Study on Network Traffic Prediction Techniques

Huifang Feng and Yantai Shu
Department of Computer Science, Tianjin University,
Tianjin 300072, China
fenghuifang2003@hotmail.com, fenghuifang2003@163.com

***Abstract* - we briefly describe a number of traffic predictors (such as ARIMA, FARIMA, ANN and wavelet-based predictors) and analyze their computational complexity. We compare their performance with MSE, NMSE and computational complexity by simulating the predictors on four wireless network traffic traces and decide the most suitable network traffic predictor based on acceptable performance and accuracy.**

Keywords: Network traffic, prediction, ARIMA, FARIMA, ANN, wavelet

I. INTRODUCTION

The predictability of network traffic is of significant interest in many domains, including adaptive applications, congestion control, admission control and network bandwidth allocation. Traffic prediction requires accurate traffic models which can capture the statistical characteristics of actual traffic. Recently, there has been a significant change in the understanding of network traffic. A number of studies have shown that long-range dependence (LRD) is an omnipresent property of several classes of network traffic [1][2]. Predicting the future traffic level from past observations is an important component to affecting traffic control under self-similar traffic conditions. Effective optimal prediction remains a technical challenge [3]. With long-range dependent network traffic and their slow convergence properties, it becomes difficult to devise effective predictors that can rigorously be shown to have desirable properties. The choice of a prediction method is a tradeoff between the prediction interval, prediction error and computational cost.

Lately, there has been some work on network traffic prediction by modeling techniques. A. Sang analyzed the theoretically available optimal predictors for multi-step prediction of network traffic using Auto-Regressive Moving Average (ARMA) and Markov Modulated Poisson Process (MMPP) models [4]. Their results also showed that traffic low-pass filtering and aggregating could bring better predictability. Y. Qiao et al. empirically studied the one-step-ahead predictability of network traffic at different resolutions [5]. Y. Liang proposed a multiresolution learning NN traffic predictor to predict real-world VBR video traffic [6]. Some researcher adopted neural network methodology to predict VBR traffic [7]. Ref. [8] proposed an adaptive wavelet predictor for dynamic bandwidth allocation for VBR video traffic.

Several prediction models have been proposed in the literature dealing with various network traffic forecasting, but it remains unclear which predictors would provide the desired performance while staying simple, adaptive, and accurate. In this paper, we provide an in-depth study on the performance of a wide range of parametric predictors (i.e. Auto-Regressive Integrated Moving Average (ARIMA) and fractional ARIMA (FARIMA) predictors) and nonparametric predictors (i.e. artificial neural network (ANN) and wavelet-based predictors) to ascertain the ideal network traffic predictor.

The remainder of this paper is organized as follows. Section II describes network traffic predictors. We also give performance metrics, MSE and NMSE. Experimental results and comparative study with various predictors using various network traffic traces are presented in Section III. Finally, Section IV concludes the paper.

II. PREDICTORS AND PERFORMANCE METRICS

In this section we give a brief introduction to various network traffic predictors, such as ARIMA, FARIMA, ANN and wavelet-based predictors. We also give performance metrics, MSE and NMSE

A. Forecasting with ARIMA and FARIMA Models

ARIMA time series models form a general class of linear models that are widely used in modeling and forecasting time series. We summarize the mathematical properties of ARIMA processes here in order to introduce the notations used in the remainder of the paper. Let $\{X_t : t = ..., -1, 0, 1, ...\}$ be a time-series and B be the backward-shift operator, i.e. $BX_t = X_{t-1}$. Then, we let $\nabla = 1 - B$ be the differencing operator and ∇^d is called the differencing operator defined in the usual binomial expansion, i.e.

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \tag{1}$$

An ARIMA(*p,d,q*) process is a stochastic time-series process where d is the level of differencing, p is the autoregressive order, and q is the moving average order [9]. All three parameters are non-negative integers. Then the ARIMA(*p,d,q*) process can be described by the following relationship:

$$\Phi(B)\nabla^d X_t = \Theta(B) a_t \tag{2}$$

where $\{a_t : t = ..., -1, 0, 1, ...\}$ is a white noise process $WN(0, \sigma^2)$ with zero mean and variance σ^2, and

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - ... - \Phi_p B^p \tag{3}$$

$$\Theta(B) = 1 - \Theta_1 B - \Theta_2 B^2 - ... - \Theta_q B^q \tag{4}$$

Both $\Phi(B)$ and $\Theta(B)$ are polynomials in complex variables with no common zeroes, and in addition $\Phi(B)$ has no zeroes in the unit disk $\{B : |B| \le 1\}$.

Although the ARIMA model handles non-stationary time series, it still is a short-range dependent process because of its exponential decaying correlation structure.

FARIMA(*p,d,q*) is a long-range dependent model that is the natural extension of the ARIAM model [10]. However, the parameter d in FARIMA(*p,d,q*) is little different from that in ARIMA(*p,d,q*). For the ARIMA model, d can only be integer, while for the FARIMA model $d \in (-0.5, 0.5)$. The parameter d can be calculated by $d = H - 1/2$, where H is the Hurst parameter.

Let $\{X_t\}$ be the causal invertible FARIMA(*p,d,q*) process., we can write

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \tag{5}$$

and

$$a_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \tag{6}$$

where $\sum_{j=0}^{\infty} \psi_j B^j = \Theta(B)\Phi^{-1}(B)\nabla^{-d}$ and $\sum_{j=0}^{\infty} \pi_j B^j = \Phi(B)\Theta^{-1}(B)\nabla^{d}$.

From the theorems on linear prediction [11], the *h*-step-ahead forecast $\hat{X}_t(h)$ of a FARIMA process is

$$\hat{X}_t(h) = \sum_{j=h}^{\infty} \psi_j a_{t+h-j} \tag{7}$$

We can obtain their minimum mean square error of the h-step forecasts

$$\hat{\sigma}_t^2(h) = E(X_{t+h} - \hat{X}_t(h))^2 = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 \tag{8}$$

The classical modeling and prediction approach using FARIMA model includes the following steps:

Step 1: Pre-process the measured traffic trace to get a zero-mean time series.

Step 2: Estimate the value of the Hurst parameter H, and get the value of the differencing parameter d. Parameter H can be obtained using the known Hurst parameter estimates methods such as variance-time plots, R/S analysis and periodogram-based method.

Step 3: Do fractional differencing on processed time series.

Step 4: Determine p and q using Akaike's Information Criterion (AIC).

Step 5: Estimate parameters using maximum likelihood estimation.

Step 6: Forecast to predict future values, based on the estimated parameters and the model structure.

For ARIMA predictor, the prediction process includes above all step except step 2. In addition, we should modify fractional differencing into integer differencing in step 3.

The computational complexity for FARIMA and ARIMA predictors are $O(N^2)$ [12]

B. Forecasting with ANN Model

ANN is a non-linear, non-parametric and data driven modeling approach. It allows one to fully utilize the data and let the data determine the structure and parameters of a model without any restrictive parametric modeling assumptions. Since back-propagation (BP) algorithm is very common, the details are available in the references, hence these are not repeated here. The procedures for developing the neural network BP model are as follows [13]:

Step 1: Normalize the learning set.

Step 2: Decide the architecture and parameters: i.e., learning rate, momentum, and architecture.

Step 3: Initialize all weights randomly.

Step 4: Training, where the stopping criterion is either the number of iterations reached or when the total sum of squares of error is lower than a pre-determined value.

Step 5: Choose the network with the minimum error.

Step 6: Forecast future outcome.

When training a multi-layer perceptron, there are several factors that determine the computational complexity. It is assumed that the network is fully connected. The variable *i, n, o* represents the number of input nodes, hidden nodes and output nodes respectively. The total order of complexity is $O(i*o*n + n*o)$ or $O(n * o * (i+1))$ for training a single epoch [14].

C. Forecasting with Wavelet-Based Method

Wavelet analysis has become a research hot point. Wavelet analysis has good time and frequency multi-resolution, and can effectively diagnose signal's main frequency component and abstract local information of the time series. The researches and applications of wavelet analysis have already applied in computer network [5] [15].

The basic idea of wavelet-based methods for prediction is to first decompose the original signal into components and then applies some predicting method to theses individual components. High frequency components can be used to predict the near future while low frequency components can usually tell the long-term trend. In other words, the lower level of the decomposition can capture the long-range dependencies, while the higher levels capture the usual short-term dependencies. The working of this algorithm is described below:

Step 1: Choose number of data points such that we can maximize 2^L within our time series interval.

Step 2: Do the wavelet transform for time series, and then obtain the approximation coefficients and detail coefficients.

Step 3: Reconstruct the approximations and details themselves from their coefficients.

Step 4: Perform prediction for every reconstructed sub-time series on each resolution.

Step 5: Combine the coefficients predicted at each level. We can obtain the final prediction for the original data

The algorithm, which computes the orthogonal wavelet coefficients of a time-series of length *N*, is very fast and its complexity is of the order of *N*, e.g. *O(N)*[16].

D. Performance Metrics

To quantitatively assess forecasting performance, the mean square error (MSE) and normalized mean squared error (NMSE) is used measurements in prediction scheme [17].

1) MSE:

$$MSE = \frac{1}{M}\sum_{t=1}^{M}\left(X_t - \hat{X}_t\right)^2 \tag{9}$$

where X_t is the observed value of the time series at time t, $\hat{X}_t$ is the predicted value of X_t, and M is the total number of the predicted value. As the prediction accuracy increases, the MSE becomes smaller.

2) NMSE:

$$NMSE = \frac{1}{\sigma^2}\frac{1}{M}\sum_{t=1}^{M}\left(X_t - \hat{X}_t\right)^2 \tag{10}$$

where σ^2 is the variance of the time series over the prediction duration. The NMSE is widely used for evaluating prediction performance. It can be seen that, for a perfect predictor, $NMSE = 0$, and for a trivial predictor (one which statistically predicts the mean of the actual time series), $NMSE = 1$. If $NMSE > 1$, it means that the prediction performance is worse than that of trivial predictor.

III. EXPERIMENTAL RESULTS

In the following, the performance of the various models is evaluated by three actual wireless traces and one synthetic trace. Two traces (t030801 and t040318) were collected from our WiFi testbed in the Network Research Laboratory of Tianjin University. The second one (t040318) is a video traffic trace. Another one (final.anon) was from the Mobile Computing Group at Stanford University [18]. Ref. [19] describes their network topology, tracing methodology, and the characteristics of the trace of wireless traffic in detail. We also conduct the experiment on a synthetic trace using Chaotic Map [20].

We do the "minimum mean square error forecast" experiments on four traffic traces according to ARIMA and FARIMA prediction algorithm. Our ANN uses the three layers: one input layer, one hidden layer and one output. In our experiments, the numbers of nodes in the hidden layer were varied between 2 and 5. The best performance was found with 2 neurons in the hidden layer. For wavelet-based prediction, we decompose all the traffic traces using the Daubechies (db7) wavelet transform into three resolution levels as the residual series become quite smooth. As the residual series is very smooth, a simple Auto-Regressive (AR) model shows quite satisfactory prediction performance. We perform AR prediction on the each scale coefficients. Due to space constraints, we only show the performance metrics for predictors in Table I.

TABLE I: PERFORMANCE METRICS OF THE SINGLE-STEP PREDICTION FOR VARIOUS TRACES

Trace Name	ARIMA		FARIMA		ANN		Wavelet	
	MSE	*NMSE*	*MSE*	*NMSE*	*MSE*	*NMSE*	*MSE*	*NMSE*
t040318	47.34	0.31	42.99	0.27	43.79	0.28	50.91	0.32
t030801	5525	0.83	5396.	0.80	5412	0.81	5618	0.84
final.anon	1752.	0.80	1587	0.72	1534	0.70	1648	0.75
synthetic	7.28	0.57	7.09	0.55	7.01	0.55	6.92	0.54

From the statistics summarized in Tables I, we can obviously observe that:

1) The FARIMA and ANN prediction methods show quite similar results on the same testing set. But a close inspection reveals a better accuracy resulting from FARIMA predictor, with regards to the MSE and NMSE. Since FARIMA is a self-similar model with the capability to capture both the short-range dependent and long-range dependent characteristics, but at the cost of computational complexity.
2) Compared with other predictor, ARIMA model delivers worse forecasting performance. Although ARIMA models are proved to be quite powerful to model a class of non-stationary data, it can't capture long-range dependent characteristics.
3) The wavelet forecasting can't give remarkable performance as in [5]. This is due to the effect of boundary conditions in wavelet transform. We will discuss in the following paragraphs.

Using the t030801 traffic trace, we further compare the computational time of the four forecasting methods. Our experiments show that wavelet-based predictor is prior to all others. It takes about 3.1 seconds of execution to predict 100 values, compared to 5.3, 22 and 1500 seconds for ANN, ARIMA and FARIMA predictors respectively. We can obtain the same results from other traffic trace. FARIMA predictor has the longest computational time.

From these evaluations, we can conclude that ANN predictor generally show better profitability than other predictors. It delivers an acceptable performance in terms of simplicity, adaptability and accuracy, at least in our experiments. It is clear that ANN is the most suitable online predictor that can be used in dynamic bandwidth allocation, congestion control, etc. FARIMA is the best offline predictor that can use in the network design.

We discuss the effect of boundary condition in wavelet-based predictor. Application of discrete wavelet transformation (DWT) to finite-length time series brings up the crucial issue of wavelets affected by the boundary. That is, the wavelet transform is based on filtering a time series, and there must be a method for computing the remaining wavelet coefficients over the forecast period, when one end of the vector is encountered. In order to study the importance of boundary conditions in wavelet transform in the context of time series forecasting, we apply the Daubechies wavelet transform on the network traffic using smooth-padding of order 0 (Sp0, constant extension at the edges), smooth-padding of order 1 (Sp1, first derivative interpolation at the edges), symmetric boundary condition (Sym, boundary value symmetric replication), and periodic boundary conditions (Ppd, periodic extension at the edges) respectively. We also perform wavelet transform by taking advantage of the real values in the future from the traffic traces, which we call the "Extensive" boundary condition.

TABLE II: PERFORMANCE METRICS OF THE SINGLE-STEP PREDICTION FOR VARIOUS BOUNDARY CONDITIONS

Trace Name	Sp0		Sp1		Sym		Ppd		"Extensive" boundary	
	MSE	*NMSE*	*MSE*	*NMSE*	*MSE*	*NMSE*	*MSE*	*NMSE*	*MSE*	*NMSE*
t040318	50.91	0.32	103.65	0.67	525.16	3.40	57.69	0.37	2.81	0.02
t030801	5618	0.84	14191	2.13	26681	3.96	5801	0.87	401.8	0.06
final.anon	1648	0.75	3938	1.80	2350. 2	1.08	1749	0.80	89.56	0.04
synthemic	6.92	0.54	19.08	1.48	275.64	21.46	7.61	0.59	1.12	0.09

In Table II we use all traffic traces to compare the five boundary conditions with regard to MSE and NMSE. The results show that, compared with other boundary conditions, the "extensive" boundary condition is the best in one step ahead forecasting. But this boundary condition treatment is unrealistic, because the real values in the future are not known at the time we make prediction. We think that the method, which using wavelet-based prediction with this boundary condition to forecast network traffic, is incorrect [5]. The wavelet-based prediction with smooth-padding of order 0 boundary condition provide more accuracy than that of other boundary extension except "extensive" boundary. That is to say, compared with other boundary conditions, smooth-padding of order 0 boundary condition is more suitable for characteristics of network traffic.

Although it is of worse prediction performance compared to other predictors, wavelet-based prediction is attracted by its lower computational complexity. We will study the more appropriate boundary conditions that can deal with long-range dependent network traffic and improve the accuracy of traffic prediction.

IV. CONCLUSIONS

In this paper, we introduced the various popular network traffic predictors, such as ARIMA, FARIMA, ANN and wavelet-based predictors. We compared their performance with MSE, NMSE and computational complexity by experiments on four wireless traffic traces and determined the most suitable online predictor based on acceptable performance and accuracy. Our results showed significant advantages for the ANN technique. FARIMA could provide the accuracy prediction, but at the cost of computational complexity. Compared to other predictor, ARIMA predictor performed considerably worse. In respect of wavelet-based predictor, our study contradicted earlier work in that the wavelet forecasting couldn't give remarkable performance. This is due to the effect of boundary conditions in wavelet transform. But it was attracted by its lower computational complexity. Maybe we can find more appropriate boundary conditions that can deal with long-range dependent and improve the accuracy of traffic prediction. This is left for further research.

ACKNOWLEDGMENT

This research was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 60472078, by a grant from the Cisco University Research Program Fund at Community Foundation Silicon Valley, by the ZhongXing Telecommunication Equipment Corporation (ZTE).

REFERENCES

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, 1994, pp. 1–15.

[2] M. Jiang, M. Nikolic, S. Hardy, and L. Trajkovic, "Impact of self-similarity on wireless data Network performance," in *proc. ICC01*, pp. 477-481, Finland, June 11-14, 2001

[3] J. Beran. Statistics for long-Memory Processes, Monographs on Statistics and Applied Probability. Chapman and Hall, New York, 1994.

[4] A. Sang and S.-Q. Li, "A predictability analysis of network traffic," *Computer Networks*, vol. 39, no. 4, 2002, pp. 329-345.

[5] Y. Qiao; J. Skicewicz, P. Dinda, "An Empirical Study of the Multiscale Predictability of Network Traffic," in *proc. HPDC'04,* pp. 66- 76.

[6] Y. Liang, "Real-Time VBR Video Traffic Prediction for Dynamic Bandwidth Allocation," *IEEE Transactions on Systems, Man, and Cybernetics,* Part C vol. 34,no.1, 2004, pp.32-47.

[7] W. M. Moh, M. J. Chen, N. M. Chu et al., "Traffic prediction and dynamic bandwidth allocation over ATM: a neural network approach," *Computer Communications*, vol.18, no.8, 1995, pp.563-571.

[8] A. D. Doulamis, N. D. Doulamis, and S. D. Kollias, "An Adaptable Neural-Network Model for Recursive Nonlinear Traffic Prediction and Modeling of MPEG Video Sources," *IEEE Transactions on Neural* Networks*,* vol. 14, no. 1, JANUARY 2003, pp.150-166.

[9] G. E. P. Box, and G. M Jenkins, G. C. Reinsel, Time series analysis: Forecasting and control, 2nd ed. Prentice-Hall, 1994.

[10] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol.83, no.1, 1981, pp.165-175.

[11] P. Brockwell, R. Davis, Time series: theory and methods, 2nd ed. New York, Springer Verlag, 1991

[12] M. Krunz and A. Makowski, "A source model for VBR video Traffic Based on M/G/∞ Input processes," in *proc. INFOCOM* '98, pp. 1441-1449.,1998

[13] C.T. Su, L.I. Tong, C.M. Leou, "Combination of time series and neural network for reliability forecasting modeling," *J. Chin. Inst. Ind. Eng.* vol.14, no.4, 1997, pp. 419–429.

[14] E. Istook, T. Martinez, "Improved backpropagation learning in neural networks with windowed momentum," *International Journal of Neural Systems*, vol. 12, no.3&4, 2002, pp. 303-318.

[15] S. Ma, and C. Ji, "Modeling heterogeneous network traffic in wavelet domain," *IEEE/ACM Transaction on Networking,* vol. 9, no. 5, 2001,pp.634-649.

[16] I. Daubechies. Ten Lectures on Wavelets. Philadelphis: SIAM, 1992.

[17] A. S. Weigend and N. A. Gershenfeld, Eds., Time Series Prediction: Forecasting the Future and Understanding the Past. Reading, MA: Addison-Wesley, 1994.

[18] http://mosquitonet.stanford.edu/software.html

[19] D. Tang and M. Baker, "Analysis of a local-area wireless network," in *proc. MOBICOM'00,* Boston, USA, August 2000, pp.1-10.

[20] R. J. Mondragón, D. K. Arrowsmith, J. M. Pitts, "Chaotic maps for traffic modelling and queueing performance analysis," *Performance Evaluation,* vol.43, 2001, pp. 223–240.