行政院國家科學委員會專題研究計畫成果報告

真實網路封包流量自我類似性質之統計模式建立

On Generator of Network Arrivals with Self-Similar Nature

計畫編號：NSC90-2213-E009-056
執行期限：90年8月1日至91年7月31日

主持人：陳伯寧　國立交通大學電信工程系所 (poning@cc.nctu.edu.tw)
共同主持人：張仲儒、陳耀宗、楊啟瑞 國立交通大學電信工程系所/資訊工程系所
計畫參與人員：姚建　國立交通大學電子工程系所
花凱龍、林建光、王聖賢　國立交通大學電信工程系所

一、中文摘要

近年來，許多網路量測研究顯示，真實網路封包序列具有自我類化特性。因此，傳統使用的短時程相依性程序，例如波以松程序等，並不適合作為網路封包序列的統計模式，唯有具有長時程相依性的自我類化程序才是一個適當的統計模式。所以，當評估一個網路裝置的性能時，如果使用短時程相依性程序而忽略長時程相依性的重要性時，所得到的評估效能將會和此裝置在真實網路上使用所得到的量測效能產生相當大的差異。

在本計畫第一年，我們已成功的定義並分析馬可夫模式的類化性質。在今年度的計畫，我們提出一個根據濾波器原理所設計的類化封包訊務產生器，並且從理論上分析其類化性質。我們注意到網路量測封包訊務的類化性質，雖然延續至超越目前工程技術可掌控的範圍，但終究會收斂為非類化訊務。因此合成並使用一個類化性質延續至極限範圍的訊務來作為模擬依據是否為必要，是一個值得商確的問題。基於此，另一種出發點為合成一個類化性質延續至所需要的範圍，而極限時則呈現非類化性質的訊務似乎更合乎實際量測訊務的行為。我們所提出的方法不僅程式相當簡單且符合以上的觀察需求，更要的是我們的模式也不會有另外兩個知名類化訊務產生器(傅立業轉換模式、隨機中點取代模式)的兩個問題: (1)所產生的封包序列可能為負整數；(2)需事先設定所需產生的封包量，換句話說，當所需要製造的封包序列長度改變時，整個封包序列必需要重新產生。

關鍵詞：統計模式、自我類似特性、長時程相關性

Abstract

Recent empirical studies have shown that the modern computer network traffic is much more appropriately modeled by long-range dependent self-similar processes than traditional short-range dependent processes such as Poisson. Hence, if long-range dependence is not considered for synthesizing experimental network traffic, it will lead to incorrect assessments of performance evaluation in network system. This arises the need of a well synthesizing trace with long-range dependence.

In the first year of this 3-year project, we defined and subsequently analyzed the degree of self-similarity for Markov packet sources. In this second-year project term, we developed a filter-based self-similar traffic generator, and theoretically analyze its self-similar property. Notably, the true measured network traffic, although appearing self-similar beyond the range of engineering manageability is still ultimately non-self-similar. Therefore, it may be arguable to synthesize and use an ultimate self-similar traffic for system performance evaluation. An alternative that generates a traffic that has the desired degree of self-similarity in a controlled range, and becomes non-self-similar beyond may be closer to the true traffic behavior. As expected, our generator can fulfill the above need. Most importantly, our generator eliminates two of the problems of two other well-known self-similar traffic generators (Paxson-FFT and Random-Mid-Point): (1)

the synthesized traffic may be negative; (2) the length of the synthesized traffic must be pre-specified. In other words, for different length of the synthesized traffic, the entire traffic must be re-generated.

Keywords: Network Traffic Model,
Self-Similarity, Long-Range
Dependence

二、計畫緣由與目的

Whether a communication system is well operated resides on its reliability in communication quality from the user point of view. To illustrate, the current wired telephone system has been held in high esteem because it provides users reliable circuit-switch-based connections. In order to ensure the reliability of a system, a certain number of testing is a must-do before its deployment. These tests must be properly conducted so that the system performance after deployment can be predictable. This arises the need of a synthesizing experimental traffic trace that well approximates the true traffic, possibly encountered in practice. As an example, the well-known Erlang B and Erlang C formulas, derived from the Markovian models, successfully characterize the user behaviors by accurately predicting the overall call blocking and queuing probability. We therefore realize the significance of a traffic model for system testing.

In early days, Poisson processes were commonly used as traffic models for packet network system. This was done under the premise that the traffic behavior in network system is similar to that in circuit-switch telephony system. Although the traffic behaviors of these two systems are both due to human behavior, the situation for the network system is more complicated because of its packet-switch nature. Other factors [8], such as network protocols, even further complicate the resultant traffic characteristic.

In [1, 2, 3, 5, 7], measurement studies have shown that the actual network traffics for different networks (e.g. Ethernet LAN, WAN, CCSN/SS7, ISDN, and VBR Video) are clearly distinguishable from the synthesized traffics by traditional Poisson or related models. Specifically, Leland and Wilson, who recorded hundreds of millions of Ethernet packets with recorded time-stamp accurate to within 100 μs, compared the measured traffic data on Ethernet LAN at Bellcore with the Markovian modeled sequences for the same load [3]. They found that in contrast to traditional models, measured traffic vary over a wide range of time scales and the predicted performance with traditional models as the input stream is quite different from the performance with measured data as the input stream. Therefore, for performance assessments and predictions of these network systems, a good model that emulates the long-range dependence of the measured data becomes necessary.

Several approaches have been proposed for synthesizing long-range dependent self-similar traffic data. In [6], Paxson synthesized self-similar traffic data by means of traffic spectrum fitting to fractional Gaussian noise. Lau, et al, [4] proposed a so-called random midpoint displacement algorithm to generate a self-similar network trace. We then noted two drawbacks of adopting these approaches. First, the required length of a traffic data should be determined prior to the generation of the traffic data; hence, when a longer traffic sequence is required, one needs to go through the entire process of data synthesization to obtain it. In other words, the traffic data cannot be generated in an on-the-fly fashion. In addition, their traffic generators may produce negative integers, unreasonable for any packet train arrival. Most importantly, the true measured network traffic, although appearing self-similar beyond the range of engineering manageability is still ultimately non-self-similar. Therefore, it may be arguable to synthesize and use an ultimate self-similar traffic for system performance evaluation. An alternative that generates a traffic that has the desired degree of self-similarity in a controlled range, and becomes non-self-similar beyond may be closer to the true traffic behavior. This leads us to develop a new approach that can compensate these drawbacks.

三、自我類似封包序列產生器

The key idea of our generator is based on power spectrum fitting. Let $S_y(w)$ denote the power spectrum of the discrete random process $Y[n]$ obtained by passing the random process $X[n]$ with power spectrum $S_x(w)$ through a filter with transfer function $H(w)$. Then $S_y(w) = |H(w)|^2\ S_x(w)$[9]. As a result, if we let the input $X[n]$ be i.i.d., and also design a filter whose transfer function satisfies that $|H(w)|^2$ approximates the power spectrum of self-similar traffics, then the filter output straightforwardly become self-similar.

The autocovariance function of an exactly second-order self-similar process with self-similar parameter H is given by

$$(c/2)[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}]$$

for some constant $c > 0$; thus, its power spectrum $F_H(w)$ is

$$c|1-e^{-jw}|^2 \sum_{k=-\infty}^{\infty} |w + 2\text{®}k|^{-1-2H}$$

for $-\text{®} \le w < \text{®}$.. By taking the major term with $k = 0$, and replacing, inside the summand, $|w/$ by $|1-e^{-jw}|$, we obtain $F_H(w) \approx c|1-e^{-jw}|^{1-2H}$ for $-\text{®} \le w < \text{®}$. Hence, the problem is reduced to find a good filter for Poisson i.i.d. input with mean λ to yield

$$\begin{aligned} S_y(w) &= |H(w)|^2\ S_x(w) \\ &= a|H(w)|^2 = a|1-e^{-jw}|^{1-2H} \end{aligned}.$$

When transforming the problem to its equivalent domain of Z-transform, we can achieve our goal by letting $H(z) = (1-z^{-1})^{0.5-H}$, and result in

$$\begin{aligned} S_y(z) &= S_x(z)H(z)H(z^{-1}) \\ &= a(1-z^{-1})^{0.5-H}(1-z)^{0.5-H} \\ &= a(2-z^{-1}-z)^{0.5-H} \end{aligned}.$$

Apparently, by taking $z = e^{-jw}$ into the formula,

$$\begin{aligned} S_y(w) &= a(2-e^{jw}-e^{-jw})^{0.5-H} \\ &= a|1-e^{-jw}|^{1-2H} \end{aligned}.$$

Now let us examine the effect of such a filter-based self-similar traffic generator. The impulse response of the filter is equal to $h[n] = \Gamma(n+H-0.5)/[\Gamma(n+1)\Gamma(H-0.5)]$ for $n \ge 0$, where $\Gamma(\cdot)$ represents the Euler gamma function. As $h[n]$ is an infinite series, which is impractical in implementation, we limit its window size to W, i.e., $h[n]$ is forced zero for $n > W$. Denote the variance of m-aggregated series with average window m and truncation window being W by $C_m(0;W)$. Figure 1 illustrates the relation between $\log[C_m(0;10^3)]$ and $\log[m]$. We found that for $0 \le \log_{10}[m] \le \log_{10}[W]$, the straight line with slope $2H-2$ fits the curve of $\log[C_m(0;10^3)]$ against $\log[m]$. The resultant H' at the filter output is listed in Tab. I. From these data, we discovered that when the average window m is less than or equal to the truncated window W, the resultant H' tends to be a little smaller than the target H, although the deviation, defined as $(H'-H)/H$, is acceptably small in all cases. As a result, the degree of self-similarity is more accurate for smaller H.

For the case m>W, $C_m(0;W)$ can be represented by $A_H(W)m^{-1}$-$B_H(W)m^{-2}$, and thus

$$\frac{\partial \log[C_m(0;W)]}{\partial \log(m)} = -1 + \frac{B_H(w)/A_H(w)}{m - B_H(w)/A_H(w)}$$

Accordingly, the degree of self-similarity is determined by the ratio

$$\frac{B_H(w)}{A_H(w)} = \frac{(H-0.5)}{H(H+0.5)}\left(W + \frac{(H-1.5)}{2}\right) - O\left(W^{1-2H}\right)$$

where $O(\cdot\ )$ is the big-O notation.

Besides, we can also establish that if we wish to obtain a resultant $H' \in (0.5,1)$ (close to the target H) up to the average window $m' > W$, then it requires that

$$-1 + \frac{B_H(w)/A_H(w)}{m' - B_H(w)/A_H(w)} > - \quad = 2H'-2$$

which implies

$$W < m' < \left(\frac{2H'}{2H'-1}\right)\frac{B_H(w)}{A_H(w)}.$$

This indicates that the output self-similarity for the truncated model can be extended up to

$$\begin{aligned} &\left(\frac{2H'}{2H'-1}\right)\frac{B_H(w)}{A_H(w)} \\ &= \left(\frac{2H'}{2H'-1}\right)\left[\frac{(H-0.5)}{H(H+0.5)}\left(W + \frac{(H-1.5)}{2}\right) - O(W^{1-2H})\right] \end{aligned}$$

Since for $W>0$ and $0.5<H<1$

$$m' > W > \frac{(H-0.5)}{H(H+0.5)}\left(W + \frac{(H-1.5)}{2}\right) > \frac{B_H(w)}{A_H(w)},$$

provided ideally that $H'=H$. The above inequality, together with $m' > W$, gives that

$$(H-0.5)W < -\frac{1.5-H}{2} - \left(\frac{2H}{2H-1}\right)O\left(W^{1-2H}\right),$$

contradicting to $W>0$. We then conclude that as long as exact self-similarity is concerned (i.e., $H'=H$), to extend the output self-similarity up to the truncated window is impossible. From Fig.1, we further show by numericals that even if H' is allowed to be a little smaller than H, the conclusion remains.

四、結果與討論

In this report, we proposed a new model for self-similar traffic sythesization, based on the filter theory. This model is long range dependent with adjustable levels of bustiness and correlation. The model is parsimonious in its number of input parameters. Specially, it depends on only three parameters. H is the self-similar parameter, which controls the bustiness and autocorrelation of the synthesized traffic, λ defines the mean of the synthesized traffic and W determines not only the length of the filter but also the valid aggregation size of self-similar nature. Though filter length W limits the valid aggregation size of self-similarity, this phenomenon turns out to match the measured behavior of true network traffic, where the self-similar nature only lasts beyond a practically manageable range, but disappears as the considered aggregated window is much further extended. Other advantages of this model are that this filter-based model can synthesize traffic on the fly and always generate non-negative integers to represent network arrivals.

We verify the validity of our filter-based model through the mathematical analysis of its variance-time relation and statistics tests of V-T plot, R/S plot and periodogram plot (cf. table II). And we conclude that our model guarantee to synthesize self-similar traffic with high degree of accuracy in terms of self-similar parameter, H.

Table I. The resultant H' versus the targeted H. Deviation = $(H'-H)/H$.

Window Length = 10^4		
Targeted H	Resultant H'	Deviation
0.5001	0.5000961	-7.7984e-006
0.55	0.5481199	-0.0034
0.6	0.5961926	-0.0063
0.7	0.6909921	-0.0129
0.8	0.7809945	-0.0238
0.9	0.8599458	-0.0445

Table II. H USED VERSUS H RESULTANT

Window Length = 10^4			
H used	H resultant (V-T Plot)	H resultant (R/S Plot)	H resultant (Periodogram)
0.5001	0.4913099	0.5423777	0.5149618
0.55	0.5243788	0.5839482	0.5433228
0.6	0.5661478	0.6248291	0.5953569
0.7	0.6860798	0.6991949	0.6902056
0.8	0.7558080	0.7792713	0.7968477
0.9	0.8662405	0.8784192	0.8822255

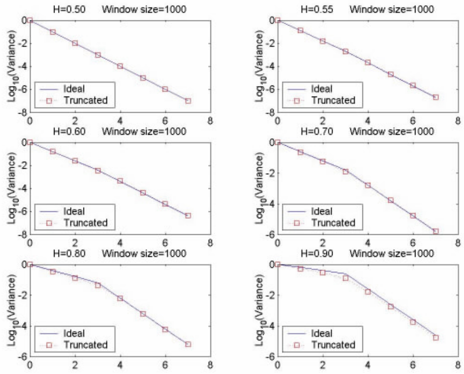

Figure 1. Variance-Time Analysis for $W = 10^3$. The slope of the blue line is equal to $2H-2$ for $m \leq W$, and -1 for $m > W$.

五、參考文獻

[1] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, pp. 1566-1579, 1995.

[2] D. E. Duffy, A. A. Mclntosh, M. Rosenstein and W. Willinger, "Statistical Analysis of CCSN/SS7 Traffic Data from Working Subnetworks," *IEEE JSAC*, pp. 544-551, June 1994.

[3] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking,* pp. 1-15, Feb. 1994.

[4] W.-C. Lau, A. Erramilli, J. Wang, and W. Willinger. "Self-similar traffic generation: the random midpoint displacement algorithm and its properties," *ICC'95*, pp. 466-472, June 1995.

[5] K. Meier-Hellstern, P. E. Wirth, Y. L. Yan and D. A. Hoeflin, "Traffic models for ISDN data users: office automation application," *In Proc. 13th ITC*, Copenhagen, Denmark, pp. 167-172, 1991.

[6] V. Paxson, "Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic," *Computer Communication Review*, pp. 5-18, Oct. 1997.

[7] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Networking,* pp. 226-244, June 1995.

[8] J. M. Peha, "Retransmission mechanisms and self-similar traffic models," *IEEE/ACM/SCS CNDSMS*, pp. 47-52, Jan. 1997.

行政院國家科學委員會補助專題研究計畫成果報告

真實網路封包流量
自我類似性質之成因探討與統計模式建立

計畫類別：□個別型計畫　　☑整合型計畫
計畫編號：NSC　－90－2213－E－009－056
執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：陳伯寧
共同主持人：張仲儒、陳耀宗、楊啟瑞
計畫參與人員：姚建、花凱龍、林建光、王聖賢

本成果報告包括以下應繳交之附件：
- □赴國外出差或研習心得報告一份
- □赴大陸地區出差或研習心得報告一份
- □出席國際學術會議心得報告及發表之論文各一份
- □國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學電信工程系所

中　華　民　國　九　十　一　年　五　月　十　六　日