows. In Subsections 3.1-3.2, we summarize the queueing model and refine the limit theorem required for the current purpose, and moreover, and in Subsection 3.3, we present our interpretations and their implication.

3.1. The Queueing Model

In the queueing system under consideration, the service times are assumed to be generally distributed and there are N i.i.d. ON/OFF input sources. Each ON/OFF source $n \in \{1,..,N\}$ consists of independent strictly alternating ON- and OFF-periods, moreover, it transmits packets to a server according to a Poisson process with interarrival time sequence $\{u_n(i), i \geq 1\}$ and rate λ if it is ON and remains silent if it is OFF. The lengths of the ON-periods are identically distributed and so are the lengths of OFF-periods, and furthermore, both of their distributions can be heavy-tailed with infinite variance. Specifically, for any distribution *F*, we denote by $\overline{F} = 1 - F$ the complementary (or right tail) distribution, and by F_1 and F_2 the distributions for ON-and OFF-periods with probability density functions f_1 and f_2 respectively. Their means and variances are denoted by μ_i and σ_i^2 for $i = 1, 2$. In what follows, we assume that as $x \to \infty$,

$$\text{either } \overline{F}_i(x) \sim x^{-\alpha_i} L_i(x) \text{ with } 1 < \alpha_i < 2 \text{ or } \sigma_i^2 < \infty, \tag{10}$$

where ~ denotes "nearly equals" and $L_i > 0$ is a slowly varying function at infinity, that is,

$$\lim_{x \to \infty} \frac{L_i(tx)}{L_i(x)} = 1 \text{ for any } t > 0.$$

Note that the mean μ_i is always finite but the variance σ_i^2 is infinite when $\alpha_i < 2$, and furthermore, one distribution may have finite variance and the other has an infinite variance since F_1 and F_2 are allowed to be different. The sizes of transmitted packets (service times) form an i.i.d. random sequence $\{v^N(i) = v(i)/\mu^N, i \geq 1\}$, where μ^N is the rate of transmission corresponding to each *N* and $\{v(i) : i \geq 1\}$ is an i.i.d. random sequence with mean 1 and variance σ_v^2, moreover, $\{v(i) : i \geq 1\}$ is independent of the arrival processes.

To derive our queueing dynamical equation, we introduce more notations. For a single source $n \in \{1, \cdots, N\}$, it follows from the explanation in Ref. [12] that the alternating ON/OFF periods can be described by a stationary binary process $W_n = \{W_n(t), t \geq 0\}$, where

$$W_n(t) = \begin{cases} 1 & \text{if input traffic is in an ON}-\text{period at time t,} \\ 0 & \text{if input traffic is in an OFF}-\text{period at time t.} \end{cases} \tag{11}$$

Moreover, the mean of W_n is given by

$$\gamma = E[W_n(t)] = P\{W_n(t) = 1\} = \frac{\mu_1}{\mu_1 + \mu_2} \tag{12}$$

In addition, we let $A^N(t)$ be the total number of packets transmitted to the server by time *t* summed over all *N* sources, $S^N(t)$ be the total number of packets that finished service at the server if it keeps busy by time *t*. Then the queue length process $Q^N(t)$ (the number of packets including the one being served at the server at time *t*) can be represented by

$$Q^N(t) = A^N(t) = S^N(B^N(t)) \tag{13}$$

where we assume that the initial queue length is zero for convenience, $B^N(t)$ is the cumulative amount of time that the server is busy by time *t*. In the following analysis, we will employ the first-in-first-out (FIFO) and *non-idling* service discipline under which the server is never idle when there are packets waiting to be served.

3.2. The Required Limit Theorem

For the above system, we are interested in the behavior of the scaled queueing process $\tilde{Q}^N(\cdot)$ under the condition that the load of the server closely approaches the service capacity when the source number *N* gets large enough, where

$$\tilde{Q}^N(\cdot) \equiv \frac{1}{\sqrt{N}} Q^N(\cdot). \tag{14}$$

Moreover, for convenience, we adapt some notations from Ref. [17]. When $1 < \alpha_i < 2$, set $a_i = (\Gamma(2-\alpha_i))/(\alpha_i - 1)$. When $\sigma_i^2 < \infty$, set $\alpha_i = 2$, $L_i \equiv 1$ and $a_i = \sigma_i^2 / 2$. In addition, define

$$b = \lim_{x\to\infty} t^{\alpha_2-\alpha_1} \frac{L_1(x)}{L_2(x)}. \tag{15}$$

If $0 < b < \infty$ (implying $\alpha_1 = \alpha_2$ and $b = \lim_{x\to\infty} L_1(x)/L_2(x)$), set $\alpha_{\min} = \alpha_1$,

$$\pi^2 = \frac{2(\mu_2^2 a_1 b + \mu_1^2 a_2)}{(\mu_1+\mu_2)^3 \Gamma(4-\alpha_{\min})} \, and \;\; L = L_2; \tag{16}$$

if, on the other hand , $b = 0$ or $b = \infty$,

$$\pi^2 = \frac{2\mu_{\max}^2 a_{\min}}{(\mu_1+\mu_2)^3 \Gamma(4-\alpha_{\min})} \, and \;\; L = L_{\min} \tag{17}$$

where min is the index 1 if $b = \infty$ (e.g. if $\alpha_1 < \alpha_2$) and is the index 2 if $b = 0$, max denoting the other index . Moreover, for each *N*, let the service rate μ^N corresponding to some positive constant θ be given by

$$\mu^N = N\lambda\gamma + \theta\sqrt{N} \ . \tag{18}$$

In addition, suppose that the distributions of F_1 and F_2 satisfy

$$F_i(x) \ (i = 1, 2) \text{ is absolutely continuous in terms of } x; \tag{19}$$

$$\text{The density } f_i(x) \ (i = 1, 2) \text{ of } F_i \text{ satisfies } \lim_{x \to 0^+} f_i(x) < \infty \ . \tag{20}$$

Theorem 3.1. *Under conditions (18)-(20) and as* $N \to \infty$, $\tilde{Q}^N(\cdot)$ *converges in distribution under Skorohod topology to a reflecting Gaussian process* $\tilde{Q}\ (\cdot)$ *given by*

$$\tilde{Q}(\cdot) = \tilde{A}(\gamma\cdot) + \lambda\tilde{T}(\cdot) - \tilde{S}(\lambda\gamma\cdot) - \theta\cdot + \tilde{I}(\cdot) \geq 0 \tag{21}$$

where the three processes $\tilde{A}(\gamma\cdot)$, $\tilde{S}(\lambda\gamma\cdot)$ *and* $\tilde{T}(\cdot)$ *are independent each other, and furthermore,* $\tilde{A}(\gamma\cdot)$ *is a Brownian motion with mean zero and variance function* $\lambda\gamma\cdot$, $\tilde{S}(\lambda\gamma\cdot)$ *is also a Brownian motion with mean zero and variance function* $\lambda\gamma\sigma_v^2\cdot$, $\tilde{T}(\cdot)$ *is a Gaussian process with a.s. continuous sample paths , mean zero and stationary increments , whose covariance function*

$$Var(\tilde{T}(t)) \textit{ can be expressed as } \begin{cases} \sim \pi^2 t^{2H} L(t) \ \ as\ t \to \infty \quad for\ \ H \neq \dfrac{1}{2}, \\ = \pi^2 t \qquad\qquad for\ \ all\ t \geq 0\ and\ H = \dfrac{1}{2} \end{cases} \tag{22}$$

where H is the Hurst parameter defined by H= $(3 - \alpha_{\min})/2$. *Moreover,* $\tilde{I}(\cdot)$ *is a non-decreasing process with* $\tilde{I}(0) = 0$ *and satisfies*

$$\int_0^\infty \tilde{Q}(s) d\tilde{I}(s) = 0 \ . \tag{23}$$

3.3. The Interpretations and Their Implication

In this subsection, we suppose that the ON-periods in the queueing system are of Pareto distribution with $1 < \beta \le 2$ and thus we know that the ON-period is heavily tailed with mean $\mu_1 < \infty$, variance $\sigma_1^2 = \infty$, $\alpha_1 = \beta$, $L_1(x) = x_m^\beta$. Moreover, we suppose that the OFF periods are of exponential distribution with mean $\mu_2 < \infty$, variance $\sigma_2^2 = \mu_2^2 < \infty$, $\alpha_2 = 1/\mu_2, L_2(x) = 1$ and suppose that $\alpha_2 > \alpha_1$. Then it follows from (15) that $b = \infty$ and then we know that $\alpha_{\min} = \alpha_1 = \beta$, which implies that *H* defined in Theorem 3.1 takes value over (1/2, 1). Hence it follows from Theorem 3.1 that the corresponding limiting queue length process $\tilde{Q}$ is a reflecting Gaussian process with LRD. Therefore, if we consider $\tilde{Q}$ as an approximating physical queue length process, then the corresponding input process superposed from large number of ON/OFF sources should be of LRD.

However, if we consider the ON-periods are of the corresponding truncated Pareto distribution, the associated mean and variance are both finite. Thus, similar to the above discussion, we know that *H=1/2,* which means by Theorem 3.1 that the derived limiting queue length process $\tilde{Q}$ is a reflecting Brownian motion (RBM), which does not have the LRD (interested readers are also directed to Refs. [8~9] for more details about Brownian approximations). Nevertheless, the constant π^2 in (21) may be very large. Therefore, if we consider $\tilde{Q}$ as an approximating queue length process, then the corresponding input process superposed from large number of ON/OFF sources should not be of LRD and the packet interarrival times more look like being generated from an i.i.d sequence of random variables, which coincides with what were observed in the simulations conducted by Refs. [4~6], and etc.

Finally, from the practical viewpoint, it should be reasonable that the ON-periods and OFF-periods from lower speed end users are bounded by some large (maybe huge) constant (e.g., even the range of time scales used in the implementations of Ref. [19] are bounded by $M = 2^{L-1}$). Therefore, the findings based on their simulations in Refs. [5~6] and etc. should be more close to the reality but not to their mathematical assumptions imposed about HTD/LRD due to the truncation effect in their simulations. More importantly, their assumptions on HTD/LRD are stemmed from Ref. [19] where only bounded time scales are concerned. So it is more suitable to use the truncated version of a heavy-tailed random variable in a simulation to be consistent with practice and to avoid more complicated HTD/LRD related analysis.

4. Conclusions and Future Work

In this paper, we have answered the question arising in a simulation or a real-word application, which indicates that the approximation of a concerned objective through a truncated random variable to replace the original heavy-tailed one may be not consistent with what should be if the HTD/LRD assumptions are imposed. So, based on this observation and our previously established approximating theory about RGP with or without LRD, we provide more reasonable interpretations to some well-known large-scale computer and statistical experiments in characterizing the superposition of internet traffic sources. More importantly, our findings also indicate that, in a simulation and related analysis, it is more suitable to use a truncated random variable to replace a heavy-tailed one in order to be more consistent with reality and to avoid more complicated discussions via HTD/LRD theory.

In addition to telecommunication, the studies concerning HTD/LRD have appeared in many fields such as financial engineering, insurance, and time series analysis (see, e.g., Refs. [1~3]). For the purpose of computer oriented simulations and real-world applications, it is also inevitable to handle the truncation problem when a heavy-tailed random variable is concerned.

Acknowledgement

This project is supported by Natural Science Foundation of China with Grant No. 10971249.

References

[1] F. Avram, Z. Palmowski, and M. Pistorius. On the optimal dividend problem for a spectrally negative Levy process. Annals of Applied Probability, 2007, 17(1): 156~180.

[2] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics. J. R. Stat. Soc. Ser. B, 2001, 63: 167~241.

[3] P. J. Brockwell and R. A. Davis. Time Series: Theory and Methods, Springer-Verlag, New York, 1991.

[4] J. Cao, W. S. Cleverland, D. Lin, and D. X. Sun. Internet traffic tends to Poisson and independent as the load increases. Nonlinear Estimation and Classification, eds. C. Holmes, D. Dension, M. Hansen, B. Yu, and B. Mallick, Springer, New York, 2002.

[5] J. Cao, W. S. Cleverland, D. Lin, and D. X. Sun. The effect of statistical multiplexing on the long-range dependence of Internet packet traffic. Tech Report, Bell Labs, Murray Hill, NJ, U.S.A., 2002.

[6] J. Cao and K. Ramanan. A Poisson limit for the unfinished work superposed point processes. Tech Report. Bell Labs, Murray Hill, NJ, U.S.A., 2001..

[7] J. D. Cavanaugh and T.J. Salo. Internetwotking with ATM WANS. In William Stallings, editor, Advances in Local and Metropolitan Area Networks, IEEE Computer Society Press, 1994.

[8] J. G. Dai and W. Dai. A heavy traffic limit theorem for a class of open queueing networks with finite buffers. Queueing Systems, 1999, 32: 5~40.

[9] W. Dai. Brownian Approximations for Queueing Networks with Finite Buffers: Modeling, Heavy Traffic Analysis and Numerical Implementations. Ph.D Thesis, The Georgia Institute of Technology, 1996. Also published in UMI Dissertation Services, A Bell & Howell Company, Michigan, U.S.A., 1997.

[10] W. Dai. Heavy traffic limit theorems for a queue with Poisson ON/OFF long-range depedent sources and general service time distribution. Preprint, 2004, invited talk at 2004 Annual Conference of Jiangsu Probability and Statistical Society, also presented at 2005 International Conference of Management and Applications, Chengdu, China and 2006 INFORMS Hong Kong International Conference, Hong Kong, China.

[11] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. IEEE/ACM Trans. Netw., 1994, 2: 1~15.

[12] T. Mikosch, S. Resnick, H. Rootzen, and A. Stegeman. Is network traffic approximated by stable Levy motion or fractional Brownian motion? Annals of Applied Probability, 2002, 12(1): 23~68.

[13] V. Paxson and S. Floyd. Wide-area traffic: the failure of Poisson modeling. IEEE/ACM Transactions on Networking, 1995, 3: 226~244.

[14] C. P. Robert and G. Casella, Monte Carlo Statistical Methods. Springer, New York, 1999.

[15] B. K. Ryu and A. Elwalid. The importance of long-range dependence of VBR traffic in ATM traffic engineering: myths and realities. In Proc. ACM SIGCOMM' 96, 1996, 3~4.

[16] B.K. Ryu and S.B. Lowen. Point process approaches to the modeling and analysis of self-similar traffic: Part I-Model construction. In Proc. IEEE Infocom 1996, 1996, 1468~1475.

[17] M. S. Taqqu, W. Willinger, and R. Sherman. Proof of a fundamental result in self-similar traffic modeling. ACM/Sigcomm Computer Communication Review, 1997, 27: 5~23.

[18] W. Whitt. Stochastic Process Limits. New York, Springer-Verlag, 2002.

[19] 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, 1997, 5(1): 71~86.