MCMC simulation of GARCH model to forecast network traffic load

Dr. Akhter Raza Syed[1], Hussain Saleem[2] and Habib-Ur-Rehman Syed[3]

**[1] Department of Computer Science, University of Karachi
Karachi, 75270, Pakistan**

**[2] Department of Computer Science, University of Karachi
Karachi, 75270, Pakistan**

**[3] Department of Mathematics, Hamdard Institute of Management Sciences,
Karachi, Pakistan**

Abstract

The performance of a computer network can be enhanced by increasing number of servers, upgrading the hardware, and gaining additional bandwidth but this solution require the huge amount to invest. In contrast to increasing the bandwidth and hardware resources, network traffic modeling play a significant role in enhancing the network performance. As the emphasis of telecommunication service providers shifted towards the high-speed networks providing integrated services at a prescribed Quality of Service (QoS), the role of accurate traffic models in network design and network simulation becomes ever more crucial. We analyze a traffic volume time series of internet requests made to a workstation. This series exhibits a long-range dependence and self-similarity in large time scale and exhibits multifractal in small time scale. In this paper, for this time series, we proposed Generalized Autoregressive Conditional Heteroscedastic, (GARCH) model, and practical techniques for model fitting, Markov Chain Monte Carlo simulation and forecasting issues are demonstrated. The proposed model provides us simple and accurate approach for simulating internet data traffic patterns.

Keywords: *GARCH, Simulation, forecasting, MCMC, network traffic, load.*

1. Introduction

In the early 1990s, two seminal papers [1] and [2] showed that traffic traces captured on both LANs and WANs exhibit Long Range Dependence (LRD) properties, and self-similar characteristics at different time scales. Those discoveries spurred a significant research effort to understand data traffic in packet networks in general, and in the Internet in particular. A number of attempts were made to develop models for LRD data traffic. Looking at packet traffic as a superposition of source–destination traffic flows, simple ON/OFF models were proposed as a first way to mimic LRD properties [3] and [4]. The statistical analysis of real traffic traces, due to the significant amount of collected data and of research projects, gave new impulse to traffic modeling. Among the numerous generic LRD models proposed in the literature, Fractional Brownian Motion (FBM) received a lot of attention, since its Gaussian nature helps in the study of the queuing behavior [5] and [6]. However, this model presents a restrictive correlation structure that fails to capture the short-term correlation of real traffic and its rich scaling behavior. Therefore, many research efforts were devoted to Multifractal models [7], whose attractiveness is due to their rich scale-invariance properties. Wavelet decomposition has been widely used as a natural approach to study scale invariance, but only recently they were introduced in the field of data networks. There are many examples of measurement-based traffic models, which try to fit the LRD properties of real traffic [8] and [9]. These models are computationally very efficient, but they are complex and difficult to tune, due to the lack of a mapping between the traffic parameters and the model coefficients. FARIMA models [10] are widely used in video trace modeling, and can be used to generate LRD sequences. These models are derived by filtering white Gaussian noise, and capture both the short and the long period correlations of traffic. However, the models are quite complex, and their structure makes it very hard to understand the relationship among the filter coefficients and the real traffic data.

On the other hand, threshold autoregressive (TAR) model [11] is proposed for the traffic exhibited non-stationary and non-linear behavior. In [11] the authors have developed the first network measurement system which integrate prediction and they have also proposed running multiple predictors simultaneously and forecasting one which exhibiting the smallest prediction error produced on its measurements. Another significant prediction research work has been introduced in [12], which analyzed the prospects for multi-step prediction of network traffic using ARMA and MMPP models. Their analysis is based on continuous time ARMA and MMPP models driven by Gaussian noise sources. Making the assumption that such models are appropriate, they have developed analytic expressions on how far into the future prediction is possible before errors would exceed a bound, and they also showed how traffic aggregation and smoothing monotonically can help to increase prediction accuracy. Apart from the above mentioned model-based prediction schemes, [13] has reported that non-model-based prediction provides better prediction than model-based prediction as long as the traffic have long range dependence or self-similar. However, the authors only compared their non-model-based prediction model with the FARIMA and FBM models. Both of these two models cannot capture traffic bursty very well and this bursty characteristic effects traffic prediction accuracy.

In this paper, time series model for data set that is related to this problem is developed. The data set is the volume in bytes of internet server access requests, aggregated over half-hour time intervals. The series exhibit strong non-stationary behavior as shown in fig 2(a). We apply Autoregressive Integrated Moving Average with Generalized Auto Regressive Conditional Heteroscedastic model. In the ARIMA model the variance is constant; therefore these models cannot capture such characteristics well. Due to this reason we introduced conditional variance model in which variance varies over time. This model can capture the required effects. The data is described in section two. Section three introduces the conditional mean and conditional variance model along with parameter estimation used in this work, fitting of the model is discussed in section four. Markov chain Monte Carlo (MCMC) simulation and prediction schemes are demonstrated in section five and six.

2. Data and its statistical features

The time series that we used here in this study represents volume of hypertext transfer protocol (HTTP) (Internet) requests to the World Wide Web server in the Computer Science Department at the University of Calgary, for the period from 12:00 A.M., Saturday, February 11th, 1995, to the end of Wednesday, October 11th, 1995. Total request volumes in bytes in successive 30-minutes intervals were aggregated to form a time series. The behavior of the internet requests presents many challenges, the key one being, non-stationary behavior. Which means the statistical properties of network load is changing from time to time (see Figure 1b). We can make the stationary signal using differencing, filtering, or by smoothing. The autocorrelation and partial autocorrelations of the data clues the underlying structure for the model to be proposed. After the preliminary model is available, its parameters need to be estimated from the measured data. Finally comparison of the performance measures of the original and the simulated signal using the proposed model needs to be done.

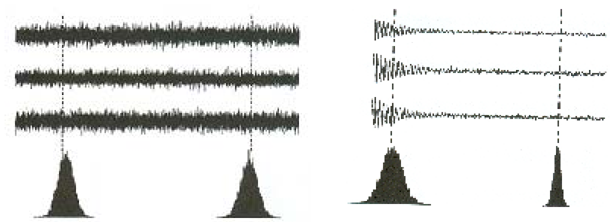


Fig. 1 Stationary and non-stationary time series (left to right)

3. The model

3.1 Conditional mean model

Autoregressive integrated moving average model assumes that errors having constant variance, whenever this variance changes from time to time this model is inappropriate. We used this model with the combination of Generalized Auto Regressive Conditional Variance Model. We call this combination as conditional mean and conditional variance model. This model can be represented by following mathematical expressions.

A stochastic model that can be extremely useful in the representation of certain practically occurring series is the autoregressive model. In this model the current value of the process is expressed as a finite, linear aggregate of previous values of the process and a shock ε_t

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi_p Z_{t-p} + \varepsilon_t \qquad (1)$$

If we express autoregressive operator of order p as $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p$ then model 1 can be expressed in its short form as

$$\phi(B) Z_t = \varepsilon_t \qquad (2)$$

Another kind of model, of great practical importance in the representation of observed time series, is the finite moving average process in which the series depends over a finite number q of previous shocks ε_t's

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q} \qquad (3)$$

The moving average operator of order q is $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q$ then model 3 can be simplified as

$$Z_t = \theta(B)\varepsilon_t \quad (4)$$

Many empirical time series behave as though they had no fixed mean. Even so, they exhibit homogeneity in the sense that apart from local level, or perhaps local level and trend, one part of the series behaves much like any other part. Models that describe such homogeneous non-stationary behavior can be obtained by supposing some suitable difference of the process to be stationary. Hence important class of model for which the d^{th} difference is a stationary mixed autoregressive moving average process is described in equation 5 below. These models are called Autoregressive Integrated Moving Average ARIMA process of order (p, d, q).

$$\nabla^d Z_t = \sum_{i=1}^{p} \phi_i (\nabla^d Z_{t-i}) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (5)$$

$$\phi(B)\nabla^d Z_t = \theta(B)\varepsilon_t \quad (6)$$

In model 5 or 6 it is assumed that the shocks or innovations are normally distributed with zero mean and constant variance σ_ε^2, however the series in question is much typical and the this assumption is not valid for it. We need another technique which model the volatility in the data appropriately.

3.2 Conditional variance model

The conditional variance of innovations, σ_t^2 is defines as

$$Var(\varepsilon_t) = \sigma_t^2$$

The general GARCH(R, M) model for the conditional variance of innovations is

$$\sigma_t^2 = \kappa + \sum_{i=1}^{R} G_i \sigma_{t-1}^2 + \sum_{j=1}^{M} A_j \varepsilon_{t-1}^2 \quad (7)$$

With constraints $\sum_{i=1}^{R} G_i + \sum_{j=1}^{M} A_j < 1$

$\kappa > 0 \quad G_i \ge 0 \quad A_j \ge 0$ Where $i = 1,2,3,...,R$

$j = 1,2,3,...,M$

When we combine the two models (5) and (7) together the resulting model is

$$\begin{aligned} \nabla^d Z_t &= \sum_{i=1}^{p} \phi_i (\nabla^d Z_{t-i}) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \\ \varepsilon_t &\sim WN(0, \sigma_t^2) \\ \sigma_t^2 &= \kappa + \sum_{i=1}^{R} G_i \sigma_{t-1}^2 + \sum_{j=1}^{M} A_j \varepsilon_{t-1}^2 \end{aligned} \quad (8)$$

3.2 Parameter Estimation

The first step in fitting the above specified model is estimation of parameters. Generally there are $'p'$ parameters $\phi_1, \phi_2, \phi_3, ..., \phi_p$ related with autoregressive process $'q'$ parameters $\theta_1, \theta_2, \theta_3, ..., \theta_q$ related with moving average process for the differencing there is one parameter $'d'$. The conditional variance model (7) has $(R + M + 1)$ parameters which are $\kappa,\ G_1,\ G_2, ..., G_R,\ A_1,\ A_2, ..., A_M$. Hence the total number of parameters related with conditional mean and conditional variance model (8) is $(p + q + R + M + 2)$. Figure 2(b) shows that the series is non-stationary hence first we have to make it stationary by estimating the parameter d. Stationarity ensures that early values of ε_t have little influence on the current value of the series. It also ensures that setting a few values of ε_t to zero at the beginning of a series does not affect the predictions very much, provided the series is moderately long. Using the estimation method in [14], the differenced parameter d can be estimated and tested by autocorrelation function.

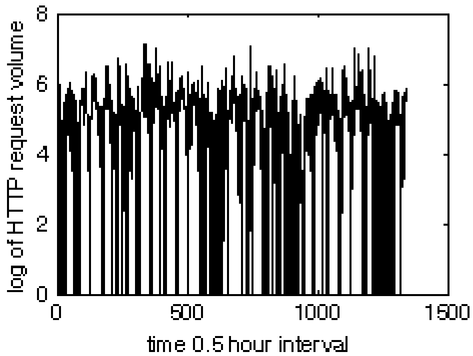


Fig. 2(a)

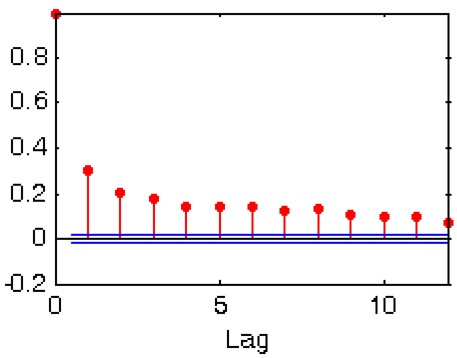


Fig. 2(b)

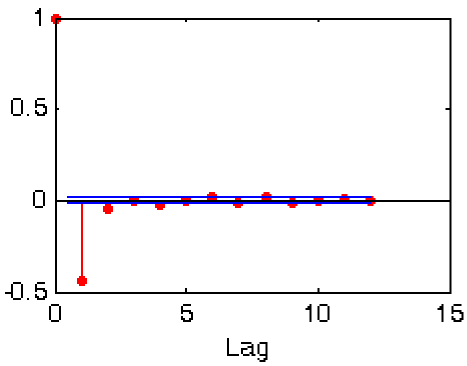


Fig. 2(c)

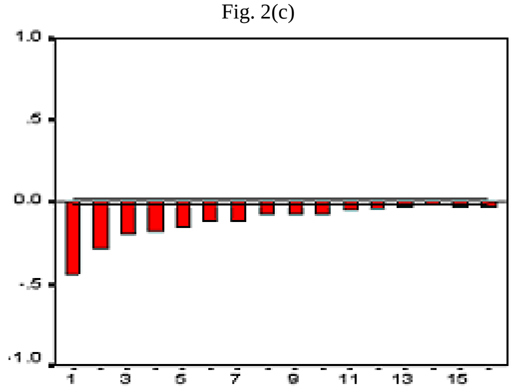


Fig. 2(d)

Fig. 2 (a) logarithms of HTTP request volume plus one, aggregated in half-hour time units first 4 weeks data (b) their autocorrelations. 2(c) autocorrelations after differencing (d) partial autocorrelations after differencing

In the autoregressive moving average model the mean is conditionally changed but the variance remains fixed. Using the autocorrelation and partial autocorrelation functions [14] of the differenced series we can determine the parameters ϕ_i's and θ_j's . From the characteristics of the Autocorrelation function, it describes the correlation between the current states of the time series with the past. The order of moving average "q" can be determine using ACF. The partial autocorrelations PACF can be used to find the order of the auto regression "p". Table 1 summarizes the possibilities.

The initial parameter estimates of conditional mean model are estimated by three step method outlined in [15]. At first, we estimate the autoregressive coefficients by computing the sample autocovariance matrix and solving the Yule-Walker equations. Then using these estimated coefficients, we filter the observed series to obtain a pure moving average process. Finally, we compute the autocovariance sequence of the moving average process, and use it to iteratively estimate the moving average coefficients. This last step provides an estimation of the unconditional variance of the innovations. The method of estimation for the parameters in conditional mean and conditional variance model is described in [16].

Table 1: Summary of Model Identification

Function	ACF	PACF
MA(q)	Function drops off to 0 after lag q	Function tails off exponentially
AR(p)	Function tails off exponentially	Function drops off to 0 after lag q
ARMA(p,q)	Function tails off exponentially	Function tails off exponentially
Noise	0	0

4. Model Fitting

The time series that we used here in this study represents volume of hypertext transfer protocol (HTTP) (Internet) requests to the World Wide Web server in the Computer Science Department at the University of Calgary, for the period from 12:00 a.m., Saturday, February 11th, 1995, to the end of Wednesday, October 11th, 1995. Total request volumes in bytes in successive 30-minutes intervals were aggregated to form a time series. Figure 2(a) illustrates the logarithms of the 1st week series after adding 1. The graph of autocorrelation of the actual data (figure 2b) shows that the series is non-stationary and differencing is required. In figure 2(c) the autocorrelations and in figure 2(d) the partial autocorrelations are represented after taking the first difference of the actual series. Figure 2(c) and 2(d) suggest that p=1 and q=1. For the different values of R and M we fit model (7), we found that the suitable values are

R=1 and M=1. Finally ARIMA(1,1,1)/GARCH(1,1) model is determined to predict the HTTP request series. We fit the conditional mean and conditional variance model (8) to HTTP requests data described above. The estimated coefficients are given in Table 2.

Results in table 2 shows that constant term in the conditional mean model is insignificant and its t-Statistic is low. We can ignore this constant in the model. Table 3 represents the estimated results after ignoring C (constant term). Observe that there is a very slight difference in both results.

Table 2: Estimated parameters including C in conditional mean model

Parameter	Estimates	S. E.	t-statistic
C	4.8684	104.93	0.0464
ϕ_1	0.35992	0.007494	48.02
θ_1	-0.97658	0.00065359	-1494
κ	9.7283e+010	1.0392e-006	200.9
G_1	0.73164	0.00076786	952.8
A_1	0.19268	0.0018513	104.7

Finally the fitted model is

$$y_t = 0.35991\ y_{t-1} + \varepsilon_t - 0.97662\ \varepsilon_{t-1} \qquad (9)$$

$$\sigma_t^2 = 9.7e10 + 0.73163\ \sigma_{t-1}^2 + 0.19273\ \varepsilon_{t-1}^2$$

Table 3: Estimated parameters without C in conditional mean model

Parameter	Estimates	S. E.	t-statistic
ϕ_1	0.35991	0.0050621	71.09
θ_1	-0.97662	0.00063085	-1548.09
κ	9.7283e+010	1.3806e-014	230.90
G_1	0.73163	0.00062131	1177.55
A_1	0.19273	0.0018166	106.09

5. Monte Carlo Simulations

The word simulation, refer to any analytical method meant to imitate a real-life system, especially when other analyses are too mathematically complex or too difficult to reproduce.

Monte Carlo (MC) simulations are stochastic techniques, meaning they are based on the use of random numbers and probability statistics to investigate problems. Strictly speaking, to call something a "Monte Carlo" experiment, all you need to do is use random numbers to solve the problem.

Algorithm $TrafficSimulator(nSim, nBurn)$

[This algorithm simulate model (9) and simulate $ySim$ for conditional mean model and σSim^2 for conditional variance model. The inputs are $nSim$ that represent the number of simulation required and $nBurn$ which represent the number of observations to be omitted for the burn in period]

Step1: set k=1, Initialize $ySim_k$ and σ_k^2 using some appropriate values [probably the mean and variance of given series]

Setp2: replace k by k+1

Step3: generate a random number from N(0, σ_{k-1}^2) and store it into ε_{k-1}

Step4: replace σ_k^2 with
$$9.7e10 + 0.73163\ \sigma_{k-1}^2 + 0.19273\ \varepsilon_{k-1}^2$$

Step5: replace $ySim_k$ with
$$0.35991\ ySim_{k-1} + \varepsilon_k - 0.97662\ \varepsilon_{k-1}$$
[repeat step 2-5 up to $k \le nSim$]

Step6: repeat for k=1, 2, 3… ($nSim$ - $nBurn$)
If $ySim_k \le 0$ then
replace $ySim_k$ with zero
[end of if structure]

Step7: replace σSim^2 with the series σ^2

Step8: [delete first $nBurn$ observations from $ySim_k$]

Figure 3 shows the simulated series with their conditional variances.

6. Predictions

Consider the conditional mean model of first order

$$(1-\phi B)\nabla Z_t = (1-\theta B)\varepsilon_t$$

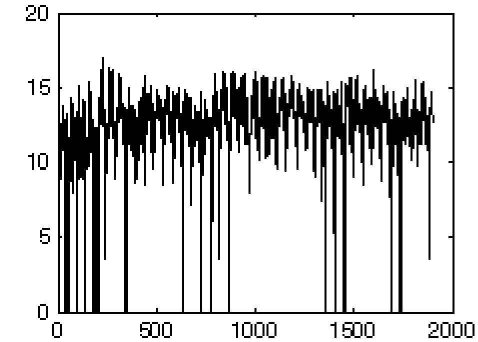


Fig. 3(a)

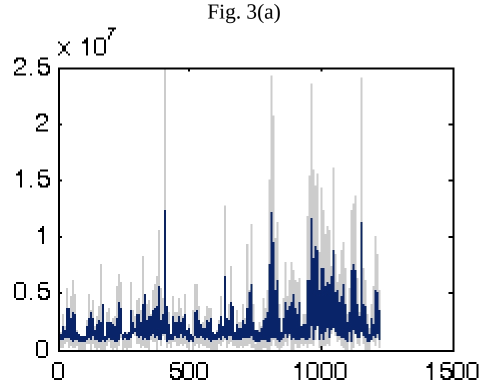


Fig. 3(b)

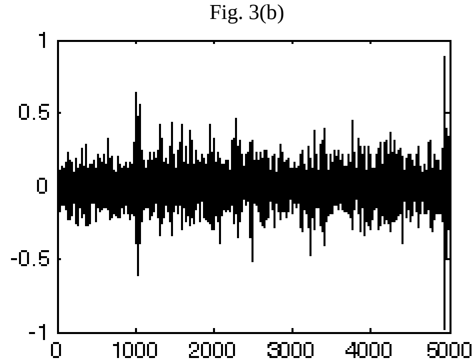


Fig. 3(c)

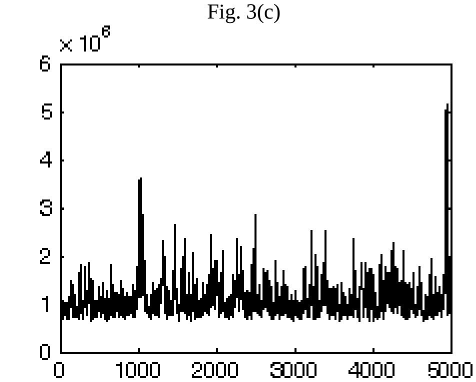


Fig. 3(d)

Fig. 3 (a) Logarithms of simulated observations (b) actual versus fitted HTTP requests (c) simulated innovations (d) simulated conditional standard deviations.

This model in difference equation form is

$$Z_t = (1+\phi)Z_{t-1} - \phi Z_{t-2} + \varepsilon_t - \theta_{t-1}\varepsilon_{t-1} \tag{10}$$

One step-ahead forecast is

$$\hat{Z}_t(1) = (1+\phi)Z_t - \phi Z_{t-1} - \theta\varepsilon_t \tag{11}$$

$l-steps$ ahead forecast will for the conditional mean model will be.

$$\hat{Z}_t(l) = (1+\phi)\hat{Z}_t(l-1) - \phi\hat{Z}_t(l-2) \tag{12}$$

where $l > 1$

Independently from the conditional mean, we can forecast the conditional variance. In the simple GARCH(1, 1) case, the optimal l-step-ahead forecast of the conditional variance, i.e. $\hat{\sigma}^2_{t+l|t}$ given by:

$$\hat{\sigma}^2_{t+l|t} = \kappa + G_1\sigma^2_{t+l-1|t} + A_1\varepsilon^2_{t+l-1|t} \tag{13}$$

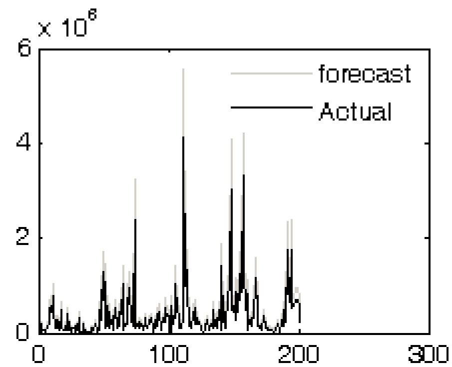


Fig. 4(a)

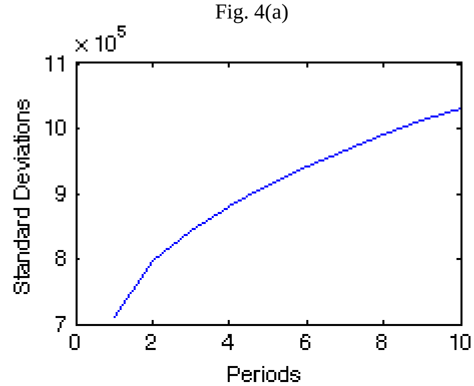


Fig. 4(b)

Fig. 4 (a) actual versus mean forecast (b) volatility forecast of HTTP requests

Figure 4(b) shows that conditional standard deviation forecast approaches the unconditional standard deviation of ε_t which is given by

$$\sigma = \sqrt{\frac{\kappa}{1-\sum_{i=1}^{R} G_i - \sum_{j=1}^{M} A_j}} \tag{14}$$

Table 4: Comparison of conditional mean conditional variance with seasonal autoregressive moving average model

	Our model	SARIMA model
R^2	0.473	0.315
AIC	30.12	30.4
BIC	30.13	32.2

We compared the proposed model with seasonal auotoregressive moving average model and found that proposed model gives better results. Table 4 gives the comparative results of the two models.

7. Conclusions

We applied ARIMA(1, 1, 1) with GARCH(1, 1) model on HTTP requests series. The method is relatively straight forward to implement and capture both the effects found in the series due to conditional mean and conditional variance. It shows that non-linear time series model can be used to forecast better than the classical linear time series models, even the linear time series model can also behave self-similarity. This model has the capability to capture volatility found in variances.

We compare the results of conditional mean and conditional variance model with the seasonal autoregressive moving average model and found that the proposed model gives better results. We cannot generalize these results for the other network series because due to the different topologies of computer network, the behavior of the traffic will be different. We are still working in this field and trying to find out the models that can perform better than the one used in this work.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger and V. Wilson, On the self-similar nature of ethernet traffic (extended version), *IEEE/ACM Transaction on Networking* **2** (1994) (1), pp. 1–15

[2] V. Paxson and S. Floyd, Wide-area traffic: the failure of poisson modeling, *IEEE/ACM Transactions on Networking* **3** (1995) (3), pp. 226–244

[3] M.E. Crovella and A. Bestavros, Self-similarity in world wide web traffic: evidence and possible causes, *IEEE/ACM Transaction on Networking* **5** (1997) (6), pp. 835–846

[4] W. Willinger, M.S. Taqqu, R. Sherman and D.V. Wilson, Self-similarity through high variability: statistical analysis of ethernet LAN traffic at the source level, *IEEE/ACM Transaction on Networking* **5** (1997) (1), pp. 71–86.

[5] Norros, On the use of fractional Brownian motion in the theory of connectionless networks, *IEEE Journal on Selected Areas in Communications* **13** (1995), pp. 953–962.

[6] Norros, A storage model with self-similar input, *Queueing Systems* **16** (1994), pp. 387–396

[7] H. Riedi, M.S. Crouse, V.J. Ribeiro and R.G. Baraniuk, A multifractal wavelet model with application to data traffic, *IEEE Transaction on Information Theory* **45** (1999) (4), pp. 992–1018.

[8] S. Ma and C. Ji, Modeling heterogeneous network traffic in wavelet domain, *IEEE/ACM Transaction on Networking* **9** (2001) (5), pp. 634–649.

[9] A. De Vendictis, A. Baiocchi, Wavelet Based Synthetic Generation of Internet Packet Delays, in: Proceedings of International Teletraffic Conference ITC17, Salvador, Brasil, December 2001.

[10] M. Krunz, A. Makowski, A Source Model for VBR Video Traffic Based on M/G/infinity Input Processes, in: Proceedings of IEEE Infocom 98, San Francisco, CA, USA, April 1998, pp. 1441–1448.

[11] C.You and K.Chandra, "Time series models for Interent data traffic", Proceedings of the 24th Conference on Local Computer Networks (LCN 99), pp. 164-171.

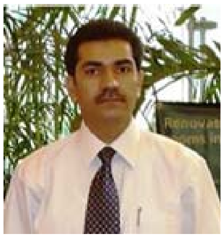
[12] W.E.Leland, M.S.Taqqu,W.Willinger and D.V.Wilson, "On the self-similar nature of Ethernet traffic (Extended Version)", IEEE/ACM Trans. Net., vol.2, no.1, pp. 1-15, Feb 1994.

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

[14] P. Brockwell and R.Davis, "Introduction to Time Series and Forecasting", Springer, 1996.

[15] Box, G.E.P., G.M.Jenkins, and G.C. Reinsel, "Time Series Analysis: Forecasting and Control", 3rd edition, Prentice Hall, 1994.

[16] T.Nakatsuma and H.Tsurumi, "ARMA-GARCH models: Bayes Estimation Versus MLE, and Bayes Non-stationary Test", department working papers with number 199619, department of Economics, Retgurs University, 1996.



Dr. Akhter Raza Syed received Ph.D. in 2011 from Dept. of Computer Science University of Karachi. He had about 21 years teaching experience and affiliated with University of Karachi. From 2004 to 2009 he was affiliated with Institute of Business Administration (IBA) as adjunct faculty. Dr. Raza had several peer reviewed publications and authored 4 books. His research interests include AI, soft computing, data mining, statistics, simulation and stochastic modeling. He is a senior member of IACSIT. He is the recipient of best teacher award by the Education Department Govt. of Sindh at the occasion of World Teachers Day 2011.

Mr. Hussain Saleem is the Assistant Professor and research scholar at Department of Computer Science, University of Karachi. He received his B.S. in Electronics Engineering from Sir Syed University of Engineering & Technology, Karachi in 1997 and Masters in Computer Science from University of Karachi in 2001. He has vast experience of about 15 years including University Teaching, Administration and Research in various dimensions of Computer Science. His field of interest is Software Science, Automation and Simulation & Modeling.

Mr. Habib-ur-Rehman Syed is an Assistant professor in Hamdard Institute of Management Sciences, Hamdard University. Did his masters in Business Administration from University of Sindh in 2002 and in Statistics from University of Karachi in 1990. He had about 20 years teaching experience