Traffic Predictability based on ARIMA/GARCH Model

Bo Zhou, Dan He and Zhili Sun
CCSR, University of Surrey,
Guildford, Surrey, United Kingdom
{B.Zhou, D.He, Z.Sun}@surrey.ac.uk

Abstract-- **The predictability of Internet traffic is a significant interest in many domains such as adaptive applications, congestion control, admission control, and network management. In this paper, we propose a new traffic prediction model called Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA/GARCH), which can capture traffic burstiness and exhibit self-similarity and long-range dependence (LRD). We discuss network traffic predictability related to different prediction applications and measure methods. We validate our prediction model by comparing with other models, includes non-model-based Minimum Mean Square Average (MMSE), pure self-similar Fractional ARIMA (FARIMA). We use the real network traces to evaluate models. The results show that MMSE computation is simplest and fastest and can apply for online prediction applications. The results also show that FARIMA predictability relies on strong degree of self-similarity, our proposed ARIMA/GARCH model get the best adaptability and accuracy. Therefore ARIMA/GARCH model can be used for exact prediction applications.**

Index terms-- **traffic prediction, traffic predictability, ARIMA, GARCH**

A. INTRODUCTION

Internet traffic prediction is of significant interests in many domains, including adaptive applications [1], congestion control [2], admission control [9], and network management [3]. The basic idea of traffic prediction is to predict traffic in the next control time interval based on the online (or offline) measurement of traffic characteristics. The goal is to forecast future variations as precisely as possible but not too complicated, based on the measured traffic history.

The most significant question of traffic prediction is traffic predictability. Traffic predictability denotes the possibility of prediction to satisfy precision requirements over desires prediction and control time interval. On one hand, a large prediction interval is needed to provide sufficient time for control actions and to offset the inevitable delays in the network. On the other hand, a small prediction error is desirable for the following reason: control actions based on erroneous prediction may inadvertently compromise the control performance. Precise prediction is preferred to overcome this problem. Unfortunately, prediction accuracy deteriorates quickly as the prediction interval increases. Clearly there is a tradeoff between a large prediction interval and a small prediction error, which reflects the tradeoff between the control time interval and the prediction accuracy.

Recently, there has been a significant change in the understanding of Internet traffic. It has been found in numerous studies that in high-speed Internet traffic exhibits self-similarity and long range dependence (LRD) [4][5][6]. Self-similarity cannot be captured by classical models. Hence new self-similar models have to be developed. The disadvantage of self-similar models is their computational complexity. And their fitting procedure is very time consuming while their parameters cannot be estimated on the short-term measurement. Hence the new fundamental problem in the traffic prediction is to find efficient self-similar models to predict future traffic variations precisely with good predictability.

In this paper, we proposed a new non-linear time series model to solve the self-similar traffic prediction problem. This model makes use of upon modeling changes in variance, called the Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA/GARCH). We compared our proposed model behaviors with another two significant traffic prediction models: Fractional Autoregressive Integrated Moving Average (FARIMA) [5][9][10] and Minimum Mean Square Error (MMSE) [13]. We analyze the quantized one-step traffic prediction, given a complete traffic history and its stationary prediction model as a priori knowledge. Thus, the result provides prediction performance and related predicted error with expected prediction interval. It reveals the promising predictability of network traffic by specified prediction model. To evaluate the best predictability, we take into account such factors as the network control time interval, resource utilization target, traffic statistics and the measurement time-scale, as traffic predictability depends on how the prediction results are used to meet the control expectation or constraint.

The main contributions of this paper can be summarized here. First, we proposed a new traffic prediction model, which can capture traffic burstiness and behave self-similarity and LRD. Second, we analyze predictability the most important and complexity question related to different applications and measure methods. Third, we validate our prediction model by comparing with the other models, includes non-model-based MMSE, pure self-similar FARIMA. We use the real network traces to evaluate models. The results show that MMSE computation is simplest and fastest and can apply for online prediction applications. The results also show that FARIMA predictability relies on strong degree of self-similarity and the proposed ARIMA/GARCH model get the best adaptability and accuracy. Therefore ARIMA/GARCH model can be used for exact prediction applications.

The rest of this paper is organized as follows. In Section B, we motivated the problem and describe related work. New ARIMA/GARCH model are proposed in Section C. Section D defined the traffic predictability validation methods. Our performance analysis is presented in Section E. We conclude our work in Section F.

B. BACKGROUND

By looking back on the development of network traffic prediction techniques, we find that most techniques have introduced black-box modeling and structural modeling to solve prediction problem. The first generation of techniques introduced linear time series models [7][8], their approaches are based on the traditional time series prediction technique, which we called Box-Jenkins approach. The basic idea of this approach is to provide a broad class of models with enough parameters to fit a variety of data sets. The emphasis is on finding something that fits according to some criterion. There are a lot of classical linear predictive models are applied to predict network traffic, such as Auto Regressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA). All these models are linear, and behave short memory.

Currently self-similarity and LRD have been observed in the network and Internet traffic [4][5][6]. However, traditional linear time series models cannot explain and capture self-similarity and LRD. Shortly, the second generation of techniques emerged that were able to capture LRD in time series as well as its non-linearity. They suggested that using Fractional ARIMA (FARIMA) models might be appropriate [5]. Thereafter FARIMA model has been shown its prediction ability in admission control [9] and dynamic bandwidth allocation [10]. Another non-linear Threshold Autoregressive (TAR) model was also introduced to model and predict network traffic [11]. Both FARIMA and TAR models are non-linear and behave long memory, they can capture self-similarity and LRD well. However both these two models are computational complexity. On the other hand, traffic predictability has been introduced and assessed by ARMA and Markov-modulated Poisson process (MMPP) models [12]. Their analysis and empirical study found that both aggregation and smoothing monotonically increased predictability. Unfortunately, ARMA and MMPP models can behave self-similarity slightly and such models cannot explain self-similar model predictability.

Determining predictability and improving prediction accuracy are the major motivation of the third generation of prediction techniques. Minimum Mean Square Error (MMSE) predictor is the simplest and fastest traffic prediction model both in theory and in practice [13]. MMSE can be used for on-line prediction applications based on its simple and fast predict characteristic. However, MMSE cannot exhibit self-similarity and LRD, and cannot be an exact model to generate synthesis traffic. Burstiness or heavy-tailed is the major reason caused network traffic self-similarity and LRD. Hence multifractal has been introduced to explain these characteristics in small time scale (microsecond or less) [14]. In small time scale, the multifractal characteristics affect TCP protocol performance, and explain why traffic behaved self-similarity (mono-fractal) and LRD in large time scale. Multifractal Wavelet Model (MWM) is a significant traffic model which is able to capture multifractal characteristic in small time scale but cannot predict traffic performance as time series models.

C. ARIMA/GARCH PREDICTION MODEL

1) Model description

ARIMA processes are the natural generalizations of standard ARMA(r,s) processes when the degree of differencing d is taking integral values [17]. ARIMA(r,d,s) process $\{X_t, t=1,2,\ldots,\}$ is defined to be

$$\phi(B)\Delta^d X_t = \theta(B)Z_t \tag{1}$$

where

$$\phi(B) = 1-\phi_1 B-\ldots-\phi_r B^r,\quad \theta(B) = 1-\theta_1 B-\ldots-\theta_s B^s .$$

$\phi(B),\theta(B)$ have no common zeroes, while r and s are non-negative integers. B is the backward-shift operator, i.e. $BX_t = X_{t-1}$. $\Delta^d = (1-B)^d$ is the differencing operator.

GARCH (p,q) processes are the explanation processes which process conditional variance of Z_t in (1) [15][16], it is define to be

$$\sigma_z^2 = \alpha_0 + \sum_{i=1}^{p}\gamma_i\sigma_{z-i}^2 + \sum_{j=1}^{q}\alpha_j Z_{t-j}^2 \tag{2}$$

$$\alpha_0 > 0, \alpha_j \ge 0, \gamma_i \ge 0$$

where σ_z^2 is the variance of Z_t and (2) has its constraint as

$$\sum_{i=1}^{p}\gamma_i + \sum_{j=1}^{q}\alpha_j < 1 \tag{3}$$

The key insight lies in the GARCH processes is the conditional variance of Z_t, where its variance is local and changing over time. The term conditional implies explicit dependence on a past sequence of observations, and the term unconditional is more concerned with long-term behavior of a time series and assumes no explicit knowledge of the past. Conditional variance property of GARCH processes can capture and explain burstiness characteristic of time series. As we know, burstiness characteristic is the major reason caused network traffic behaved self-similarity and long range dependence (LRD). Hence, ARIMA/GARCH processes are flexible with regard to the modeling of the self-similarity and LRD behaviors of network traffic.

2) Modelling a trace

The core of traffic prediction is to fit an appropriate model to the actual network traffic. Hence, we propose the following procedure to fit a non-linear ARIMA(r,d,s)/GARCH(p,q) model to the traffic trace. Explanation is provided where necessary.

Step 1: Pre-processing the measured traffic trace to get a zero-mean time series X_t, t=1,2,…

Step 2: Using differencing operator difference the time series until it appears to come from a stationary process. It is often found that that first order (d=1) differencing of non-seasonal data is adequate.

Step 3: Determining the orders of r and s using Autocorrelation function (ACF) and its partial function (PACF). Normally, the theoretical ACF of a MA(s) has a simple form in that it "cuts off" at lag s and the PACF of an AR(r) has a simple form in that it cuts off" at lag r.

Step 4: Estimating parameters of $\phi_1,\phi_2,\ldots\phi_r$ and $\theta_1,\theta_2,\ldots\theta_s$. Values of these parameters can be estimated by using maximum likelihood estimation (MLE). After parameter estimation, the best values of parameters can be determined by the Akaike's Information Criterion (AIC) [17].

Step 5: Determining the orders of p and q. Identifying order of p and q are not easy not only in theory but also in practice, many analysts assume both orders setup one as the standard order.

Step 6: Estimating parameters of p and q. We assume that the sum of γ_p and α_q is close to 1 based on GARCH constraint (3). Specifically, we setup GARCH(1,1) as the "standard" model depend on step 5, so initial estimation are

$$\sigma_z^2 = \alpha_0 + \gamma_1\sigma_{z-1}^2 + \alpha_1 Z_{t-1}^2, \quad \gamma_1 + \alpha_1 < 1 \tag{4}$$

Then we estimate constant α_0 of the conditional variance by estimating the unconditional variance at first,

$$\sigma_z^2 = \frac{1}{T}\sum_{t=1}^{T} Z_t^2 = Z_t^2 \tag{5}$$

The (4) can be extended as,

$$\alpha_0 = \sigma_z^2(1-\gamma_1-\alpha_1) \tag{6}$$

After determining the initial constant α_0, we can estimate parameters γ_1 and α_1 by MLE and select best values by AIC.

Finally, we get the fitted ARIMA(r,d,s)/GARCH(p,q) model as form (1) and (2).

3) Forecasting a trace

Assumptions of causality and invertibility allow us to write:

$$Z_t = \sum_{u=0}^{\infty}\pi_u X_{t-u} \tag{7}$$

$$where \quad \sum_{u=0}^{\infty}\pi_u B^u = \phi(B)\theta^{-1}(B)\Delta^d$$

and

$$\begin{gathered} Z_t = \sum_{v=0}^{\infty}\psi_v^2\sigma_{t-v}^2 \\ where \quad \sum_{v=0}^{\infty}\psi_v B^v = (1-\sum_{i=1}^{p}\gamma_i)(1-\sum_{j=1}^{q}\alpha_j)^{-1} \end{gathered} \tag{8}$$

We use minimum mean square error (minimum MSE) study ARIMA/GARCH models prediction. Let $\widehat{X}_{t+1}$ denote the one-step prediction made at origin t of X_{t+1} at future time t+1. Then we modify (7) and (8) and have:

$$\begin{gathered} \widehat{X}_{t+1} = \sum_{u=1}^{\infty}\pi_u X_{t-u+1} \\ \widehat{\sigma}_{Z+1}^2 = \sum_{v=1}^{\infty}\psi_v^2\sigma_{t-v+1}^2 \end{gathered} \tag{9}$$

where $\widehat{\sigma}_{Z+1}^2$ denote the one-step prediction of variance of Z_{t+1}. And the MSE of the one-step prediction is

$$MSE = E((X_{t+1} - \widehat{X}_{t+1})^2) \tag{10}$$

Hence, from a fitted ARIMA/GARCH model of a time series, one may obtain its one-step prediction by (9) and the MSE by (10).

When the decisions of network control are based upon traffic prediction, some packets can be lost in the network if a predicted traffic value is less than the actual value. Therefore, to specify the accuracy of traffic prediction in networks, we only need to calculate the upper probability limit of a forecast, as opposed to providing the upper- and lower-limits as done in normal prediction technique. Hence we propose an adapted prediction method to provide upper probability limit.

Our algorithm determines the adaptive one-step prediction $\widehat{X}_{t+1}$ by adding a bias ξ_w to the predicted value of minimum MSE prediction. Mathematically,

$$\widehat{X}_{t+1}^{w} = \widehat{X}_{t+1} + \xi_w \quad \textbf{(11)}$$
$$where \quad P[e_t \le \xi_w] = w, \quad 0.5 \le w < 1$$

e_t is the prediction errors, and w is the upper probability limit. Hence, w is the probability that predicted value is more than the observed value. Assuming that the prediction error e_t has a Normal distribution, we can obtain the relationship between ξ_w and w. Obviously, $\xi_w = 0$ when $w = 0.5$.

We propose a four-step procedure to predict network traffic as following:

Step 1: Building ARIMA/GARCH model to describe the traffic.

Step 2: Doing minimum MSE prediction.

Step 3: Determining the value of upper probability limit w according to the QoS necessary in the particular network.

Step 4: Doing traffic predictions by the adapted prediction method with upper probability limit.

D. PREDICTABILITY DEFINITION

As description in introduction section, predictability is the most important prediction issue we try to discuss in this section. We describe three major predictability measure methods in this section. After then we redefine predictability in theory considered different network applications.

Mean Square Error (MSE) and its normalized function (NMSE) is the first method we introduce to measure model predictability. MSE is a popular predictability not only in theory but also in practice [17]. MSE is a measure of the absolute error. This measure suffers from the problem that the amplitude of the signal to be predicted plays a strong role in the size of the measurement error. To avoid this problem, a relative error measurement is considered. A typical approach is to normalize the MSE relative to the variance of the time series to be predicted. The result is called the normalized mean square error (NMSE), its mathematic expression can be extended from (10),

$$NMSE = \frac{E((X_{t+1} - \widehat{X}_{t+1})^2)}{\sigma^2} \quad \textbf{(12)}$$

Because $\sigma^2 = E((X_{t+1} - \overline{X}_{t+1})^2)$, where $\overline{X}_{t+1}$ is the mean value of X_{t+1}. So (12) can be expressed as

$$NMSE = \frac{E((X_{t+1} - \widehat{X}_{t+1})^2)}{E((X_{t+1} - \overline{X}_{t+1})^2)} \quad \textbf{(13)}$$

From (13), it can be found the smaller the NMSE, the better the predictability.

Signal to Error Ratio (SER) is the second predictability measurement we describe. SER quantify the prediction quality by following expression,

$$SER = \frac{E(X_{t+1}^2)}{E((X_{t+1} - \widehat{X}_{t+1})^2)} \quad \textbf{(14)}$$

Normally, comparing with expression of NMSE, we compute SER^{-1}, its expression is extended from (14),

$$SER^{-1} = \frac{1}{SER} = \frac{E((X_{t+1} - \widehat{X}_{t+1})^2)}{E(X_{t+1}^2)} \quad \textbf{(15)}$$

From (15), it can be found the smaller the SER^{-1}, the better the predictability.

The SER^{-1} method is similar to NMSE. They both compute prediction error and normalized the error related to the characteristics of time series (mean or variance).

The final measurement is different from the above methods. This method can be described as following steps.

We assume the normalized one-step prediction error is

$$\overline{err} = \left| \frac{\widehat{X}_{t+1} - X_{t+1}}{X_{t+1}} \right| \quad \textbf{(16)}$$

This normalized error should exceed a percentage ε (e.g. 10%) with a probability $P_{\overline{err}}(\varepsilon)$, where $P_{\overline{err}}(\varepsilon) = \Pr(\overline{err} > \varepsilon)$. We call $P_{\overline{err}}(\varepsilon)$ is Prediction Error Critical Probability (PECP). Obviously it found that the smaller the PECP, the better the predictability.

All the three above methods are widely used for predictability measure not only in theory but also in practice. However all of these methods have their common limitation: measuring prediction error only. Although prediction error is the most important predictability standard in theory, it is not enough in practice. Considered there are different kinds of applications of prediction, and each application has its prediction requirements. For example: on-line congestion controls pay more attentions on computation simple and fast than accuracy, or long-term network designs pay more attentions on prediction accuracy and adaptability than computation speed. Thus we have to decide traffic models predictability not only in measuring prediction error but also in others criteria: prediction computation complicacy, history length and prediction length, prediction adaptability.

Computation complicacy can be seen from prediction models themselves. In a prediction model, more parameters need to be estimated, therefore, it makes it more computationally complex. Otherwise, different parameters estimation methods also cause different computation speed. For example, an estimation of parameters by computing simple autocorrelation matrix in MMSE prediction model will be faster than estimation of those parameters by MLE in others long memory models.

History length and prediction length is another useful predictability measure method. If we want to select a prediction model for on-line prediction application, we have to choose the model, which needs short traffic history to estimate parameters. In contrast, we have to choose long traffic history to model and predict long-term traffic and wish our selected model could behave traffic self-similarity and LRD characteristics.

A good prediction model should be adapted to changing traffic, which we call this adaptability. As time progresses, more traffic information are available and more characteristics are known. We should select the adaptive traffic prediction model, which can improve and update its parameters to capture new available information and behave new characteristics.

E. PERFORMANCE ANALYSIS

We demonstrate the performance of our proposed ARIMA/GARCH modeling approach when applied to the real network trace. Then we validate the prediction performance of our proposed models comparing with classical models: MMSE and FARIMA. We discuss how to address predictability in different applications as well.

1) Real network traces

Our study is based on two of real network trace. The NLANR set consists of short period packet header traces chosen at random from among those collected by the Passive Measurement and Analysis (PMA) project at the National Laboratory for Applied Network Research (NLANR) [18]. The PMA project consists of monitors located at aggregation points within high performance network such as Abilene. Each of the traces is approximately about 90 seconds long and consists of IP packet headers from a particular interface at a particular PMA sites. We chose 35 continued NLANR traces provided by one PMA site (names PUR). The traces were collected in the 1st May 2005. We connected continued NLANR traces together, aggregated trace on time scale 1 second, the unit is Mbits/second.

The AUCKLAND set also come from NLANR. These traces are IP packet header traces captured at the University of Auckland's Internet uplink between 22nd and 23rd April 2004. These also represent aggregated WAN traffic, but here the durations for most of the traces are on the order of an hour (3600 seconds). We chose 2 continued AUCKLAND traces, aggregated trace on time scale 1 second and unit is Mbits/second as well.

2) ARIMA/GARCH Modelling

We validate our models performance and compare their characters with real traces characters. Because the real traces behaved LRD and self-similarity, we try to use variance-time plot [4], R/S estimation [4] and Wavelet estimation [19] to justify our modeling results.

Variance-time plot is comparing trace variance with its scale size. For a trace behaved SRD, its variance decreases as the scale size increase, and the variance decrease rate is linear depend on the scale size increase rate. In contrast, if the trace behaved LRD, the variance decreases slowly depend on the scale size increase rate with a constant slope. Here the slope indicates the LRD. Analysis results are shown in Figure 1(a).

R/S estimation captures Hurst parameter, which indicates self-similarity. The basic idea is based on rescaled adjusted range (R/S). We assume there is a reservoir with flexible input and allow withdraw output. We hope the reservoir never to be empty. Where R called the adjusted range, and it is the maximum capacity of input rate; S called the rescaled value, and it is the mean output rate. For the trace behaved self-similarity, R/S always increase depend on the scale size increase rate with the constant Hurst slope (Hurst parameter indicates the self-similarity when $0.5 < Hurst < 1$). Analysis results are shown in Figure 1(b).

Wavelet estimation estimate LRD parameter by wavelet methodology. The key of wavelet is using a "mother wavelet" to smooth and filter the trace, try to localize in both time and frequency domain. Comparing the covariance value after wavelet analysis with the scale of "mother wavelet" coefficient it used, we can find the slope value which indicates the LRD parameter. When applied the wavelet estimation to the increments of a self-similar trace, the estimate slope can indicate the Hurst value as well. Analysis results are shown in Figure 1(c).

The Hurst parameter of NLANR trace is observed as 0.7706, which is not exceeding 0.85. That means that the real trace does not exhibit strong LRD, and the fact that infinite history is not possible in practice. In contrast, the Hurst parameters of AUCKLAND trace are detected as 0.9016 and it means strong LRD. From above results, we can find our proposed ARIMA/GARCH models can behave the LRD and self-similarity characteristics same as the real network traces. ARIMA/GARCH model can explain the network traffic burstiness characteristic by its key characteristic: changing variance over time (conditional variance).

3) Predictability validation

We discuss traffic models' predictabilities, by comparing our proposed ARIMA/GARCH model predictability with classical MMSE model and FARIMA model. We try to find out suitable prediction models for varied prediction applications based on the methods described in section 4 to validate models' predictability.

Models computation speed is one of characteristics in predictability. We use simulation environment to validate the simulation speed. The computer for simulation is a Pentium III 850 MHz, 256MB memory computer. The software we do simulation is based on Matlab version 7.0 (R14) [20]. Three models' prediction computation speed is varied. The MMSE predictor is the fastest and the others two are slower than MMSE but little difference with each other. For example: for 1000 steps prediction, the MMSE predictor needs about 0.1 second to estimate results, the FARIMA predictor needs about 2 second to estimate results and our proposed ARIMA/GARCH predictor needs about 2 second to get results. Noticed the computation duration includes modeling and forecasting procedures.

Prediction history size is another characteristic in predictability. Long history size will cause good prediction results in theory. However, some applications only provide a short history of trace for modeling and this will affect some models' predictability. On the other hand, long history size will cause bad prediction results when traces behaved little LRD or varied LRD characteristic in the history scale. For all these reasons, we try to vary the prediction history size and predict traces in the long prediction duration (prediction length=1000), and validate the models predictability.

We measure the NMSE from the traffic prediction results at first. Compared the NMSE whole prediction procedure with both traces in Figure 2, we can see some interesting results. Figure 2(a) shows the NLANR prediction performance and Figure 2(b) shows the AUCLAND. For the MMSE prediction model, we can see its NMSE decreases when trace history size increases. This can be explained by basic prediction theory: more information will give better prediction. For FARIMA prediction model, we can see a clearly cutoff curve in its NMSE results. The best history size is 100 for NLANR and 10 for AUCKLAND. When the trace history becomes long, the FARIMA predictability becomes varied. Its predictability for NLANR becomes the worst but AUCKLAND becomes the best. This can be explained by FARIMA fractal characteristic. From Hurst parameter estimations, we can find both trace behave self-similarity and LRD. Strong LRD can be observed in AUCKLAND but not in NLANR. This characteristic causes that FARIMA get good prediction performance in AUCKLAND but not in NLANR. For ARIMA/GARCH prediction model, we also can find a cutoff curve in its NMSE results. The best history size is 10 for NLANR and 100 for AUCKLAND. When the trace history becomes long, the ARIMA/GARCH predictability becomes bad but normally best in three models. This can be explained by its conditional variance characteristic. Conditional variance can exhibit traffic burstiness, and burstiness is the major reason caused self-similarity and LRD. When trace history size increases, there are some burstiness cannot be model by GARCH, this causes the prediction performance decreases.

After measure predictability by NMSE, we use SER^{-1} to measure prediction performance. Compared the SER^{-1} whole prediction procedure with both traces in Figure 3, Figure 3(a) shows the measure results for NLANR trace and Figure 3(b) shows the measure results for AUCKLAND trace. Comparing Figure 3 with Figure 2, we can find their results are very similar.

Finally, to validate predictability by PECP measurement, we set the error exceed limit as 10%, that means we can measure the probability of error exceed 10%. Compared the PECP whole prediction procedure with both traces in Figure 4, it is shown the NLANR prediction results in Figure 4(a) and AUCKLAND prediction results in Figure 4(b). In Figure 4(a), it is clearly found that ARIMA/GARCH gets the best prediction performance, and MMSE and FARIMA both get similar prediction results. There is a significant cutoff point for FARIMA and ARIMA/GARCH, where their prediction performance will be the best in the whole scale. In Figure 4(b), we can find ARIMA/GARCH gets the best prediction performance, and FARIMA gets better prediction performance than MMSE as well. Clearly, there is also a significant cutoff point for FARIMA and ARIMA/GARCH, where their prediction performance will be the best.

To sum up above results, we find out some interesting points.

MMSE prediction model is a non-model based prediction approach. Its prediction performance will increase when the trace history size increases. Its predictability is the worst in three prediction models, but its computation is the simplest and fastest. So this model can be suitable for on-line (short-term) prediction applications.

FARIMA prediction model is a fractal based time series model. Its prediction performance is depend on the degree of traffic behaved self-similarity and LRD. The stronger the LRD, the better the prediction performance is. Its predictability does not rely on history size. There is always cutoff point in its whole prediction scale and the prediction performance will be the best in this point. Its computation is complex and speed is slow.

ARIMA/GARCH prediction model is our proposed prediction model in this paper. It is a conditional variance based time series model. Its prediction performance is the best in three prediction models and less predictability influence by self-similarity and LRD than FARIMA. And its predictability does not rely on history size. This characteristic is similar as FARIMA prediction model. Unfortunately, its computation is complex and speed is slow. Considered ARIMA/GARCH model get the best adaptability and accuracy, it can be suitable for off-line (long-term) prediction applications.

F. CONCLUSION

Network traffic prediction is of significant interests in many network applications. Self-similarity and LRD have been observed in network traffic, and these characteristics cannot be captured by the traditional traffic prediction models. So the problem in the network traffic prediction is to find efficient self-similar models to predict future traffic variations precisely, and get good predictability.

In this paper, we propose a new ARIMA/GARCH model in network traffic prediction. The basic characteristic of this model is conditional variance. This characteristic can capture the traffic burstiness and heavy-tailed in nature, and burstiness and heavy-tailed are the reasons caused self-similarity and LRD in theory. We investigate self-similarity and LRD captured by ARIMA/GARCH model and real network traffic, find our proposed model can capture all these characteristics not only in theory but also in practice.

The most import problem of traffic prediction is traffic predictability. Hence, we discuss our proposed ARIMA/GARCH model predictability, and compare its predictability with classical MMSE and FARIMA models. We use varied methods to measure models predictability. We find the MMSE computation is the simplest and fastest in three models but its prediction accuracy is lowest. We can apply MMSE model for online (short-term) traffic prediction application. We find the FARIMA computation is complex and its predictability relies on the degree of self-similarity, it can only model and predict traffic behaved strong self-similarity and LRD. Finally, we find that our proposed ARIMA/GARCH model get the best adaptability, it can model and predict traffic behaved different degree of self-similarity and LRD. And ARIMA/GARCH prediction accuracy is also the best in three models. Hence we suggest our proposed ARIMA/GARCH model is suitable for offline (long-term) traffic prediction applications.

REFERENCE

[1] M.Stemm, S.Seshan, R.H.Katz, "A network measurement architecture for adaptive applications", in *proceedings of INFOCOM 2000 (March 2000),* vol. 1, pp.285-294.

[2] V.Jacobson, "Congestion avoidance and control", in *proceeding of the ACM SIGCOMM 1988 Symposium*, pp. 314-329, August 1988.

[3] S.F.Bush, "Active virtual network management prediction", in *proceedings of the 13th Workshop on Parallel and Distributed Simulation (PADS'99),* pp. 182-192, May 1999.

[4] W.E.Leland, M.S.Taqqu, W.Willinger, D.V.Wilson, "On the self-similar nature of Ethernet traffic (extended version)", in *IEEE/ACM Trans. Networking,* vol.2, no.1, pp.1-15, Feb 1994.

[5] W.Willinger, M.S.Taqqu, R.Sherman, D.V.Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level", *IEEE/ACM Trans. Networking,* vol.5, no.1, pp.71-86, Feb. 1997.

[6] M.E.Crovella, A.Bestavros, "Self-similarity in World Wide Web traffic: Evidence and possible causes", *IEEE/ACM Trans. Networking*, vol.5, no.6, pp.835-846, Dec. 1997.

[7] S.Basu, A.Mukherjee, S.Klivansky, "Time series models for internet traffic", in *proceedings of Infocom 1996*, pp.611-620, March 1996.

[8] N.C.Groschwitz, G.C.Polyzos, "A time series model of long-term NSFNET backbone traffic", in *proceedings of the IEEE International Conference on Communications (ICC'04)*, vol.3, pp.1400-4, May 1994.

[9] Y.Shu, Z.Jin, J.Wang, O.W.Wang, "Prediction-based admission control using FARIMA models", in *Proc. IEEE ICC'00*, vol.3, New Orleans, USA, June 2000, pp.1325-1329.

[10] N.Sadek, A.Khotanzad, T.Chen, "ATM dynamic bandwidth allocation using FARIMA prediction model", in *proceeding of International Conference on Computer Communications and Network 2003 (ICCCN 2003),* p.349-363, Oct 2003.

[11] C.You, K.Chandra, "Time series models for Internet data traffic", in *proceeding of the 24th Conference on Local Computer Networks (LCN99)*, pp.164-171, 1999.

[12] A.Sang, S.Li, "A predictability analysis of network traffic", *Computer networks*, vol. 39, pp. 329-345, 2002.

[13] Y.Gao, G.He and J.C.Hou, "On Leveraging traffic predictability in active queue management", *Proc. IEEE INFOCOM 2002*, June 2002.

[14] R.H.Riedi, M.S.Crouse, V.J.Ribeiro, R.G.Baraniuk, "A multifractal wavelet model with application to network traffic", *IEEE Trans. Information Theory*, vol.45 no.3, pp.992-1018, April 1999.

[15] T.Bollerslev, R.Y.Chou, K.F.Kroner, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence", *Journal of Econometrics*, vol.52, 1992.

[16] C.Gourieroux, "ARCH Models and Financial Applications", *Springer Series in Statistics*, Springer-Verlag New York, 1997.

[17] P.J.Brockwell, R.A.Davis, "Introduction to Time Series and Forecasting", 2nd ed, Springer New York, 2002.

[18] National Laboratory for Applied Networking Research, NLANR network analysis infrastructure. http://pma.nlanr.net/Traces/

[19] P.Abry and D.Veitch, "Wavelet analysis of long-range-dependent traffic", *IEEE Trans. Information Theory*, 44(1), 2-15, 1998.

[20] Matlab website, http://www.mathworks.com/

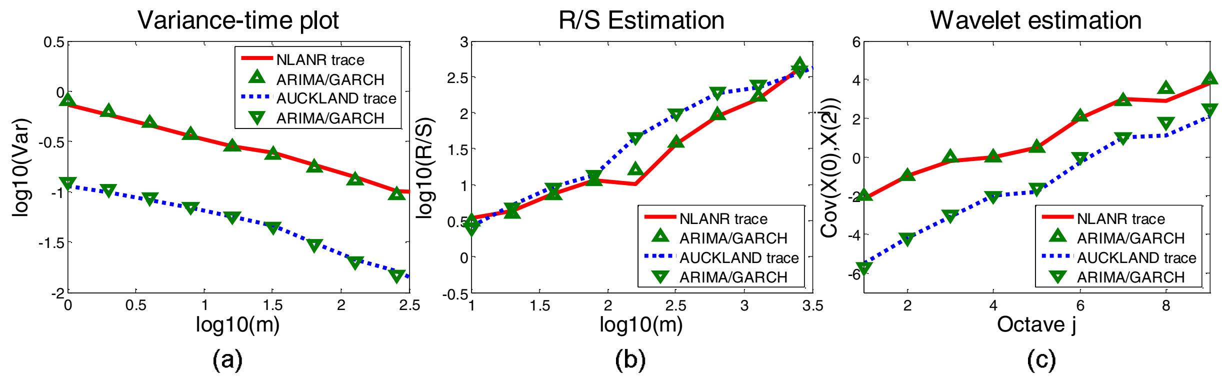

Figure 1 ARIMA/GARCH modeling validation

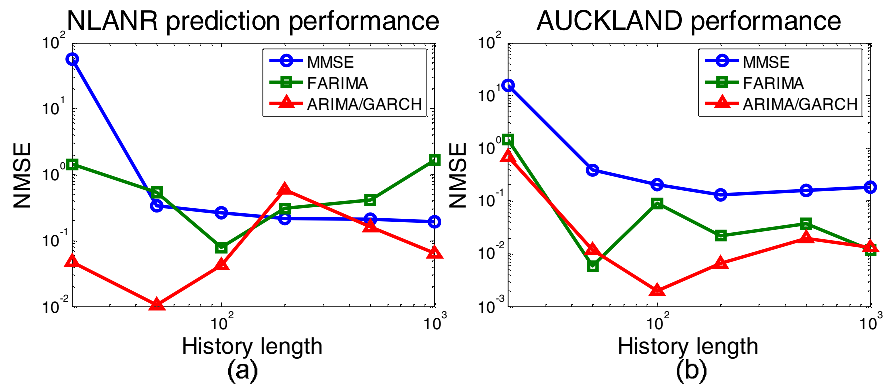

Figure 2 NMSE measure predictability

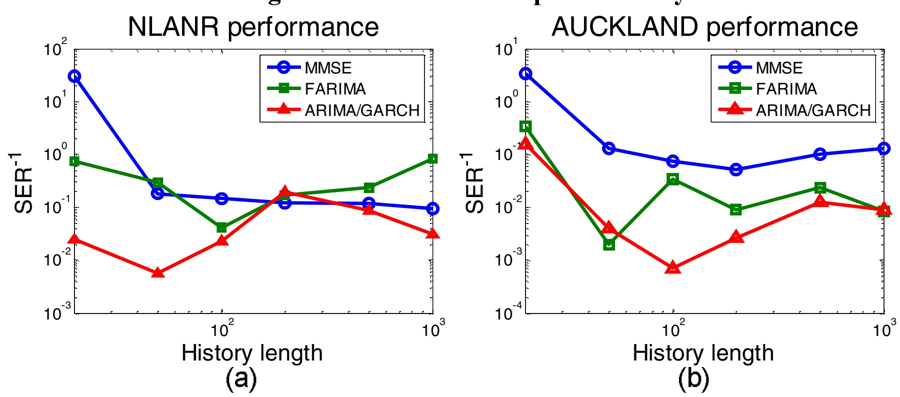

Figure 3 SER^{-1} measure predictability

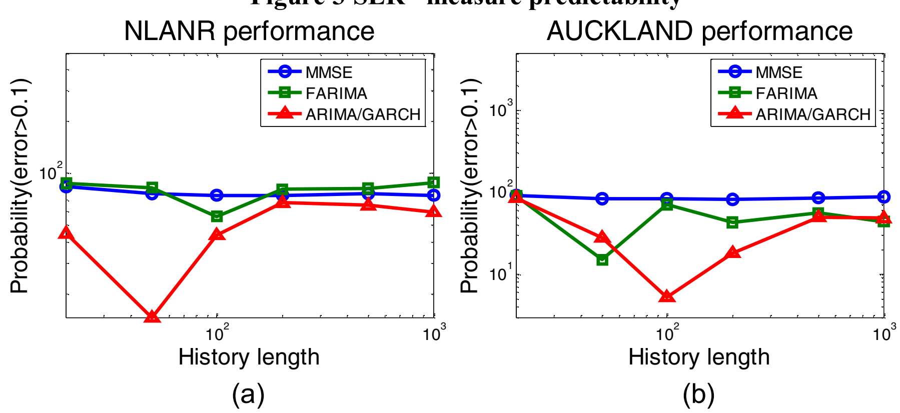

Figure 4 PECP measure predictability