A Whole Correlation Structure of Asymptotically Self-Similar Traffic in Communication Networks

Ming Li, Weijia Jia and *Wei Zhao

Dept. of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, SAR China
Email: {mingli, wjia}@cs.cityu.edu.hk
*Dept. of Computer Science, Texas A&M University, College Station, USA
*Email: zhao@cs.tamu.edu

Abstract

A closed form of autocorrelation functions about asymptotically self-similar processes is presented. The verification shows that this form best the real-traffic data on Ethernet investigated satisfactorily.

Keywords: teletraffic modeling, long-range dependence, autocorrelation functions

1. Introduction

World Wide Web (WWW) is flourish today with various applications such as information search and e-commerce. In addition, many classical applications also share the network resources such as voice/video on-demand, File Transfer Protocol (FTP), Telnet etc. These make the WWW traffic data complicated and stochastic. A data sequence in communication networks is called traffic trace $x(t)$, indicating the packet counts (i.e., number of packets at t). Recent experimental researches have revealed that the self-similarity is the nature of WWW traffic [1, pp. 196-197], [2]. That is, the behaviors of WWW traffic are well modeled by second-order self-similar processes with long-range dependence [1-3], [5] and [7-8].

Because autocorrelation functions play an important role in stochastic processes [4-5] and can be used to model the traffic data practically [6], it is important to study the representation of autocorrelation functions of self-similar processes. This paper presents a novel asymptotic model for self-similar traffic in Internet (such as WWW) using autocorrelation functions.

The paper is structured as follows: Section 2 gives the definitions related to the contents of the paper. Section 3 discusses the derivation of the form of autocorrelation functions with fixed finite lag for asymptotically self-similar processes. And Section 4 shows the verification with real-traffic data.

2. Definitions and Problem Put Forward

The term of self-similar processes (traffic) means second-order self-similar processes (traffic) with long-range dependence. The self-similar processes are defined by autocorrelation functions and they are classified into two models: One is exactly self-similar model and the other asymptotically self-similar model [5], [8]. Exactly second-order self-similar model is too narrow to model real traffic [8, p. 713] while the asymptotically self-similar model with fixed finite lag has not been specified exactly [5, pp. 101-102]. This is a problem discussed in this section.

Definition 1 [4]: Let

$$X = (X_t: t = 1, 2, \cdots)$$

be a stochastic process. If there exist the mean $E(X_t)$ and the variance $Var(X_t)$ $\forall\ t \in I_1$ (= 1, 2, $\cdots$), X is called a second-order process, where E is the expectation operator, Var is the variance operator.

Definition 2 [4]: Let

$$X = (X_t: t = 1, 2, \cdots)$$

be a second-order process. If

$$E(X_t) = \mu = const$$

and the autocorrelation function is the function of lag k

$$r(k) = \frac{E(X_{t+k} - \mu)(X_t - \mu)}{\sigma^2},\ k \in I, \quad (2\text{-}1)$$

where I is the set of integers, X is called a wide sense stationary process.

The following properties P1 and P2 have to be satisfied by all types of autocorrelation functions [4, pp. 107-108].

P1: $r(k)$ is an even function.

P2: $r(0) \geq |r(k)|$.

Definition 3: Let $\mathcal{R}$ be the set of real numbers, $f(x)$ and $g(x)$ be functions defined on $\mathcal{R}$ and c be a limit in $\mathcal{R}$. We say that $f(x)$ is *asymptotically equivalent to* $g(x)$ under the limit $x \to c$ if $f(x)$ and $g(x)$ are such that [9]

$$\lim_{x \to c} \frac{f(x)}{g(x)} = 1.$$

We write

$$f(x) \sim g(x)\ (x \to c), \quad (2\text{-}2)$$

where c can be infinity. It is proved [9] that if

$$f(x) \sim g(x)\ (x \to c)$$

and

$$g(x) \sim h(x)\ (x \to c)$$

then

$$f(x) \sim h(x)\ (x \to c).$$

That is,

$$f(x) \sim g(x) \sim h(x)\ (x \to c). \quad (2\text{-}3)$$

$f(x)$ is called slowly varying function if $\lim_{t \to \infty} \frac{f(tx)}{f(t)} = 1$ for all x.

Definition 4 [8]: A process X is called exactly second-order self-similar with parameter $H \in (0.5, 1)$, if its autocorrelation function is

$$r(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]\},\ k \in I. \quad (2\text{-}4)$$

Definition 5 [5], [8]: A process X is called asymptotically second-order self-similar with parameter $H \in (0.5, 1)$, if its autocorrelation function is with the form

$$r(k) \sim ck^{2H-2} \ (k \to \infty), \qquad (2\text{-}5)$$

where $c > 0$ is a constant.

According to Definition 5, the constraints for self-similar processes are summarized as C1 and C2.

C1: $\sum_k r(k) = \infty$.

C2: $r(k)$ decays hyperbolically.

A slight generalization of definition may be obtained by replacing the constant c with slowly varying functions [5], [7], [8]. However, "for most practical purposes, this generalization is not needed" [5, pp. 42]. Therefore, we do not consider the generalization mentioned in the paper.

It can be easily verified that $r(k)$ of (4) satisfies P1, P2, C1 and C2. Define the set $\mathcal{E}_1$ as

$$\mathcal{E}_1 = \{r; r(k) = 0.5[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}]\}. \qquad (2\text{-}6)$$

Then, autocorrelation functions of exactly second-order self-similar processes are elements of $\mathcal{E}_1$. The advantage of exactly self-similar model is that the whole correlation structure is specified and it is easy to use. Unfortunately, the exactly self-similar model is too narrow to model real traffic [7]. Attention has been paid to the asymptotically self-similar model. Define

$$S = \{r; r(k) \sim ck^{2H-2} \ (k \to \infty)\}. \qquad (2\text{-}7)$$

Then, $\mathcal{E}_1 \subseteq S$ and an element of S is an autocorrelation function of self-similar process. However, the form of autocorrelation functions in S for any fixed finite lag is not specified. So far as we know, the only closed form of autocorrelation functions of self-similar traffic was reported by [10] with the following expression for *compressed-video* sequences

$$r(k) = L k_i^{-\beta} \ U(k \ge k_i) + \sum_{i=1}^{J} w_i \ exp(-\lambda_i k) U(k < k_i),$$

$$k = 1, 2, \cdots, 0 < \beta < 1. \qquad (2\text{-}8)$$

where

$$\sum_{i=1}^{J} w_i = 1$$

and

$$L k_i^{-\beta} = \sum_{i=1}^{J} w_i \ exp(-\lambda_i k_i),$$

L is a constant, k_i is the lag value corresponding to the "knee" of the curve of an autocorrelation function approximated, λ_i is the rate of $exp(.)$ and $U(.)$ is the indicator function.

3. Derivation

This section focuses on the derivation of a form of autocorrelation functions with fixed finite lag specifying for asymptotically self-similar processes.

Statement 1: The autocorrelation function

$$(|k| + 1)^{2H-2}$$

is asymptotically equivalent to k^{2H-2} under the limit $k \to \infty$:

$$k^{2H-2} \sim (|k| + 1)^{2H-2} \ (k \to \infty), \ k \in I. \qquad (3\text{-}1)$$

Proof is clear and omitted.

Statement 2: A process X is called asymptotically second-order self-similar with parameter $H \in (0.5, 1)$, if its autocorrelation function is

$$r(k) \sim c(|k|+1)^{2H-2} \ (k \to \infty), \quad (3\text{-}2)$$

where $c > 0$ is a constant.

Proof is clear and omitted.

According to Statement 2, we obtain the following correlation structure

$$r(k) = c(|k|+1)^{2H-2}. \quad (3\text{-}3)$$

The $r(k)$ in (3-3) satisfies P1, P2, C1 and C2. Define the set $\mathcal{A}_1$ as

$$\mathcal{A}_1 = \{r;\ r(k) = c(|k|+1)^{2H-2}\}. \quad (3\text{-}4)$$

Then

$$\mathcal{A}_1 \subseteq S.$$

Let

$$R_H = \mathcal{A}_1 \cup \mathcal{E}_1 \subseteq S. \quad (3\text{-}6)$$

Now, we construct an interesting form of autocorrelation functions with fixed finite lag.

Statement 3: The autocorrelation function

$$(|k|^{\alpha} + 1)^{2H-2},\ \alpha \in (0, 1] \text{ and } H \in (0.5, 1)$$

is nonsummable at infinity and decays hyperbolically.

Proof: As $-1 < 2H - 2 < 0$ and $0 < \alpha \le 1$,

$$(|k| + 1)^{2H-2} \le (|k|^{\alpha} + 1)^{2H-2}.$$

Then

$$\sum_k (|k|^{\alpha} + 1)^{2H-2} = \infty.$$

On the other hand, as

$$(|k|^{\alpha} + 1)^{2H-2} \sim (|k|^{\alpha})^{2H-2} \ (k \to \infty),$$

$(|k|^{\alpha} + 1)^{2H-2}$ decays hyperbolically.

According to Statement 3, $\mathcal{A}_1$ is extended as

$$\mathcal{A}_2 = \{r;\ r(k) = c(|k|^{\alpha}+1)^{2H-2}\}. \quad (3\text{-}7)$$

Clearly,

$$\mathcal{A}_1 \subseteq \mathcal{A}_2$$

and $\mathcal{A}_2$ is an extension of $\mathcal{A}_1$. Let

$$R_{H,\alpha} = \mathcal{A}_2 \cup \mathcal{E}_1. \quad (3\text{-}8)$$

Then,

$$R_H \subseteq R_{H,\alpha}. \quad (3\text{-}9)$$

The key point is that the whole correlation structure of each element of $R_{H,\alpha}$ is specified.

4. Verification

A data file called pAug.TL [11] is a sequence of real-traffic data. Fig. 1 (a) illustrates its time sequence. From the point of view of the parameter estimation, the non-parameter estimation of the autocorrelation function of pAug.TL in Fig. 1 (b) is regarded its target autocorrelation function. To evaluate the result of modeling, mean square error

$$M^2(\hat{r}) = E[(\hat{r} - r)^2] \triangleq m(H, \alpha)$$

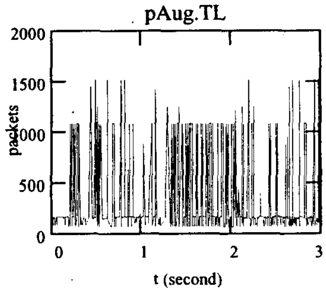
is used as a criterion, where $\hat{r}$ is the estimation of r. Under the least mean square error, we obtained the following closed form of autocorrelation function for pAug.TL

$$r_{ext}(k;\ 0.61, 0.064) = (|k|^{0.064} + 1)^{-0.78} \quad (4\text{-}1)$$

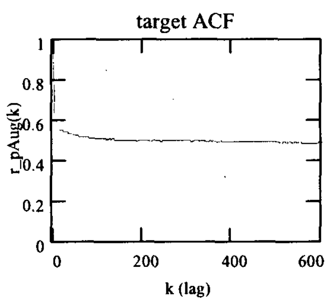
where $r_{ext}(k;\ H,\ \alpha)$ stands for the parameter estimation of the autocorrelation function of pAug.TL. Fig. 1 (c) shows the result of Equation (4-1). For the function of (4-1), its mean square error is

$$M^2(r_{ext}) = m(0.61, 0.064) = 5.73\times10^{-5}. \quad (4\text{-}2)$$

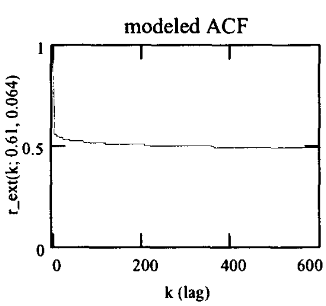
The value of $M^2(r_{ext})$ interprets that the estimation result of (4-1) best fits the target one as shown in Fig. 1 (d).



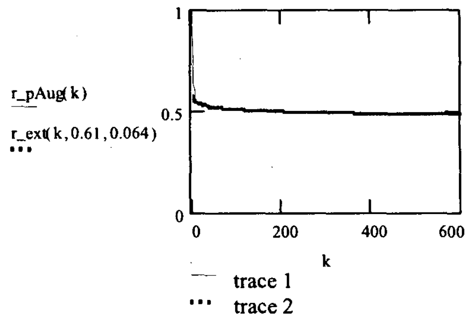
(a) An illustration of pAug.TL



(b) Target autocorrelation function sequence of pAug.TL



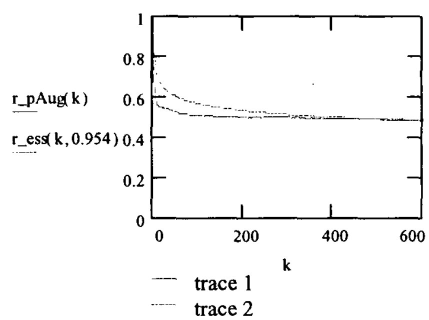
(c) Modeled autocorrelation function



(d) The modeled autocorrelation function best fits its target

Fig. 1 Illustrations for modeled autocorrelation function and result of best fitting

To evaluate the benefit of our model of (3-7), the approximation with exactly self-similar model (2-1) is given in Fig. 2, where $r_{ess}(k;\ 0.954)$ is the best result at the extent of the least mean square error being 0.01. In comparison with the result (4-2), the benefit of our model, $\hat{r}(k) = (|k|^{\alpha}+1)^{2H-2}$, $\alpha \in (0, 1]$, $H \in (0.5, 1)$ and $k \in I$, is obvious.



$M^2(\hat{r}) \approx 0.01$

Fig. 2 Matching $r_{pAug}(k)$ with exactly self-similar model

5. Conclusions

One of difficulties in traffic modeling with the asymptotically self-similar model is that the form of autocorrelation functions of asymptotically self-similar processes with fixed finite lag was not specified. This paper presented a form of functions for asymptotically self-similar processes.

Acknowledgement

Authors appreciate the useful discussions with Professors Xiaotie Deng and To-Yat Cheung. This work was partially sponsored by City University of Hong Kong under grants.

References

[1] William Stalling, High-Speed Networks, TCP/IP and ATM Principles, Prentice Hall, 1998

[2] Mark E. Crovella, Azer Bestavros, Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes, *IEEE/ACM Trans. on Networking*, vol. 5 (6), Dec., 1997, pp. 835-846

[3] Will E. Leland, Murad S. Taqqu, Walter Willinger and Daniel V. Wilson, On the Self-Similar Nature of Ethernet Traffic (Extended Version), *IEEE Trans on Networking*, vol. 2, no. 1, Feb., 1994, pp. 1-15

[4] M. B. Priestley, Spectral analysis and Time Series, Academic Press, 1981

[5] Jan Beran, Statistics for Long-Memory Processes, Chapman & Hall, 1994

[6] San-Qi Li, Chia-Lin Hwang, Queue Response to Input Correlation Functions: Continuous Spectral Analysis, *IEEE/ACM Trans. on Networking*, vol. 1, no. 6, Dec. 1993, pp. 678-692

[7] Boris Tsybakov, Nicolas D. Goergannas, On Self-Similar Traffic in ATM Queues: Definitions, Overflow Probability Bound, and Cell Delay Distributions, *IEEE Trans. on Networking*, vol. 5, no. 3, June 1997, pp. 397-409

[8] Boris Tsybakov and Nicolas D. Georganas, Self-Similar Processes in Communications Networks, *IEEE Trans on Information Theory*, vol. 44, no. 5, Sep., 1998, pp. 1713-1725

[9] J. D. Murray, Asymptotic Analysis, Springer-Verlag, 1984

[10] Changcheng Huang, Michael Devetsikiotis, Ioannis Lambadaris and A. Roger Kaye, Modeling and Simulation of Self-Similar Variable Bit Rate Compressed Video: A Unified Approach, *Proc., ACM SIGCOMM'95*, 1995, pp. 114-125

[11] Data traces can be obtained via anonymous ftp from bellcore.com under /pub/wel/lan_traffic