

Estimation Techniques for Monitoring and Controlling the Performance of the Computer Communication Networks

¹Ibrahiem M.M. El Emary and ²Adanan I. Al Rabia
¹Amman Al Ahliyya University, Amman, Jordan
²Al Balqa Applied University, Al Salt, Jordan

Abstract: This study was concerned with making a comparative study between four types of time series model with the goal of choosing the optimum one to predict the performance of a computer communication network. The investigated four types of time series are: Least Square (LS), Fourier series, Exponential Weighted Moving Average (EWMA) and the Auto Regressive Integrated Moving Average (ARIMA). Comparative study is based on comparing some of statistical measurement for these four time series models as: mean standard deviation and variance. We apply these various types of time series on two types of network to predict their performance; the first one is called token bus while the second one is called the token ring. Concluded results prove that both EWMA and ARIMA perform better than the others LS and Fourier since both EWMA and ARIMA forecast the network performance parameters with an accuracy of 98% near the actual values.

Key words: Exponential weighted moving average, auto regressive integrated moving average, least square, relative error, standard deviation

INTRODUCTION

Computer communication networks require monitoring to detect performance anomalies, but problem detection and diagnosis is a complex task requiring skilled attention. Although human attention was never ideal for this task, as networks of computers grows larger and their interactions more complex, it falls far short. Existing computer-aided management systems require the administrator manually to specify fixed "trouble" thresholds.

The process of managing a system performance is an iterative one performed by a human, the steps involved might not be as well-defined as described here, but the basic iterative steps still exist. The performance management process is composed of a definition phase, a diagnostic phase and a therapy phase. The definition phase consists of determining what performance criteria are to be evaluated and how the data will be collected and analyzed. Diagnosis begins by making measurements of the system. This is followed by analysis of the collected data. Analysis is done manually or with assistance from another computer. During analysis, the performance is recognized and categorized. After the analysis, if the results suggest that some improvements in performance are possible, a therapy phase is beginning. During this phase, adjustments to the system are made in an attempt to alleviate the problem. The remaining phases are repeated until there are no known performance problems in the system.

This study describes a time series modeling technique that can be used for more automatically

detecting the problems of computer network performance. Techniques that are used are based on four types: EWMA, LS, FOURIER and ARIMA, they allow detection of performance problems in a computer network by providing a means of detecting performance criteria values that are out of normal range

Our effort in this study was focused on making a comparative study between these four types of time series in order to select the optimal one and use it in the management process. The comparative studies depend on using some of statistical criteria as: Mean, standard deviation and variance. Numerical results show that EWMA is the optimum type of time series models since it can forecast or predict the network performance with accuracy approximately equal to 98% of the actual values. According to this output result, network an administrator can use this approach to monitor the network performance and decide if there is a problem or not.

Exponentially weighted moving average time series model (EWMA): In a seasonal time series model (Exponential Weighted Moving Average EWMA), the values being modeled have four components: constant, trend, seasonal, and random^[1]. The model can be used to account for the first three of these components (constant, trend, seasonal). The constant component is the portion of the data that is always present. The trend component reflects the fluctuation in the data that extends throughout the entire time series. The seasonal is the regular variations that occur every period of time. Finally the random component accounts for fluctuations in the data due to undetectable causes.

The basic form of the seasonal time series model is written as:

$$X_t = b_1 + b_2 t + C_t + \epsilon_t \quad (1)$$

Where:

b_1 : A constant component.

b_2 : A trend component

C_t : A seasonal factor

ϵ_t : A random error component

The effect of the seasonal factors C_t is to de-seasonalize the current reading X . The length of the seasonal variation is fixed at length L . The parameters b_1 , b_2 and C_t , $t=1, 2, \dots, L$ must be estimated. These estimates are updated at the end of each period. The model adapts to changes in the data by the use of three smoothing constants α , β , and γ . The α , β , and γ smoothing constants are used to smooth the: constant, trend and seasonal components of the time series model^[2].

The estimates $b_1^{\hat{}}$, $b_2^{\hat{}}$, and $c_t^{\hat{}}$, $t=1 \dots L$ is computed as follows:

$$b_1^{\hat{}}(t) = \alpha [X_t - c_t^{\hat{}}(t-L)] + (1-\alpha) [b_1^{\hat{}}(t-1) + b_2^{\hat{}}(t-1)] \quad (2)$$

$$b_2^{\hat{}}(t) = \beta [b_1^{\hat{}}(t) - b_1^{\hat{}}(t-1)] + (1-\beta) [b_2^{\hat{}}(t-1)] \quad (3)$$

$$C_t^{\hat{}}(t) = \gamma [X_t - b_1^{\hat{}}(t)] + (1-\gamma) [c_t^{\hat{}}(t-L)] \quad (4)$$

Where, $0 < \alpha, \beta, \gamma < 1$

Autoregressive integrated moving average time series model (ARIMA): ARIMA is based on finding a model that accurately represents the past and future patterns of a time series^[1]:

$$Y_t = \text{Pattern} + e_t \quad (5)$$

Where:

Y_t = The forecasting value

e_t = A random error

The pattern can be random, seasonal, trend, or a combination of them.

ARIMA model building methods use a simple, versatile model notation; they are designated by the level of auto regression, integration, and moving averages. This standard notation identifies the orders of auto regression by p , integration or differencing by d , and moving average by q .

Auto regression which is called an ARIMA (1, 0, and 0) or AR (1) is represented by:

$$Y_t = \theta_0 + \Phi_1 Y_{t-1} + e_t \quad (6)$$

where, θ_0 and Φ_1 are coefficients chosen to minimize the sum of squared errors.

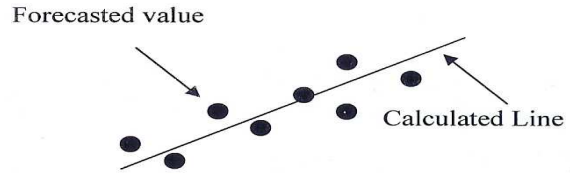


Fig. 1: Idea of least square

Y_t : The forecasting value

e_t : A random error

Moving average which called an ARIMA (0,0,1) or MA (1), is represented by:

$$Y_t = \mu - \theta_0 e_{t-1} + e_t \quad (7)$$

Where:

Y_t : The forecasting value

θ_0 : An estimated coefficient

e_t : A random error

Integrated series which called and ARIMA (0,1,0) is represented by:

$$Y_t = Y_{t-1} + e_t \quad (8)$$

Where:

Y_t : The forecasting value of time t

Y_{t-1} : The forecasting value at time $t-1$

e_t : A random error

Least square time series model (LS): Least Square (LS) is another example of the time series models. The idea of the least square method is to find a line which goes through all points and be close as possible to each one of points as shown in the Fig. 1.

The line equation that represents LS is given by^[1]:

$$Y = mX + b \quad (9)$$

Where:

Y : Forecasted values on y axis

X : Value of time on x axis

So, we must find the line equation with minimum ϵ which is the error square sum i.e.:

$$\epsilon = \sum_{i=0}^t (y_i - y)^2 = \sum_{i=0}^t (y_i - (mX - b))^2 \quad (10)$$

where, y_i : the forecasted value of time i .

From calculus, the square formula reaches its minimum value when its derivatives equal zero. In the next derivation, we derive equations of the mean and the constant b of the line equation which must be close to the forecasted values. The summation boundaries are

from 0 to n. When we determine the derivatives of two parameters m and b, we get:

$$m = n \left(\frac{\sum_{i=0}^n xy}{\sum_{i=0}^n y} \right) - \left(\frac{\sum_{i=0}^n y}{n} \right) \left(\frac{\sum_{i=0}^n x}{n} \right)^2 - \left(\frac{\sum_{i=0}^n x}{n} \right)^2 \quad (11)$$

$$b = \left[\left(\frac{\sum_{i=0}^n y}{n} \right) - m \left(\frac{\sum_{i=0}^n x}{n} \right) \right] \quad (12)$$

Where:

- b: Constant
- y: Forecasted values on y-axis
- X: Values of time on x-axis
- m: Line slope
- n: Period of time

Fourier time series model: The fourth series type that we are studying of the Fourier series. The Fourier series is the series of the form^[3]:

$$X_t = A_0 + \sum_{n=1}^{\infty} (A_n \cos(nx) + B_n \sin(nx)) \quad (13)$$

The constants A_0 , A_n and B_n are called the coefficients of $F_n(x)$.

Where:

$$i = 0, 1 \dots n$$

$$A_0 = \frac{1}{n} \sum_{t=0}^{n-1} X_t$$

$$A_n = \frac{2}{n} \sum_{t=0}^{n-1} X_t \cos(W_j t)$$

$$B_n = \frac{2}{n} \sum_{t=0}^{n-1} X_t \sin(W_j t)$$

Where:

- X_t : The predicted value
- t: The time of forecasting
- W_j : The frequency

Descriptions of the simulated computer networks: In this section, we deal with two types of computer networks which are simulated according to Fig. 2. Their performance was tested by the previous four types of time series model. The first one is called a token ring computer network, while the second one is called a token bus. The ring topology is a physical, closed loop consisting of point-to-point links. Each node on the ring acts as a repeater. It receives a transmission from the previous node and amplifies it before passing it on. A token ring network was a special frame called a token that rotates around the ring when no stations are actively sending information. When a station wants to transmit on the ring, it must capture this token frame.

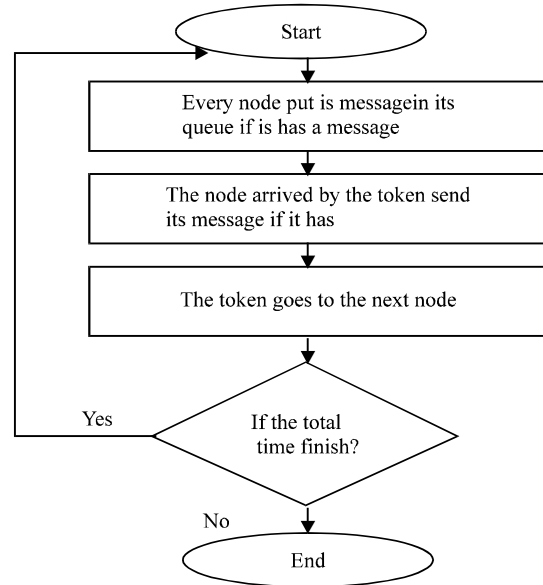


Fig. 2: Flowchart of operation for a token ring computer network

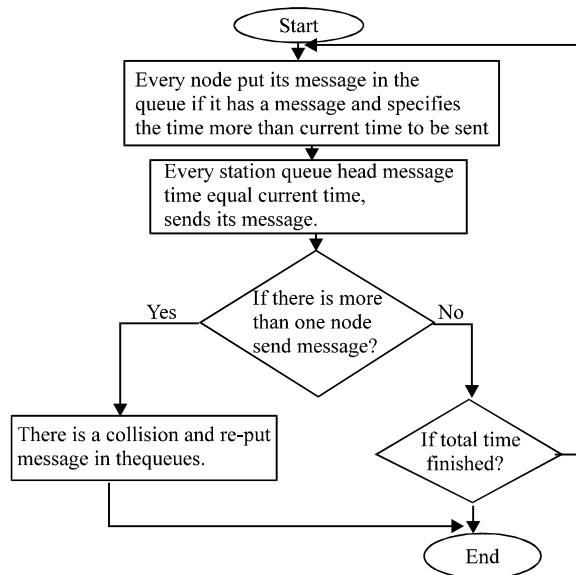


Fig. 3: Flowchart illustrating the operation of a token bus network

The owner of the token is the only station that can transmit on the ring. A flowchart that illustrates the operation of a token ring computer network is shown in Fig. 2^[4].

On the other hand, in a token bus network, all devices are attached to the same transmission medium. A flowchart that illustrates the complete concept of token bus operation is shown in Fig. 3.

Comparison between the various types of time series models: The data used in this experiment come from self-build a simulator for a token bus computer

communication network. This simulator produces a data for: network utilization, network delay, network collision frequency, each node collision frequency, and each node delay. These data are saved into a file where this experiment takes the first half amount of data for training and the remaining half used for testing. The experiment produced forecasted data for network utilization, network delay, network collision frequency, each node collision frequency, and each node delay from the actual data had been entered before. The data taken passed through four time series models where each one produced a separate different forecasting data to a different file. Figure 4 shows the standard deviation of the bus busy time for the token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA standard deviation are close to zero because the predicted value close to the actual values, while the LS and Fourier far. Figure 5 shows the relative error of the bus busy time for the token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA relative error is too small since the predicted value is close to the actual values, while the Fourier relative error is large relatively.

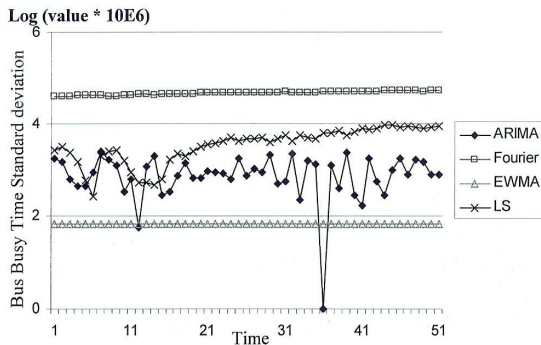


Fig. 4: Standard Deviation of bus busy time for the token bus network using four types of time series models

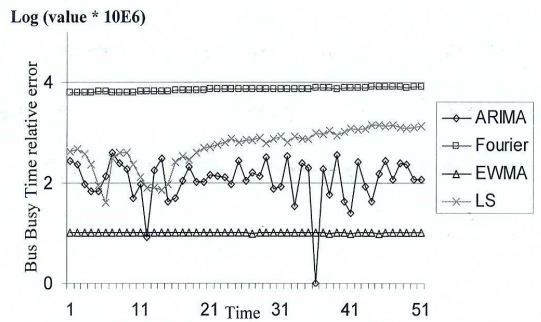


Fig. 5: Relative Error of bus busy time for the token bus network using four types of the series models

Figure 6 shows the standard deviation of network delay for the token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA standard deviation are close to zero because the predicted value close to the actual values, while the LS and Fourier far.

Figure 7 shows the relative error of the network delay for the token bus computer network using four types of time series models. From the Figure, it is seen that the ARIMA and EWMA relative error is too small since the predicted value close to the actual values, while the Fourier relative error is relatively large.

Figure 8 shows the standard deviation of the utilization of the token ring computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA standard deviation are close to zero while the LS and Fourier far.

Figure 9 depicts the relative error of the utilization of the token ring computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA relative error is too small while the Fourier relative error is huge relatively.

Figure 10 shows the standard deviation of the delay for node 0 token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA standard deviation are close to zero.

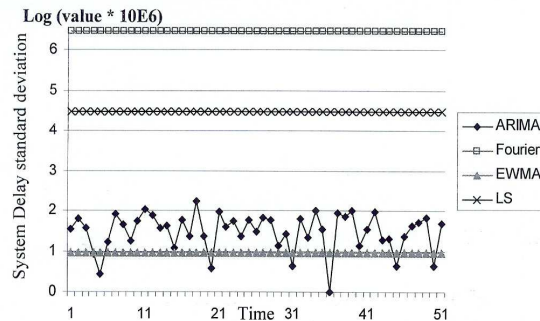


Fig. 6: Standard deviation of the network delay for the token bus computer network using four types time series models

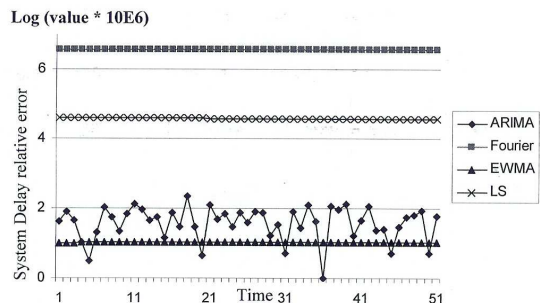


Fig. 7: Relative error of the network delay of the token bus computer network using four types of time series models

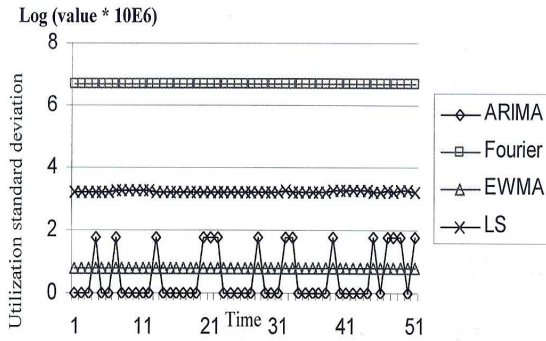


Fig. 8: Standard deviation of the utilization of the token ring computer network using four types of the series models.

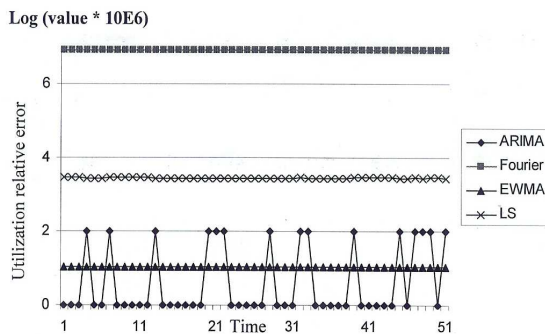


Fig. 9: Relative error of the Utilization of the token ring computer network using four types of the time series models

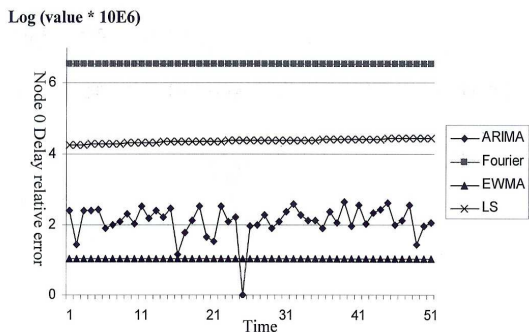


Fig. 10: Standard deviation of delay for node 0 token bus computer network using four types of time series models

Figure 11 depicts the relative error of the delay for need 0 token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA relative error is too small while the Fourier relative error is relatively large.

Figure 12 shows the Standard deviation of the collision frequency of the token bus computer network using four types of time series models. From the Figure, we see that the ARIMA and EWMA standard deviation are close to zero while the LS and Fourier far.

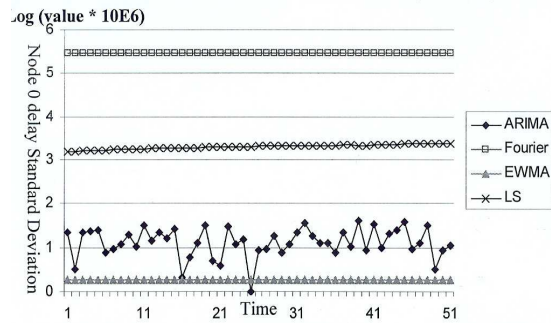


Fig. 11: Relative error of delay for node 0 token bus computer network using four types of time series models

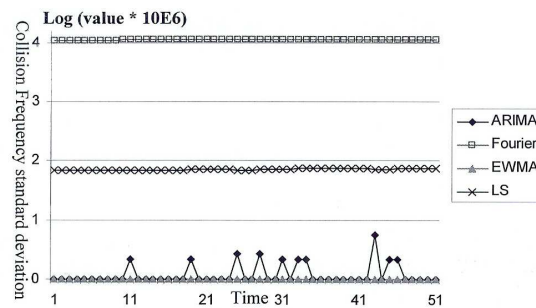


Fig. 12: Standard deviation of the collision frequency of the token bus computer network using four types of time series models

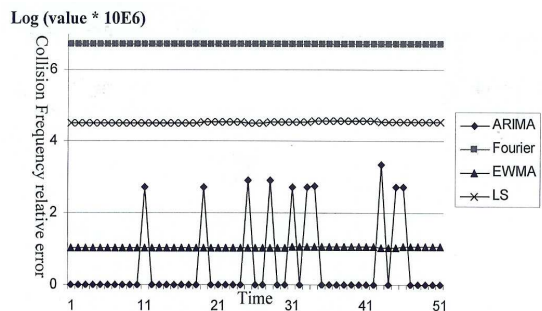


Fig. 13: Relative error of the collision frequency of the token bus computer network using four types of time series models

Figure 13 shows the relative error of the collision frequency of the token bus computer network using four types of time series models.

From the figure we can see that the ARIMA and EWMA relative error is too small while the Fourier relative error is relatively large.

We can see from the experiments obtained from Fig. 4-13 that, the EWMA is the best because its values of performance measurements are closer to the actual values than ARIMA, LS, and Fourier. ARIMA is better than the LS and Fourier and finally LS is better than Fourier. EWMA has four components constant, trend,

seasonal, and random. In EWMA, the constant which is always presented have the biggest effects to the total result of EWMA and updated periodically while the other components have small effect on the result. So, it goes through a steady state.

ARIMA consist of three models AR, I, and MA, each one computes one parameter and passed it to the next one. The change of actual values will affect on the ARIMA result time. LS is a line equation method which tries to find the slope and constant of the line equation that goes through the most of historical points. So, it will be affected by the far points in history. Finally, because the Fourier time series model has periodically shape, it will take the waveform and still progress over time as its start. In all models, the value factors affect the result of the time series and give those closest between them.

CONCLUSION AND FUTURE WORKS

In this work, we monitor the performance of both token ring and a token bus computer network. Any failure of network component will affect one of the network measurements (Utilization, system delay, total channel busy time ... etc.). If we know about abnormal changing in these measurements, this means that there is some problem in the network. We compare between four time series models (ARIMA, EWMA, LS, and Fourier) in order to choose the optimal one for estimating the performance of the above two types of computer networks. We found that the EWMA time series is the best one, next the ARIMA with small differences between them. This conclusion has been obtained as a result of making a statistical comparison between these four types of time series models.

As a future work, we suggest to use the EWMA time series model in building an Expert system can assist in putting a threshold value of the network performance parameters instead of putting it manually by the network administrator. Here the accuracy of management and administration will be raised to a noticeable level which can be reflected in raising the performance of a computer communication networks.

REFERENCES

1. Ahmad, B.H., 2005. On the application of time series models for detecting the problems of computer communication networks. Al-Balqa' Applied University- Jordan, Master Thesis.
2. William, S., 2003. Cryptography and Network Security Principles and Practice. 3rd Edn., Pearson Education International.
3. Malandain, E. and P. Skarek, An Expert System for Accelerator Fault Diagnosis.
4. William, S., 2004. Data and Computer Communications. 7th Edn. Pearson Education International.
5. William, S., 2002. High-speed networks and internets: Performance and quality of service. 2nd Edn., Prentice Hall.
6. Ioannis, V., N. Bassileiades, I.S.M. Molina, S.O.I. Futo, Z. Pasztor, J.S.I. Velbitskiyi, S. Yershov, S. Golub and I. Netesin. System Architecture of a Distributed Expert System for the Management of a National Data Network. Department of Informatics, Aristotle University of Thessaloniki, Department of Artificial Intelligence, Technical University of Madrid, International Software Technology Research Center Technosoft.