Simulation of Self-similarFlowBased on Fractal Gaussian Noise Method

Li Jie

School of Computer Science and Engineering

Xi'an Technological University

Xi'an, 710032, China

e-mail: lijiekpt@163.com

Tang Junyong

School of Computer Science and Engineering

Xi'an Technological University

Xi'an, 710032, China

e-mail: 83090141@qq.com

Lu Ying

School of Computer Science and Engineering

Xi'an Technological University

Xi'an, 710032, China

e-mail: luyingkk@sina.com

Abstract—**The conventional network traffic flow models are mostly based on Poisson model. With the continuous development of network services, studies found that the actual network traffic has a long-range dependence (LRD) now and in a very long time, which is a kind of self-similarity. In this paper, RMD and Fourier algorithm were adopted to simulate and analyze a self-similar model of FGN. They generated the necessary sequence of self-similar traffic. Then the article uses R/S method to verify H value of the generated sequence of self-similar traffic in order to verify the self-similarity of the self-similar traffic sequence. The existence of self-similarity is verified by experiments, and the advantage and disadvantage of RMD and Fourier algorithm are analyzed.**

Keyword-Self-similarity; Traffic Flow; FGN; R/S

I. INTRODUCTION

The study of network traffic flow plays an important role in improving network structure, improving network performance and improving network security. The traditional business model is based on the poisson process, which is a short-range correlation model. In the 1990 s, Leland and others in the literature[1]for the first time explicitly put forward the network traffic with self-similarity, and proves that the self-similarity of network traffic flowis network business data flow within a long time both have a

long self-similarity.After that, the researchers observed and analyzed some existing networks, and found that no matter how the network topology, number of users and service types changed，this self-similarity always exist. Based on *FBM* model, the simulation generation experiment of flow sequence was carried out using *RMD* and Fourier transform method respectively, and the self-similar sequences generated by *R/S* algorithm were used to verify and analyze the results.

II. DEFINITION OF THE SELF-SIMILAR PROCESS OF NETWORK TRAFFIC

A statistical process of stable covariance is investigated $X=\{Xt:t=0,1,2,3...\}$，the mathematical expectation of the process is that $\mu=E[X_t]$ is constant, the variance $\sigma^2=E[(X_t-\mu)^2]$ is a finite value, the auto correlation function $r(k)=E[(X_t-\mu)(X_{t+k}-\mu)]/\sigma^2$, $k=1,2,3,\ldots$, it's only related to k. X_t can be understood as the number of network business entities that arrive in t unit time. It is assumed that the autocorrelation function of X has the following form[2]:

$$r(k)\sim k^{-\beta},0<\beta<1,k\to\infty \tag{1}$$

For each $m=1,2,3,\ldots$, make $X^{(m)}=\{X^m{}_t:t=0,1,2,3,\ldots\}$ represents a new sequence. $X^m{}_k=1/m(X_{km-m+1}+\ldots+X_{km})$, $k=1,2,3,\ldots$ stands for the aggregation of length m. If $X^{(m)}$

also defines a generalized stationary random process, $r^{(m)}(k)$ represents the auto correlation function of $X^{(m)}$. If for all m, there are:

$$r^{(m)}(k)=r(k) \sim k^{-\beta}, 0<\beta<1 \qquad (2)$$

When $k \rightarrow \infty$, X is the exact second order self-similar process, and it is said that $H=1-\beta/2$ is its self-similar parameter.

III. FBM FLOW MODEL AND SIMULATION MODELING

In the network modeling of traffic flow, the parameters of use less as far as possible to the actual requirements of the network traffic modeling, another is required to produce synthetic data sequence to show similar features and measurement data. In the modeling of network traffic flow, it is required to model the traffic flow of the actual network with as few parameters as possible.In addition, the required synthetic data sequences are required to display similar characteristics to the measured data. The fractal Brownian motion FBM is a kind of convenient traffic model. The model is based on the Fractal Gaussian Noise(FGN) process, which is an incremental process of FGN[3].

In the simulation experiment, the FBM model is modeled and studied by the random mid-point method and the discrete Fourier transform method, which results in thetraffic flow of the Fractal Gaussian Noise.Moreover, the network traffic sequences generated by different algorithms are compared, and their performance is analyzed. The network traffic sequences generated by different algorithms are compared for different self-similarity coefficients, and then analyze their performance.

A. Random Mid-Point Method

The random mid-point method (RMD method) is used to generate simulation sequences by gradual subdivision method.The greatest advantage of this method is its rapidity. Algorithm principle: Firstly, itdivides the time interval [0,T], and then calculates the arithmetic mean of the left and right points，iterates to get the midpoint value and addsan offset. Based on *matlab* simulation platform, RMD algorithm steps are written as follows [4,5] :

The original business flow data is performed by the weighted iteration of the pseudo-random sequence (by mathematical algorithm, the random number is calculated according to the law of random number).The foreground sequence is generated, performance analysis is performed on the foreground sequence, and the model parameters are constantly adjusted so that the performance of the foreground sequence matches the original sequence, and finally the practical RMD model is formed.

1）In *matlab, randn()* in *M* language is used to generate a pseudo-random sequence with mean value of 0 and variance of 1 as the original sequence(G=G(n)(n=1,2,3....2^m));

2）The original sequence is weighted，and the weighted coefficient is $\theta_i = \sqrt{\frac{(1-2^{2H-2})}{2^{2Hi}}}$,background sequence is $D(i)=G(n)\theta_i$(i is the number of iterations)；

3）The number of iterations is m，generates foreground sequence $X=X(n)(n=1,2,3,....2^m)$, Let $X = 0(0)$, $X(2m) = G(1)$

When$j$=1:

$$X(j*2^{m-i})=X(2^{m+1-i})/2+ G(2^{m+1-i})\sqrt{\frac{(1-2^{2H-2})}{2^{2Hi}}}$$

When $j!$=1:

$$X(j*2^{m-i})=(X((j-1)*2^{m-i})X((j+1)2^{m-i}))/2+G(2^{m+1-i}+(j+1)/2)$$

$$\sqrt{\frac{(1-2^{2H-2})}{2^{2Hi}}},$$

this is j=1，3，5，......2^i-1

4）The foreground sequence after iteration is difficult to analyze mathematically，and let's do the cosine of it, $SY=cos(X)$, The resulting sequence is finally obtained-$SY(n)(n=1,2,3,...2m)$.

B. Fourier Transform Method

Fourier transform method is to obtain the business data by Fourier transform of the spectrum of Fractal Gssian-Noise, which has high accuracy. The basic idea is to produce a spectrum, and then fast Fourier inversion for this

spectrum.

Based on *matlab* simulation platform, the steps of fast Fourier inversion conversion are as follows [6] :

1) Tectonic sequence {f1,f2,…fn/2}, fi=f（2πj/n，H）in it, The power spectrum of the FGN process from 2π/n to π is respectively corresponding to the corresponding frequency. The amplitude of the frequency domain sample of the self-similar sequence is $\sqrt{f}$.

2) Hybridization ｛f_1｝, a separate exponential random variable with a mean of 1 times each ｛ f_i ｝, gets the sequence ｛ $f_i^{'}$ ｝.

3) Construct a sequence ｛$z_{1,}$｝. Take the complex sequence $\{z_1, z_2, \ldots z_{n/2}\}$, among $|z_i| = (f_i')^{1/2}$, the phase of z_i is randomly distributed between $0 \sim 2\pi$.

4) Solving the complex sequence $\{z'_1, z'_2, \cdots, z'_{n/2}\}$,

$z'_j = 0$, j=0

$z'_j = z_j$, 0<j≤n/2

$z'_j = \bar{z}_{n-j}$, 0<j≤n/2

$\bar{z}_{n-j}$ is the complex conjugate of z_{n-j} .

5) The inverse Fourier transform of ｛ z'_j ｝ is obtained, and the approximate sample point {x_i} of *FGN* is obtained. The focus of this method is the calculation of spectrum. That is to say the power spectrum[5] of *FGN* process is:

f （ω, H）$= A$（ω, H）$[|\omega|^{-2H-1} + B$（ω, H）]

In the formula, $0 < H < 1, -\pi < \omega < \pi$

A（ω, H）$= 2sin(\pi H)\Gamma(2H+1)(1-cos\omega)$

B（ω, H）$= \Sigma[(2\pi j+\omega)^{-2H-1} + (2\pi j-\omega)^{-2H-1}]j$ j=1,2,、、、 , ∞

The main difficulty in calculating power spectrum is the infinite sum. Now precise formula hasn't been found, and an approximate formula is used instead.

B（ω, H）$\approx B_3$（ω, H）$= a^d_1 + b^d_1 + a^d_2 + b^d_2 + a^d_3 + b^d_3 + (a^g_3 + b^g_3 + a^g_4 + b^g_4)/8\pi H$

In the formula, d=-2H-1, g=-2H, $a_k = 2k\pi+\omega$, $b_k = 2k\pi-\omega$

IV. EXPERIMENT AND ANALYSIS OF SELF-SIMILAR SEQUENCES

A. The Generation of Self-similar Sequence

When using *MATLAB* simulation. ， *RMD* algorithm and Fourier transform method were used respectively.The self-similar sequence with a similar parameter of 0.7,0.8 and 0.9 was set, and the sequence length was 8192.FIG. 1(a) is the result of *RMD* algorithm when *H*=0.9, and (b) is the result of Fourier transform method when *H*=0.9.
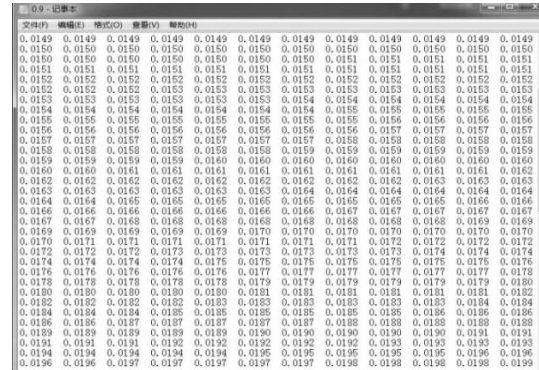
(a) When *H*=0.9, *RMD* Algorithm Generates Results
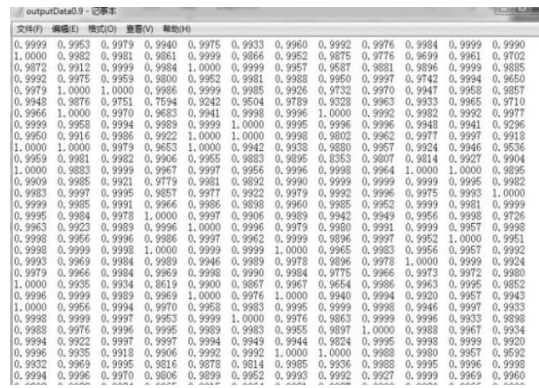
(b) *H*=0.9 When Fourier Transform Method Generates Results.

Figure 1. Self-similar Sequence Generated by *RMD* Algorithm and Fourier Algorithm

B. The Common Parameter Estimation Method in the Process of Self-similarity

In this paper, *R/S* method and variance time curve method are used to test the self-similarity of generation sequences. For self- similar time series $X = \{X_1, X_2, X_3, \ldots, X_L\}$, calculate the partial

sum $Y(n) = \sum_{i=1}^{n} X_i$ and variance of the sequence $S(n)$. Then calculate *R/S* method statistics:

$$\frac{R}{S}(n) = \frac{1}{S(n)}\left[\max_{0 \le t \le n}(Y(t) - \frac{t}{n}Y(n)) - \min_{0 \le t \le n}(Y(t) - \frac{t}{n}Y(n)) \right] \quad (3)$$

It is known by statistical properties，$\frac{R}{S}(n) \approx cn^H$. We can do the logarithm of both sides: $\log \frac{R}{S}(n) \approx \log c + H \log n$，

among, $\log c$ is a constant. $\log \frac{R}{S}(n)$ is the vertical axis, $\log n$ is the horizontal axis curve fitting, Because the curves of the two curves are linear, the curves can be fitted with straight lines. And then the slope of the line is an estimate of the *hurst* exponent.

C. Self-similarity Detection of Generating Sequence

For the random sequences generated by *RMD* algorithm, this paper uses the *R/S* method to estimate the Hurst parameters.In the *RMDS* algorithm, when *H*=0.7, 0.8 and 0.9, the self-similar parameter *H* value of the sequence is detected. As shown in fig.2, two dotted lines in the scatter plot generated by the test results show a line with a slope of 1 and a slope of 0.5. There are many scattered points in the graph which can be obtained by linear fitting. The slope of the line is the *hurst* parameter, whose value is between 0.5 and 1, and the specific data is shown in table 1. *H* specifies the theoretical value for the algorithm, and H' is the detection of *H* value.In order to better verify the matching degree of the estimated value and the theoretical value, the relative error can be used to represent. This：$\Delta H = (H' - H)/H$ ·
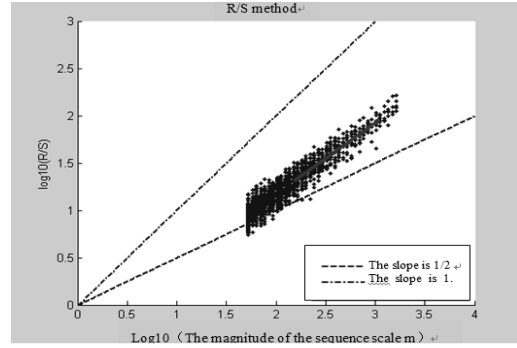
(a)　　H=0.7
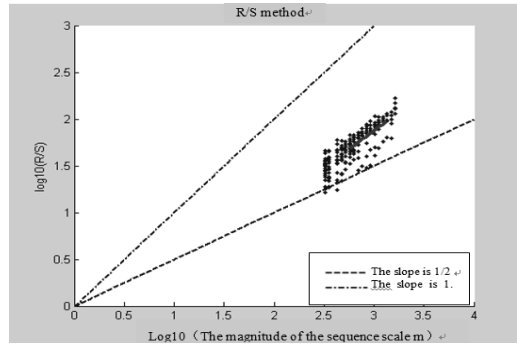
(b)　　H=0.8
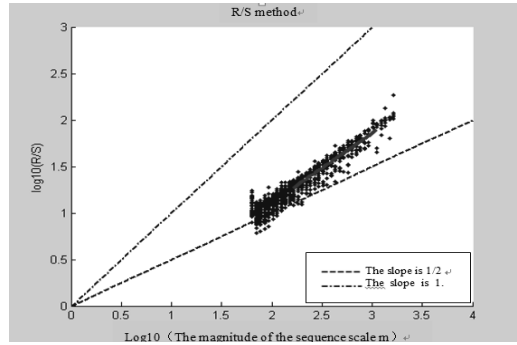
(c)H=0.9

Figure 2.　Scattersof three sample sequencegenerated by *R/S* detectedalgorithm

TABLE I.　TEST RESULTS OF SELFSIMILAR SEQUENCE GENERATED BY R/S ALGORITHM(RMD ALGORITHM)

Sample	Theory H	Detection H'	Compatibility(H'-H)/H
1	0.70	0.7242	0.00 03%
2	0.80	0.7869	-0.00 01%
3	0.90	0.8768	-0.00 02%

TABLE II.   TEST RESULTS OF SELF-SIMILAR SEQUENCE
GENERATED BY R/S ALGORITHM(FOURIER ALGORITHM)

Sample	Theory H	Detection H'	Compatibility(H'-H)/H
1	0.70	0.7013	0.00002%
2	0.80	0.8008	0.00001%
3	0.90	0.8997	-0.000003%

It can be seen from table I and table II that the obtained value is basically matching with the theoretical value of the original self-similar flow, and its relative error is small, thus proving the correctness of the simulation algorithm.The self-similar sequences generated by *RMD* and Fourier transform are compared. We can see that: The *RMD* algorithm has no complicated functions to participate in the calculation, the calculation process is clear, the speed is fast, and the self-similar sequence can be generated quickly, but the accuracy of the generation sequence of Fourier transform method is relatively high. The *RMD* algorithm and Fourier transform method can directly select the variable values that can affect the self-similar characteristics of the generation sequence; However, the *RMD* algorithm generates the sample data sequence through the continuous segmentation interval, and the generated sequence is within two endpoints. Therefore, it is important to choose which endpoint is the endpoint of the initial segmentation, which has an influence on the self-similar characteristics of the resulting data sequence, which will bring instability to the later self-similarity parameter estimation.

V. CONCLUSION

The self-similarity of network traffic determines the behavior characteristics of the network, and only self-similar modeling can accurately describe the actual situation of the network.Based on *FBM* model, the flow sequence is generated by *RMD* and Fourier transform method respectively. This experiment shows that the simulation algorithm is effective.

REFERENCES

[1] Leland W E,aqqu M S,Willinger W,et a1.On theself-similar nature of Ethernet traffic(extended version)[J].IEEE/ACM Transactions on Networking,1994,(1),P:1-15.

[2] Lee YH,KassamSA.Generalized median filtering and related nonlinear filtering techniques[J].IEEE Transactions on Acoustica,Speech,Signal Processing,1985,33(3):672-683.

[3] Mandelbort B B,Van Ness J.Fractional Brownian motions,factional noises and applications[J].1968,10(4):422-437.

[4] Zhu Xiao-wen, Tan Xian-hai, Chen Baixiu. Research on modeling of network traffic based on(MK)-FBM[J]. Railway Computer Application, 2009, 18(6): 10-13.

[5] Li Lin-fengQiu Zheng-ding .An Iterative Method to Estimate Hurst Index of Self-similar Network Traffic[J]. Journal of Electronics& Information Technology.2006, 28(12):136-158.

[6] Xu Mingwei, Tong Aijun. Network performance testing based on self-similar modeling[J]. Computer Engineering and Applications, 2002, 38(5): 56-59.

[7] Fu Lei-yang, Wang Ruchuan, Wang Haiyan, et al. Implementation and application of computing self-similar parameter by R/S approach to analyze network traffic[J]. Transactions of Nanjing University of Aeronautics amp; Astronautics, 2007, 39(3): 358-362.

[8] Wang Xin. Modeling and prediction of self-similar network traffic [D]. Beijing: THU,2003.

[9] Wang Xifeng, GAO Ling,Zhang Xiaoluan. Analysis and research on self-similar network traffic forecast[J]. Computer Technology and Development, 2007, 17(11): 42-45.