Arima Model for Network Traffic Prediction and Anomaly Detection

H.Zare Moayedi
School of electronic Eductaion In IT ,
Shiraz University
Hzm_h@yahoo.com

M.A.Masnadi-Shirazi
Dept of Electrical Engineering,
Shiraz University
Masnadi@Shirazu.ac.ir

Abstract

This paper presents the use of a basic ARIMA model for network traffic prediction and anomaly detection. Accurate network traffic modeling and prediction are important for network provisioning and problem diagnosis, but network traffic is highly dynamic. To achieve better modeling and prediction it is needed to isolate anomalies from normal traffic variation. Thus, we decompose traffic signals into two parts normal variations, that follow certain law and are predictable and, anomalies that consist of sudden changes and are not predictable. ARIMA analysis and modeling for network traffic prediction is able to detect and identify volume anomaly or outliers.

1. Introduction

Studies of the traffic measurement results were trying to understand the network traffic characteristics to form models for simulations and traffic predications. Since the early 1970's, many efforts have been carried out to reveal the traffic in a mathematical method [1] and [2] measurement and analysis of the existing network with the aim to characterize and model the traffic has been done by [3]and

[5] that are mainly focused on locale area networks. In [1] and [2] the selected summaries of traffic on a per-protocol basis in a wide area network has been reported.

Network traffic prediction is of significant interests in many domains, including adaptive applications congestion control, admission control, and network management. 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, based on the measured traffic history. [9].

The most important problem of traffic prediction is traffic predictability. Traffic predictability denotes the possibility for prediction to satisfy some precision requirement over desired prediction and control time interval.

The organization of the paper is as follows; section 2 reviews the ARIMA Model. In section 3 we simulate a network with random process. In section 4 we simulate a network with ARIMA process without trend. Simulation of ARIMA Model with trend has been presented in section 5, The all above Simulation have been carried out with normal variations. ARIMA Model simulation with Anomalies is brought in section 6. Finally, the conclusion is given in section 7.

2. Arima Model

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 [1], their approaches are based on the traditional time series prediction technique, which is 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 that are used to predict network traffic, such as Auto Regressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA).

978-1-4244-2328-6/08/$25.00

2.1. Fractal Arima (Farima)

FARIMA is a model-based prediction model and it has a long range dependence (LRD) behavior. Its mathematical description can be shown as follows:

Let $\{X_t\}$, $t = 0,1,2,\ldots$ denote a stochastic process; the general ARMA (*r,s*) model can be expressed as,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_r X_{t-r} = Z_t - \theta_1 Z_{t-1} - \ldots - \theta_s Z_{t-s}$$
$$\text{where} \quad Z_t \sim N(0, \sigma^2{}_z) \qquad \text{(E 1)}$$

Define the lag operator B as $BX_t = X_{t-1}$, where $B^r X_t = X_{t-r}$. Also assume that Δ denotes the difference operator, i.e. $\Delta X_t = X_t - X_{t-1}$, equivalently $\Delta^d = (1-B)^d$ which can be expressed by the binomial expansion:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k \qquad \text{(E 2)}$$

We also define polynomials $\phi(B)$ and $\theta(B)$ as:

$$\phi(B) = (1 - \phi_1 B - \ldots - \phi_r B^r)$$

$$\theta(B) = (1 - \theta_1 B - \ldots - \theta_s B^s) \qquad \text{(E 3)}$$

The ARMA (r,s) model has the form

$$\phi(B) X_t = \theta(B) Z_t$$
$$\text{where } Z_t \sim N(0, \sigma^2{}_z) \qquad \text{(E 4)}$$

FARIMA is the natural extension of the ARMA process when we allow real values for parameter *d*. X_t is a stationary invertible FARIMA(r,s) process if

$$\phi(B) \Delta^d X_t = \theta(B) Z_t \qquad \text{(E 5)}$$
$$\text{where } Z_t \sim N(0, \sigma^2{}_z)$$

and d is a real number $-0.5 < d < 0.5$ (it is called the ARIMA model if *d*=1).

In FARIMA model, corroding invariability, we can write equation in the following new form

$$X_t = \sum_{u=0}^{\infty} \pi_u Z_{t-u} \quad \text{where}$$

$$X_t = \sum_{u=0}^{\infty} \pi_u B Z_t^u$$

$$\sum_{u=0}^{\infty} \pi_u B^u = \phi(B) \theta^{-1}(B) (1-B)^d \qquad \text{(E 6)}$$

From the theorems on prediction [11], a one-step prediction of a FARIMA process is

$$X_{t+1} = \sum_{u=1}^{\infty} \pi_u X_{t-u+1} \qquad \text{(E7)}$$

2.2. Arima Prediction Models

In this section we briefly introduce the key characteristic of the ARIMA model which is modeling the changes in variance. Then, we give ARIMA model mathematics description and explain its modeling and prediction procedure.

2.2.1. Models for Changing Variance. In ARIMA and FARIMA, Z_t is white noise and its variance is equal to σ_Z^2 In general, this variance value is global and never changes over time [10].

$$\sigma_Z^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j Z_{t-j}^2, \quad \alpha_0 > 0, \quad \alpha_j \geq 0 \quad \text{(E8)}$$

2.2.2. Arima Model. In this paper, model the non-linear time series model: ARIMA network traffic prediction.The ARIMA (*r,s*) mathematical expression can be shown.

$\phi(B)(1-B)^d X_t = \theta(B) Z_t$ and

$$\sigma_Z^2 = \alpha_0 + \sum_{i=1}^{p} \gamma_i \sigma_{Z-i}^2 + \sum_{j=1}^{q} \alpha_j Z_{t-j}^2 \qquad \text{(E 9)}$$

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

Where d is integer, this is different from FARIMA model. The $\phi(.)$ denotes the r degree polynomials and the $\theta(.)$ denotes the s degree polynomials.

ARIMA parameters can be estimated by Box-Jenkins modelling. the general Box-Jenkins approach is to difference the time series until it appears to come from a stationary process it is often found that first-order (d=1)differencing of non-seasonal data is adequate, although second-order (d=2) differencing is occasionally required. Suppose fitting the differencing parameter d, other parameters r and s of ARIMA can be estimated by Box-Jenkins modelling as well. Determining the orders of r and s depend on auto correlation function ACF and its Partial auto correlation function (PACF). Normally, the theoretical (ACF) of an MA (s) has a very simple form in that it "cuts off" at lag s and the PACF of an AR (r) has a very simple form in that it "cuts off" at lag r. After determining the order of ARIMA, we use maximum likelihood estimation (MLE) to estimate initial parameters values, and use Akaka's information Criterion (AIC) to select best fitting parameters [10],[11].

2.2.3. Predictability Definition. We describe three major predictability measure methods in this section. 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. Its mathematical expression can be shown as follows.

$$MSE = E(\mathrm{X}_{t+1} - \overline{X_{t+1}})^2 \qquad \text{(E 10)}$$

where $E(.)$ is the expected value, and $\overline{X_{t+1}}$ is the predicted value of X_{t+1}.

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).

$$NMSE = E(\mathrm{X}_{t+1} - \overline{X}_{t+1})^2 / \sigma^2 \qquad \text{(E 11)}$$

where $\sigma^2 = E((X_{t+1} - \overline{X}_{t+1})^2)$, and $\overline{X}_{t+1}$ is the mean value of X_{t+1}. Thus,

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

$$SER = E(\mathrm{X}^2{}_{t+1}) / E(\mathrm{X}_{t+1} - \overline{\mathrm{X}}_{t+1})^2 \qquad \text{(E 12)}$$

Normally, comparing with expression of NMSE.

$$SER^{-1} = 1 / SER \qquad \text{(E13)}$$

The SER^{-1} method is similar to NMSE. They both compute the prediction error and normalize 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 the following steps.

We assume the normalized one-step prediction error is

$$\overline{err} = \left| (\mathrm{X}_{t+1} - \mathrm{X}_{t+\_1}) / \mathrm{X}_{t+1} \right| \qquad \text{(E 14)}$$

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 $P_{\overline{err}}(\varepsilon)$ are Prediction Error Critical Probability (PECP). Obviously it is found that the smaller than ε leads to the PECP, the better predictability.

3. Simulation of the network with random process

In this section a sample network is simulated with a uniform random number distributed between the intervals [10,100]. 900 samples of the random number can simulate the data from a network interface card for about 15 minutes data which is shown in Figure 1.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

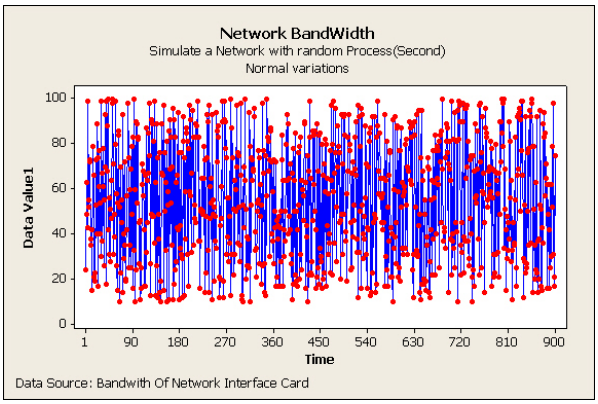

Figure 1. Simulation of the network with random process(second) normal variation

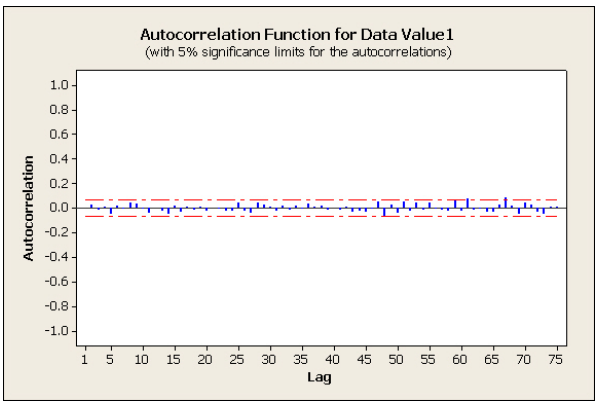

Figure 2. Autocorrelation function for random process

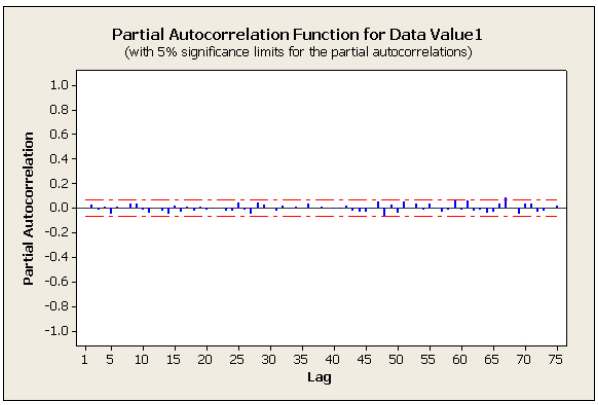

Figure 3. Partial autocorrelation function for random process

The ACF and PACF of data obtained and plotted in Figures 2 and 3.

4. Simulation of the network with Arima Process without trend

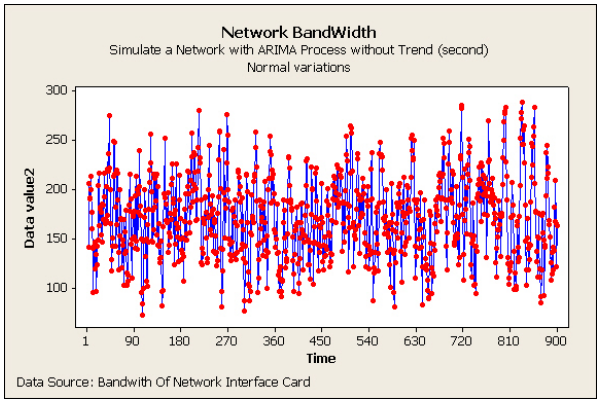

Figure 4. Simulation of the network with Arima process without trend

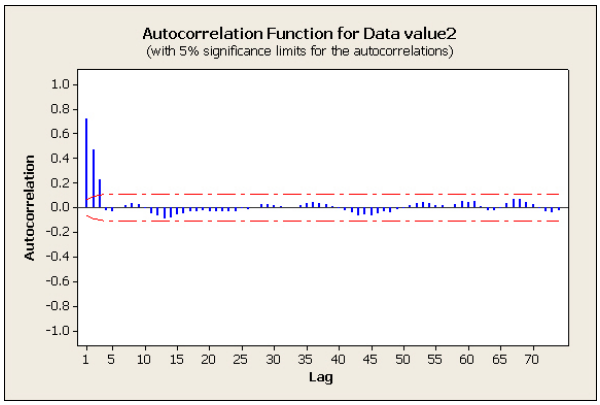

Figure 5. Autocorrelation function for Arima process without trend

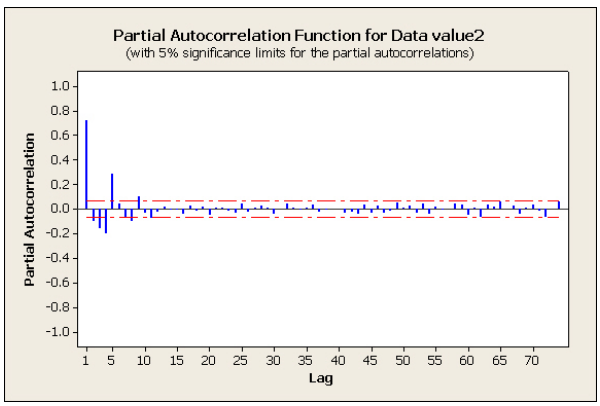

Figure 6. Partial autocorrelation function for Arima process without trend

From the uniform data, the ARIMA model data (3,3) ,(r=3,s=3) are generated and plotted in Figure 4

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

using the corresponding equations.The ACF and PACF are shown in figures 5 and 6.

5. Simulation of network with Arima model with trend

To show Anomaly in the signal, we added a line trend to the data which is show in Figure 7. The corresponding ACF and PACF plot are shown in Figures 8 and 9.

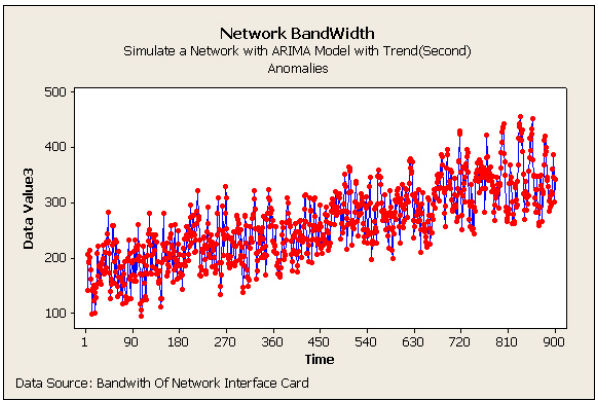

Figure7. Simulation of the network with Arima process with trend

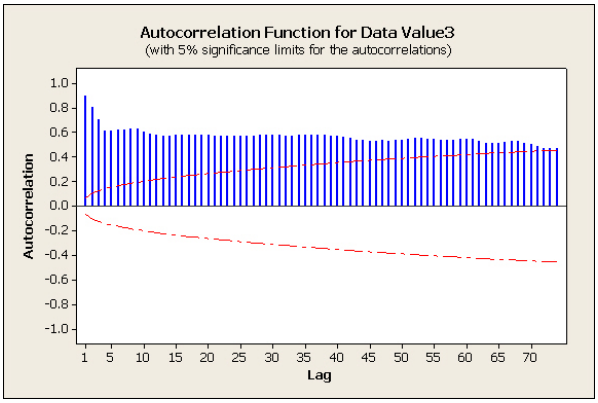

Figure 8. Autocorrelation function for arima process with trend

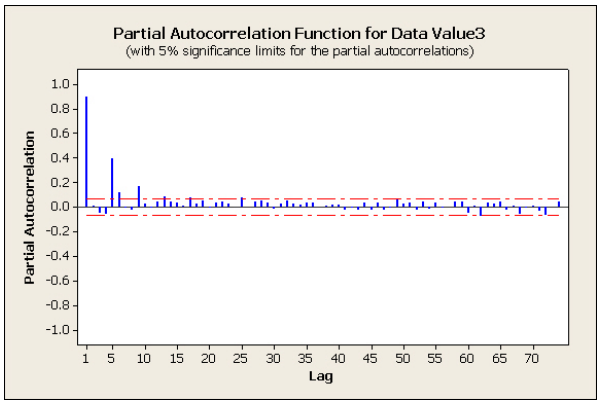

Figure 9. Partial autocorrelation function for Arima process with trend

6. Simulation of the network arima model with anomalies

In Figure 10 the anomaly in the system is considered as some attacks to the systems some specific interval. The corresponding ACF, PACF plots are shown in Figures 11and 12.

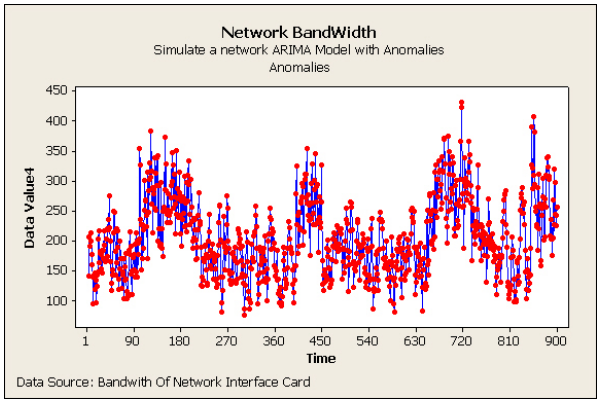

Figure 10. Simulation of the network arima with anomalies

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

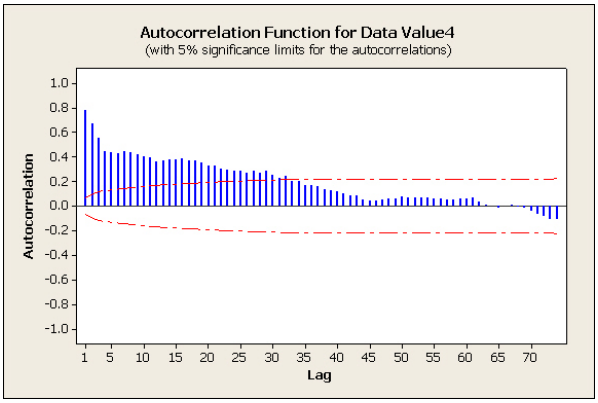

Figure 11. Autocorrelation function for Arima with anomalies

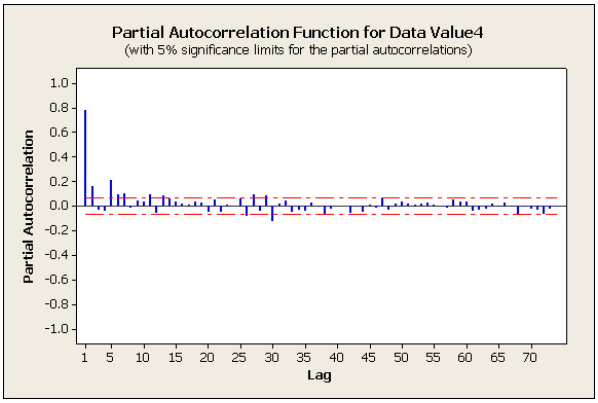

Figure 12. Partial autocorrelation function for arima with anomalies

7. Conclusions

In this paper we simulated network traffic by using ARIMA model. All the normal traffic and traffic with anomalies modeling trend and attacks have been simulated and the corresponding results have been illustrated. The non stationary behavior and the anomalies have been detected in the ACF and PACF plots, with trend and attacks.

8. References

[1] D. L. Jagerman, Benjamin Melamed, Walter Willinger, "Stochastic Modeling of Traffic Processes", 1996.

[2] W. E. Lelan Bellcore, W.Willinger Bellcore, M.d S.Taqqu Department of mathematics Boston University , D. V.Wilson Bellcore ,"On The Self-Similar Nature Of Ethernet Traffic ",March 3 , 1993

[3] P. D. Amer , R. N. Kurnar, R. Kao, J. T. Phillips, and L.N. Cassel, "Local Area Broadcast Network Measurement: Traffic Characterization," *Proc. Of Spring COMPCON'87*, pp. 64-70, IEEE Computer Society, San Francisco, California, February, 1987.

[4] Riccardo Gusella, "The analysis of Diskless Workstation Traffic on an Ethernet," Technical Report UCB/CSD 87/379, Computer Science Division, University of California, Berkeley, California, November, 1987

[5] Raj Jain and Shawn Routhier, "packet Trains: Measurement and a New Model for Computer Network Traffic," *IEEE Journal on Selected Areas in Communication*, 4(6), pp. 986-995, September, 1986

[6] R. Caceres, "Measurements of Wide Area Internet Traffic", Report UCB/CSD 89/550, Computer Science Division, University of California, Berkeley, California, 1989.

[7] J. Crowcroft and I. Wakeman, "Traffic Analysis of some UK-US Academic Network Data", *Proceedings of INET'91*, Copenhagen, June 1991.

[8] David J. Ewing, Richard S. Hall, Michael F. Schwartz, "A Measurement Study of Internet File Transfer Traffic", Tech. Rep. CUCS -571-92, University of Colorado, Dept. of Computer Science, Boulder, Colorado, January 1992.

[9] Zhili Sun ,"Prototyping and Expreimental validation Of network Layer Function In the NGI", University Of Surrey (P55), June 3 ,2006.

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

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

978-1-4244-2328-6/08/$25.00 © 2008 IEEE