An ARMAX/GRACH time series model for IP traffic trace

Dan He, Zhili Sun and Bo Zhou

Centre for Communication Systems Research (CCSR)
University of Surrey, Guildford, Surrey, GU2 7XH, UK
{D.He,Z.Sun,B.Zhou}@surrey.ac.uk

Abstract: Large-range dependence (LRD) is essential phenomena both in LAN and WAN data traffic. Modeling such traffic traces is significant to understand the nature of the original traffic and to synthesize simulation traffic traces. Existing work such as multi-fractal wavelet model (MWM) was proposed to model IP traces and has reached much exactness. However, the nature of IP traffic exploits burst that is not fully captured by the proposed models. This paper proposes a new time series model ARMAX/GARCH to exploit even more burst network traffic. Through experiment analysis, our model shows that ARMAX/GARCH is able to identify all the important statistical properties of IP traffic traces. By applying partition function and the multi-fractal spectrum as metrics to evaluate both MWM and our model, our model reveals superior than MWM model does while the moment order is positive.

Keywords: Large-range dependence, multi-fractal, traffic modeling, ARMAX/GARCH

1. INTRODUCTION

IP traffic modeling brings significant impact in terms of design, control, and performance analysis of the different type of traffic, e.g., router buffer dimensioning, delay-sensitive service provisioning, and congestion control. Traffic studies have revealed that network traffic exhibits strong variability and burstiness on many time scales. Self-similar or fractal nature has been found not only in the Ethernet Local Network Area [1], but also in many Internet applications such as FTP, WWW [2]. These results sharply conflict the traditional traffic modeling assumptions Where Fractional Brownian motion (fBm) models, FARIMA models, Cox’s M/G/∞ models, and on/off models [3] have been developed to model network traffic. However, extreme analysis of data traffic reveals that more complex scaling behaviour, while cannot be explained in a self-similar framework (mono-fractal), seeking more sophisticated mathematical framework such as multi-fractals is investigated [4] [5]. Moreover, the multi-fractal wavelet model (MWM) has been proven to suit the analysis and synthesis of network traffic data [6].

This paper proposes a new time series model, called a generalized autoregressive conditional heteroskedasticity (GARCH) model [8] [9] to be used for traffic modeling and trace synthesis. The ability of GARCH explicitly models the dependant variance as a linear function of past squared residuals and of its past values by using the Auto-Regressive Moving Average (ARMA (p,q)) model to describe the self-similar network traffic. We applied the
statistical properties of the model to the actual measured network traffic. Through comparing
the Hurst parameter \((H)\) generated by our model with the \(H\) value produced by the original
traffic trace, the proposed model can produce exact approximation. Moreover, after we
compared multi-fractal characteristics of synthetic trace in multi-fractal spectra with those of
\(\beta\)MWM and hybrid MWM respectively, the proposed model shows better approximation.

The rest of the paper is organized as follow: Section 2 overviews the basic concepts of
GARCH model. Section 3 reviews the ARMAX model including the construction procedure
and parameter estimation of the ARMAX/GARCH. Section 4 shows the application of the
model to collected data traffic with its validation based on its multi-fractal properties. Finally,
section 5 is conclusion and future work.

2. GARCH MODEL

The Autoregressive Conditional Heteroskedasticity (ARCH) model was initially proposed
in Engle (1982). An autoregressive process of order \(p\), denoted \(AR(p)\) process, for the square
of \(\epsilon\), forms [10]:

\[
\epsilon_t^2 = c + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + u_t = c + u_t + \sum_{n=1}^{q} \alpha_n \epsilon_{t-n}^2
\]  (1)

where \(u_t\) and \(\epsilon_t\) is white noise:

\[
E(u_t) = 0, E(u_t u_{t'}) = \begin{cases} \lambda^2 & t = \tau \\ 0 & \text{otherwise} \end{cases}
\]

\[
E(\epsilon_t) = 0
\]

\[
E(\epsilon_t \epsilon_{t'}) = \begin{cases} \sigma_t^2 & t = \tau \\ 0 & \text{otherwise} \end{cases}
\]  (2)

The autoregressive conditional heteroskedastic process \(\epsilon_t\) is denoted \(\epsilon_t \sim ARCH(q)\). We
suppose the conditional variance of \(\epsilon_t\) is \(h_t\), and here \(\epsilon_t\) can be expressed in equation:

\[
\epsilon_t = \sqrt{h_t} \cdot z_t
\]  (3)

where \(\{z_t\}\) is an i.i.d. sequence with zero mean and unit variance [11].

Therefore,

\[
h_t = c + \sum_{n=1}^{q} \alpha_n \epsilon_{t-n}^2
\]

since \(h_t\) cannot be negative, thus \(c > 0\) and \(\alpha_n > 0\), for \(n=1,2\ldots q\).

In practice, ARCH modeling requires a large lag \(q\); thereby a large number of parameters
need to be estimated. This is not good for network traffic data process. In order to reduce
computation, Bollerslev [12] extended original ARCH($q$) model to a generalized ARCH (GARCH) by introducing past conditional variances. Specifically, GARCH($p,q$) stands for conditional variance of innovations or residuals ($\varepsilon_j$) [10]:

\[ h_t = \sigma_t^2 = \kappa + \delta_1 h_{t-1} + \delta_2 h_{t-2} + \cdots + \delta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 \]

\[ = \kappa + \sum_{m=1}^{p} \delta_m h_{t-m} + \sum_{n=1}^{q} \alpha_n \varepsilon_{t-n}^2 \]

where $\varepsilon_j$ is an uncorrelated sequence but is required to be co-variance stationary[13]:

\[ \sum_{m=1}^{p} \delta_m + \sum_{n=1}^{q} \alpha_n < 1 \]

We found that GARCH(1,1) model produces a sufficient description of many network traffic. To obtain the parameters ($\kappa, \delta_m, \alpha_n$) [10] [12], we use the Maximum Likelihood Estimation (MLE).

3. ARMAX MODEL

Box and Jenkins proposed Auto-Regressive Moving Average (ARMA($p,q$)) to analyze and forecast univariate time series data. ARMA actually is a combination of a pure “Autoregressive” (AR) process and a “Moving Average” (MA) process [14].

3.1. ARMA model

An ARMA($p,q$) process incorporates both autoregressive and moving average terms:

\[ Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \cdots + \varphi_q \varepsilon_{t-q} \]

Or, using lag operator formula,

\[ (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t = c + (1 + \varphi_1 L + \varphi_2 L^2 + \cdots + \varphi_q L^q) \varepsilon_t \quad \text{(4)} \]

Equation (4) is divided by \( (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) \) to obtain

\[ Y_t = \mu + \psi(L) \varepsilon_t \]

where

\[ \psi(L) = \frac{(1 + \varphi_1 L + \varphi_2 L^2 + \cdots + \varphi_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)} \quad \text{and} \quad \sum_{j=0}^{\infty} |\psi_j| < \infty \]

\[ \mu = c / (1 - \phi_1 - \phi_2 - \cdots - \phi_p) \]

However, the stationary of an ARMA process depends fully on the autoregressive parameters ($\phi_1, \phi_2, \ldots, \phi_p$), rather than on the moving average parameters ($\varphi_1, \varphi_2, \ldots, \varphi_q$) while
the order \( p \) and \( q \) can be determined by the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The properties of ACF and PACF of MA\((q)\), AR\((p)\), and ARMA\((p,q)\) process are described below:

<table>
<thead>
<tr>
<th>Function</th>
<th>MA((q))</th>
<th>AR((p))</th>
<th>ARMA((p,q))</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF</td>
<td>Spikes up to lag ( q ), and cuts off afterwards</td>
<td>Slow decay like infinite tails off, declined exponential and/or cosine waves after ( q - p )</td>
<td>Declined exponential and/or cosine waves after ( q - p )</td>
</tr>
<tr>
<td>PACF</td>
<td>Slow decay like infinite tails off, declined exponential and/or cosine waves</td>
<td>Spikes up to lag ( p ), and cuts off afterwards</td>
<td>Declined exponential and/or cosine waves after ( p - q )</td>
</tr>
</tbody>
</table>

### 3.2. ARMAX model

ARMAX model is an extension by taking other time series as input variables, referred as a dynamic regression model. We applied the first and second-order lags of the measured time series as the regression matrix to estimate the model coefficients, and to reconstruct the traffic trace on the basis of the estimated parameters. The ARMAX\((p,q,n)\) can be described as following [15]:

\[
Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \phi_j \varepsilon_{t-j} + \sum_{k=1}^{n} \beta_k X(t,k)
\]

where \( \beta \) is coefficient vector of regression component matrix \( X \). In practice, the coefficients \( (\phi_i, \phi_j, \beta_k) \) are estimated by MLE method.

### 4. ARMAX/GARCH model evaluation

In this section, we examine the effectiveness of the model in simulation of some real data trace. After fitting all parameters required for ARMAX/GARCH model, synthetic data trace is generated and compare with the real traces in the scaling properties analysis.

### 4.1. Model fitting

Data traces were collected from the 1989 Bellcore Ethernet traces. We focus on two traces of them, say; pAug89 and Oct89Ext4 represented LAN and WAN IP traffic respectively. These data trace provide a well-known benchmark useful for checking the self-similarity and factuality of network traffic although they are slightly dated. As [1] indicates, each observation represents the number of bytes sent over the Ethernet (rep. WAN) every 10 milliseconds. Therefore, the time series obtained imply the traffic flow in the unit time.
First, we analyze the graphical properties of the ACF and the PACF to evaluate the order \( p \) and \( q \). For trace pAug89, the optimal \( p=0, q=1 \), while for trace Oct89Ext4, \( p,q \) are both estimated as one. Normally, GARCH(1,1) is enough to model the conditional variance, thus the ARMAX(0,1)/GARCH(1,1) model is determined to fit pAug89 and ARMAX(1,1)/GARCH(1,1) to Oct89Ext4. Specifically, the mathematical expressions of both models are given below:

\[
\text{ARMAX}(0,1): \quad Y_t = c + \varepsilon_t + \phi_t \varepsilon_{t-1} \\
\text{ARMAX}(1,1): \quad Y_t = c + \phi_t Y_{t-1} + \varepsilon_t + \phi_t \varepsilon_{t-1} \\
\text{GARCH}(1,1): \quad h_t = \kappa + \delta h_{t-1} + \alpha \varepsilon_{t-1}^2 \\
\kappa > 0 \quad \delta > 0 \quad \alpha > 0 \\
\alpha + \delta < 1
\]

where \( Y_t \) represents the fitted data trace, \( \varepsilon_t \) is the error of the modeled trace compared with the true trace (i.e., innovations), and \( h_t \) is conditional variance \( \varepsilon_t \). Second, in order to infer the values of \( \varepsilon_t \) based on \( h_t \) according to their relation (3); the probability distribution of \( \{z_t\} \)should be examined first. A number of experiments show that \( \{z_t\} \)is standard normal (i.e., Gaussian) i.i.d. random variable, provided the number of sample data is large enough. Therefore, we model the data trace assuming the \( z_t \sim N(0,1) \).

4.2. Self-Similar analysis

To show the self-similar property of the trace, we use the variance-time plot to acquire a qualitative characterization of the correlations held in the generated data traces. By observing the log-log plot (Fig.1), we can see that the two traces exhibit LRD with Hurst parameters \( (H) \approx 0.79 \) and \( H \approx 0.85 \) respectively. We also illustrate the variance vs. time with linear fit of GARCH modeled pAug89 and Oct89Ext4 data traces (Fig.1). We observe that there are very close match in the two fit lines, indicating that ARMAX/GARCH model can capture the LRD properties, which real data traces hold.
4.3. Multi-fractal analysis

To further check whether these modeled data traces are truly multi-fractals, we measure the partition function $T(q)$ [16] as the slope of a linear fit of the log-log plot of the sample moments $S_j(q)$ at resolution $2^{-j}$ against the scale $j$, because $T(q)$ is defined as:
\[ T(q) = \lim_{n \to \infty} \frac{1}{n} \log_2 \left[ \sum_{k=0}^{2^n-1} \left( \Delta_{k}^n [Y] \right)^q \right] \]

And \( S_j(q) \) is given as:

\[ S_j(q) = \sum_{k=0}^{2^j-1} \left( \Delta_{k}^j [D] \right)^q \]

Thus, the relation of form \( 2^{(-j)p(q)} \equiv S_j(q) \) is established. In particular, we focus on the interarrival times series of pAug89 trace, as measured by Leland et al. [17], rather than the data length sequences of that trace. In Fig. 2(a), the \( S_j(q) \) of real data trace is almost linear against scale \( j \) except a slight deviation at the very finest resolution. Hence, we can conclude that interarrival time series of pAug89 is multi-fractal. Then, we measure the multi-fractal properties of the synthetic trace. Form the linearity of the log-log plots in Fig. 2(b); we observe that the synthetic trace manifests a multi-fractal scaling with the exception of strongly negative \( q \) and large \( j \).
From Fig.3, the $T(q)$ vs. $q$ and we notice that $T(q)$ is always concave. In the interest that whether ARMAX/GARCH model is better model than Multi-fractal Wavelet Model (MWM), we compare their $T(q)$’s in Fig. 3, and see that ARMAX/GARCH excels βMWM, while hybrid MWM shows its superiority in matching the $T(q)$ of real data trace.

To apply the Legendre transform [16], we convert partition function $T(q)$ into the multi-fractal spectrum $f_G(\alpha)$ of Fig. 4. According to Large Deviation Principle (LDP) [18],
multi-fractal spectrum provides the large range of deviations from the ‘most frequent’ singularity $\alpha$, and hence gives much information that how often the rare events such as bursts (small $\alpha$) occurs. Note that $q$ values are changing from 0 to 6.4 and all curves behave to be concave. Clearly among all traces, the spectrum of ARMAX/GARCH generated trace matches best that of the real data, whereas the spectra of MWM synthetic traces show a degree of divergence on the left. It indicates that the probability of observing small $\alpha$ in the MWM data is somewhat too small. Thus, the occurrence of bursts estimated by MWM is less frequent than that in the actual condition. Additionally, the spectrum of ARMAX/GARCH behaves a little bit worse when $\alpha$ is large, which corresponds to negative $q$ values. That means ARMAX/GARCH generated trace has the values that are too small. In fact, the generated trace has a lot of negative values, which is not acceptable because interarrival time must strictly positive.

![Multi-fractal spectra](image)

Fig. 4 Multi-fractal spectra ($q = 0 \text{ to } 6.4$)

5. CONCLUSION AND FUTURE WORK

The ARMAX/GARCH model is used for network traffic modeling for describing and synthesizing the LRD real data traces. Applying the model to the data of LAN and WAN IP traffic per unit time, ARMAX/GARCH model can fully capture the self-similarity to the popular internet trace files. The ARMAX/GARCH model is successful to build the multi-fractal traffic data. By using both partition function and multi-fractal spectrum applied to the synthesized trace it exploited close match to those of the real data. The MWM synthetic traces are compared to corresponding trace constructed by the ARMAX/GARCH model for examining both multifractality. The results show that the proposed model is slightly worse than that of hybrid MWM synthetic trace in partition function measure while it is still quite good. And the multi-fractal spectrum of generated trace shows excellent when $q$ is positive. We conclude that the ARMAX/GARCH model is a useful competent for the analysis and characterization of IP network traffic flow.
A drawback of the proposed mode is that the generated traces have a number of negative values due to Gaussian process. The future work is correction to the proposed model and evaluates the impact on network queuing performance.

REFERENCES

4. M.S. Taqqu, V. Teverovsky, and W. Willinger, "Is network traffic selfsimilar or multifractal?" Fractals, 5:63-73, 1997