LRD Traffic Predicting Based on ARMA

Bo Gao[1], Qinyu Zhang[1], and Naitong Zhang[2]

[1] Shenzhen Graduate School, Harbin Institute of Technology. Shenzhen 518055, China

[2] School of Electronics and Information Engineering, Harbin Institute of Technology. Harbin 150001, China

gaobo@hitsz.edu.cn, {zqy,ntzhang}@hit.edu.cn

Abstract. The prediction of long range dependence (LRD) is the critical problem in network traffic. The traditional algorithms, such as Markov model and ON/OFF model, may provide high computation cost and low precision. In this study, a novel method based on empirical mode decomposition (EMD) and ARMA model was proposed. The results show that EMD could offer the function of canceling the LRD in traffic data. After transforming LRD to SRD (short range dependence) by EMD processing, the LRD traffic data could be predicted with high accuracy and low complexity by ARMA model. Meanwhile, the results indicate the usefulness of EMD in the applications of network traffic prediction.

Keywords: LRD, EMD, ARMA, Predicting.

1 Introduction

The amount of network traffic increases day and day with the rapid development of the Internet. The LRD (long range dependence) characteristic of network traffic leads to larger packet loss rate and queuing delay in practice than theoretic analysis according to traditional queuing theory. Obviously, LRD property of network traffic should be studied deeply in order to insure the QoS of network. Unfortunately, the SRD (short range dependence) model can not capture the LRD characteristic which is one trait in network traffic. In order to predict the network traffic accurately, researchers studied mainly in two aspects: the first one was using SRD model to fit LRD traffic, but the results were not ideal; the second was trying to discover some LRD model to capture the network traffic LRD characteristics, but the time and space complexity of related algorithm was high.

The SRD models that are mainly related to Markov and regress model, have been already confirmed that they fit the LRD network traffic badly [1-4]. The LRD model included FARIMA (fractal ARIMA) and FBM model (Fractal Brownian Motion), etc. FARIMA model in [5] was used to structure LRD sequence and proved that this kind of model could fit the LRD network traffic efficiently by analyzing the autocorrelation function of sequence generated by model itself. Some scholars adopted other complex models, such as multi-fractal wavelet, wavelet neural network

P. Ren et al. (Eds.): WICON 2011, LNICST 98, pp. 472–479, 2012.

[6-8] and chaos model [9]. Although all of these models had good precision, the algorithm complexity was high. Besides those statistical model mentioned, ON/OFF model [4] that is one kind of physical model can be applied to predict LRD sequence and has specific physical meaning, nevertheless the precision and flexibility of the physical model is worse than statistical one. Researchers proposed methods that could transfer LRD sequence into SRD. Reference [6] figured that the multi-fractal model could obtain SRD traffic from LRD. It means that we can adopt SRD model to predict LRD which is hardly analyzed directly with SRD linear model. Furthermore, reference [7] and [8] found the wavelet transform coefficients of FBM and FGN had SRD property in the same scales, instead of LRD.

This paper presents a method base on EMD and ARMA which can transfer the LRD sequences into SRD sequences, then model and predict the LRD traffic. We proved the theory in our former paper [10] and presented the simulation results. In this paper we continued studying the prediction about those IMF components decomposed by EMD. The paper is organized as follows. In Section 2, we summarize the basic definition of LRD, EMD and ARMA model. Then the simulation results are given in Section 3, but we find IMF1 prediction results are not accurate. Then Section 4 presents a method to promote IMF1 prediction accuracy. Section 5 concludes the paper.

2 EMD and ARMA

2.1 LRD and Empirical Mode Decomposition

Firstly, let us have a look at the definition of LRD. Assume that there exists a stationary stochastic process $\{X(t), t > 0\}$ with the autocorrelation function expressed as $r(\tau)$. If $\int r(\tau)d\tau = \infty$, that means, the autocorrelation function is non-integrable, then $\{X(t), t > 0\}$ is an LRD process. Thus we can know the present data are related to all of historical data against SRD process which satisfies $\int r(\tau)d\tau < \infty$.

Then we talk about the definition and characteristic of EMD. EMD is used to decompose the original signal to a number of intrinsic mode functions (IMF) in Hilbert-Huang transform (HHT) [11]. The EMD method is a necessary pre-processing before signals are transformed by the Hilbert transform. It will separate the data into a collection of IMFs defined as the functions satisfying the following conditions: (a) in the whole data set, the number of extrema and the number of zero-crossing must either equal or differ at most by one, and (b) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

With above definition, we can decompose any function, for example $x(t)$, as follows: Identify all of the local extrema, and then connect all the local maxima by a cubic spline line as the upper envelope. Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data between them. The mean of the upper and lower envelopes is designated as $m(t)$, and the difference between the data and $m(t)$ is $h(t) = x(t) - m(t)$. If $h(t)$ satisfies IMF condition, we obtain the first IMF (named IMF1), or we see $h(t)$ as $x(t)$ and repeat the process above until:

$$\left\{ \sum_{t=0}^{T} \frac{\left[h_{1,k-1}(t) - h_{1,k}(t) \right]^2}{h_{1,k-1}(t)^2} \right\}^{\frac{1}{2}} < 0.1 \tag{1}$$

In (1), $h_{1,k}(t)$ means the difference value between $x(t)$ and the mean of envelops after k times iterations.

Then, we separate the first IMF designated as $C_1(t)$ from the rest of the signal by

$$l(t) = x(t) - C_1(t) \tag{2}$$

Treat *l(t)* as the new data and subjected to the same sifting process as described above. The procedure can be repeated to obtain all the subsequent and finally we have

$$x(t) = \sum_{i=1}^{n} C_i(t) + r(t) \tag{3}$$

In [10], we proved that the LRD data did not possess LRD property after EMD and the computer simulation results also showed that every IMF component was SRD. Therefore, the LRD model can be replaced by some SRD model which is relatively simple, such as ARMA.

2.2 ARMA Model

ARMA is a famous SRD model which is widely applied in the field of data modeling and predicting. Its whole name is autoregressive moving average and the definition is as follows: assume that $\{X(n), n = 0, \pm 1, \pm 2, \dots\}$ is one stationary process with mean zero and $\xi(n)$ is a white noise process with variance σ, then for any n, if

$$X(n) - \phi_1 X(n-1) - \cdots - \phi_p X(n-p) = \xi(n) - \theta_1 \xi(n-1) - \cdots - \theta_q \xi(n-q) \tag{4}$$

then we call that $X(n)$ is a ARMA process or ARMA(p, q). In (4), the parameters ϕ, θ and σ need to be estimated according to some known data. To identify the value of p and q, we apply the AIC criterion which was proposed by H. Akaike [13] in 1974. The AIC function is

$$\mathrm{AIC} = \ln \hat{\sigma} + \frac{2 \cdot (p+q+1)}{N} \tag{5}$$

where $\hat{\sigma}$ is the estimation of $\xi(n)$'s variance σ and N is the number of observation data samples. Choose p and q from some certain range (p and q are both integer), such as the interval (0, ln$N$), to make AIC minimum, thus we can obtain the order of ARMA model. Then use the inverse function method [14] to estimate other parameters.

3 Prediction by Using ARMA

According to our research results in [10], we could apply ARMA to predict the LRD data which may be transformed to a series of SRD data. The simulation environment we adopt is Intel Pentium 4 CPU 2.8G Hz, two pieces of 512M memory, Matlab 7.0 and Microsoft Windows XP Professional SP2.

The LRD data are selected from Bellcore in 1989 named BC-pOct89 [12] which are the same with [10]. In [10] we decomposed the data to 8 IMFs every of which has 450 data. Now use ARMA to model and predict these 8 IMFs one by one in this paper. For each IMF component, select first 300 data as the training data to estimate parameters and establish the model, and the other 150 data are used to prediction process for detecting the difference between the forecast value and the actual one. The MSE (mean squared error) is applied to evaluate the result of predictions. Assume the value of MSE is *M*, then we have:

$$M = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{x}(i) - x(i)\right]^2 \tag{6}$$

In (6), n is the number of traffic data, $\hat{x}(i)$ is the ith prediction value and $x(i)$ is the ith actual traffic data value. After the research we found that each forecast error from IMF2 to IMF8 is much lower than the error of IMF1 and with the IMF order increasing all of the MSEs decay nearly at an exponential speed. Fig. 1 shows the prediction results.

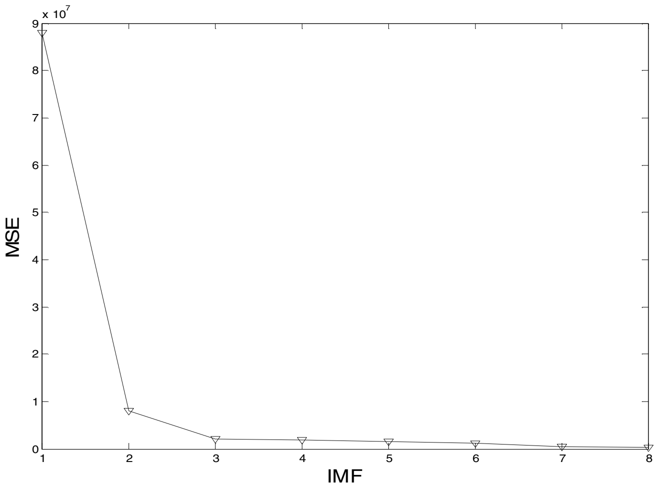

Fig. 1. The MSE of each IMF component after ARMA

We add the IMF components from IMF2 to IMF8 as a whole body to predict because of their higher accuracy, as a result, not only the number of our models is reduced, but also the prediction efficiency is raised. Fig. 2 is the prediction result of IMF1 and Fig. 3 is the prediction result of the sum of IMF2 to IMF8.

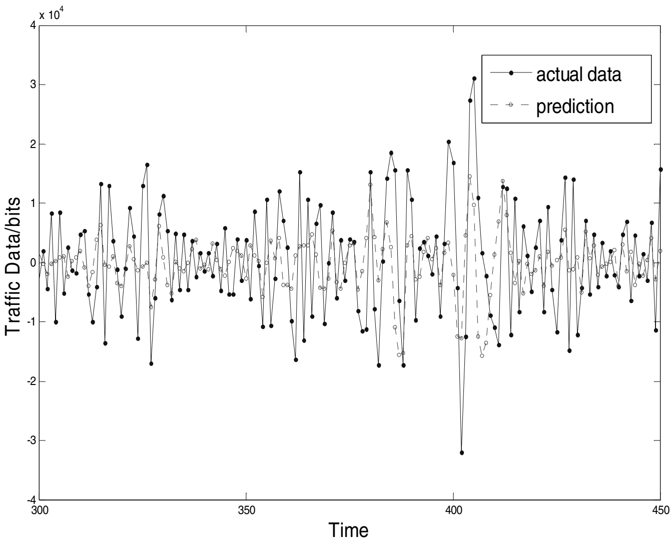

Fig. 2. IMF1 and its prediction

4 Promotion Predicting Precision of IMF1

For the problem of IMF1's poor prediction accuracy, after study we found that the difference operation could improve the prediction accuracy of IMF1.

ARMA model is mainly applied for stationary process, while in the actual engineering there are so many non-stationary processes. Non-stationary data will lead to the increase of ARMA model prediction error. Therefore, in order to reduce the IMF1 prediction error, we should add pre-processing operation on the IMF1 data to reduce its non-stationary characteristic. There are a number of data smoothing methods, such as difference operation, so we use difference method to smooth IMF1. Firstly, IMF1 data are operated by difference calculation some times, until the precision is satisfying. Secondly, we input the processed data to ARMA model for prediction, then the predicting data are obtained.

We could assume that $\{X(n)\}$ is a discrete time sequence, so we know the first order difference of $X(n)$ is $Y(n) = X(n+1) - X(n)$, the second order difference of $X(n)$ is $Z(n) = Y(n+1) - Y(n) = X(n+2) - 2X(n+1) + X(n)$, and the rest may be deduced by analogy. After the difference calculation, the new sequence's non-stationary characteristic can be inhibited. However, the amplitude of the new sequence has been changed, as a result, MSE can not be able to measure the prediction performance. Therefore, we select NMSE (Normalized Mean Squared Error) to measure the prediction performance for IMF1 components.

If the value of NMSE is N, then

$$N = \frac{1}{\hat{\sigma} n} \sum_{i=1}^{n} \left[\hat{x}(i) - x(i) \right]^2 \tag{7}$$

where $\hat{\sigma}$ is expressed to variance of prediction data, n is the number of traffic data, $\hat{x}(i)$ is the ith prediction value and $x(i)$ is the ith actual traffic data value. NMSE is the normalized MSE, so it can eliminate the impact of forecast errors with amplitude changing.

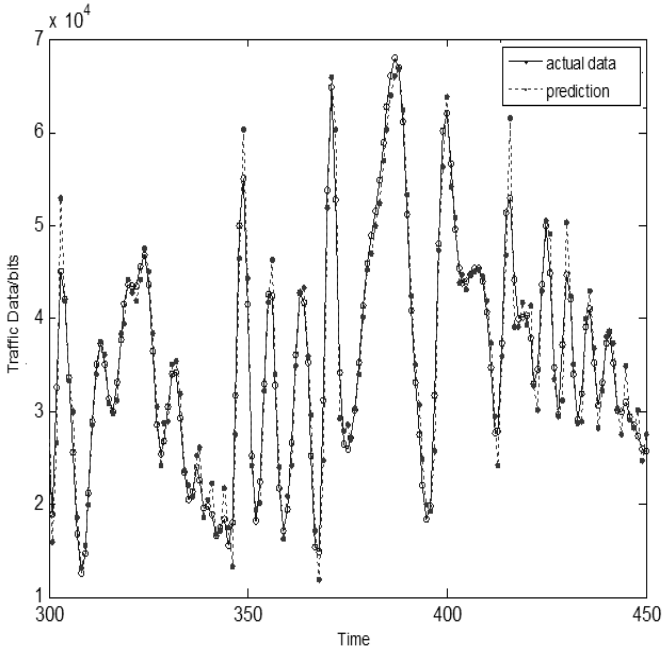

Fig. 3. The sum of IMF2 to IMF8 and the sum's prediction

Fig. 4 is the NMSEs of ARMA predicting by using 0-5th order difference operation. We can know that with the increase of difference order, the value of NMSE decreases at a fast speed, like exponential function. This means that difference plays a great role in promotion the IMF1 prediction accuracy and the effect is obvious. But the complexity of this method increases system overhead. In practice, by considering the time cost of model building and requirement of prediction accuracy, people should select the proper difference order to make the prediction system overall optimal. To illustrate feasibility of our proposal method, we choose 5th-order difference operation to process the traffic data. The prediction results obtained is shown in Fig. 5, from which we can see that the difference of the prediction results and the true value is very small. Table 1 is NMSE value of 0-5th order difference data prediction.

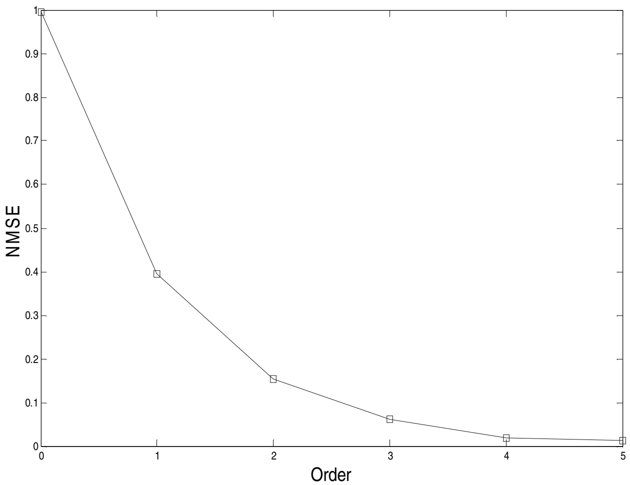

Fig. 4. NMSE of IMF1 prediction with every order difference

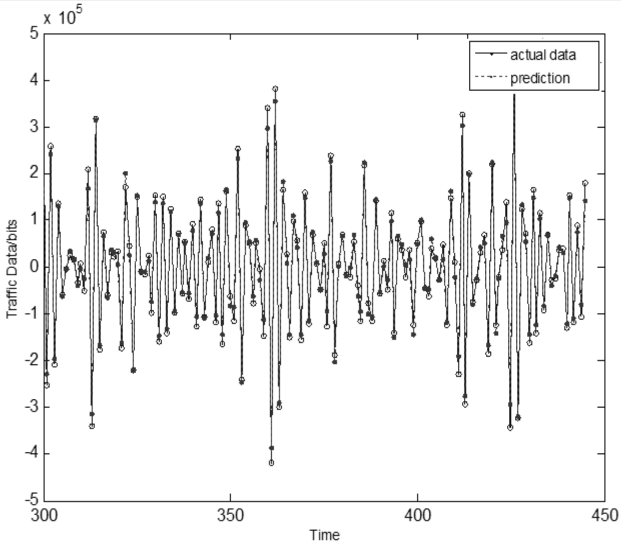

Fig. 5. IMF1 5th order difference and its prediction

Table 1. NMSE of IMF1 after 0-5th order difference

Order	0	1	2	3	4	5
NMSE	0.9958	0.3951	0.1543	0.0612	0.0191	0.0136

5 Conclusion

In this study, for the problem of the high complexity and the poor prediction accuracy of LRD traffic prediction, we research and propose one method which could be used in predicting LRD traffic data. The theoretical results indicate that LRD process decomposed by the EMD performs a SRD characteristic in each IMF component. Thus, some LRD model can be replaced by the simple models, such as ARMA, which could reduces the complexity of the model. The simulation results showed that IMF2~IMF8 components prediction had higher accuracy than IMF1 component and choosing the appropriate difference order could eliminate the non-stationary characteristic of IMF1 component. Then the accuracy of IMF1 component prediction was significantly improved.

Acknowledgments. This research was supported in part by the National Nature Science Foundation of China (NFSC) under grant No. 60672150 and 60702034.

References

1. Ji, Q.J.: Can Multifractal Traffic Burstiness be Approximated by Markov Modulated Poisson Processes? In: Proceedings 12th IEEE International Conference on Networks, pp. 26–30 (2004)
2. Shahram, S.H., Tho, L.N.: MMPP Modeling of Aggregated ATM Traffic. Electrical and Computer Engineering. In: IEEE Canadian Conference, Canada, pp. 129–132 (1998)
3. Zhou, B.X., Yao, Z.Q.: A Method to Stabilize Network Traffic. Journal on Communications 25(8), 14–23 (2004)
4. Rosario, G.G., Stefano, G., Marco, I., Michele, P.: On the Implications of the OFF Periods Distribution in Two-State Traffic Models. IEEE Communications Letters 3, 220–222 (1999)
5. Liu, J.K., Shu, Y.T., Zhang, L.F., et al.: Traffic Modeling Based on FARIMA Models. Electrical and Computer Engineering. In: IEEE Canadian Conference, Canada, pp. 162–167 (1999)
6. Rudolf, H.R., Matthew, S.C., Vinay, J.R., Richard, G.B.: A Multifractal Wavelet Model with Application to Network Traffic. IEEE Transactions on Information Theory 45, 992–1018 (1999)
7. Qigang, Z., Xuming, F., Qunzhan, L., Zhengyou, H.: WNN-Based NGN Traffic Prediction. In: Proceedings Autonomous Decentralized Systems, pp. 230–234 (2005)
8. Ahmad, B., Mohammad, F.A.: Generalized Wavelet Neuro-Fuzzy Model and its Application in Time Series Forecasting. In: International Symposium on Evolving Fuzzy Systems, pp. 253–258 (2006)
9. Mikio, H., Gang, W., Mitsuhiko, M.: Applications of Nonlinear Prediction Methods to the Internet Traffic. In: Proceedings IEEE International Symposium on Circuits and Systems, pp. 169–172 (2001)
10. Bo, G., Qinyu, Z., Yongsheng, L., Naitong, Z.: One Method from LRD to SRD. In: IEEE Wireless Communications, Networking and Mobile Computing (WiCOM 2009), Beijing, China, pp. 1–4 (2009)
11. Huang, N.E., Shen, Z., Long, S.R.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. Proc. Royal. Soc. London A 454, 903–995 (1998)
12. Bellcore Lab, `http://ita.ee.lbl.gov/html/traces.html`
13. Akaike, H.: A New Look at the Statistical Identification Model. IEEE Trans. on Automatic Control 19(6), 716–723 (1974)
14. Liu, J.K., Wang, G.S.: Applied Stochastic Processes. Science Press, Beijing (2004)