*https://dx.doi.org/10.22161/ijaers.6727* *ISSN: 2349-6495(P) | 2456-1908(O)*

# Traffic Forecasting for Monitoring in Computer Networks using Time Series

Henrique Dornel[1], Eliane da Silva Christo[1], Kelly Alonso Costa[2], Danilo Pinto Moreira de Souza[1]

[1]Postgraduate in Computational Modeling in Science and Technology, Fluminense Federal University, Volta Redonda RJ - Brazil
[2]Postgraduate in Production Engineering, Fluminense Federal University, Volta Redonda RJ – Brazil

***Abstract*—*With the expansion of connectivity and information exchange, monitoring Internet traffic becomes a priority in network management to identify anomalies and resource use. This paper presents a study of data traffic forecasting on a computer network, by using known approaching methods for Time Series analysis. The objective of this work is to monitoring the connection of users to network-based applications, including resource availability and network stability of a Brazilian educational institute. To estimate the traffic at a given time, the adjustments made with Exponential Smoothing, AR and ARIMA models were compared in order to detect possible future abnormal behavior of network usage. The results indicate that the chosen models, mainly the ARIMA, can be used to predict both input and output traffic of a network, also allowing the generation of alerts in real time. It is possible to predict how Internet traffic will be in the next few moments in order to detect possible anomaly on the network in a short period of time when they differ considerably from the forecast made for that specific period. Efficient network monitoring favors the quality of applications and services available to users, helping the network manager to make decisions for maintenance and constant improvement.***

***Keywords*—*Computer network, Information security, Time series analysis*.**

## I. INTRODUCTION

Internet access in many organizational environments has grown especially in the last decade with the expansion of network technologies. In an educational institution, for example, Internet and network resources are fundamental in several academic and administrative activities. The growth of Internet usage and the search for information in research bases requires an increase in the connection bandwidth by constantly expanding the data traffic in the network, which directly impacts on the performance of the services provided.

Furthermore, evil-wishing users can take advantage of security vulnerabilities and failures on the network or systems to practice intrusions and attacks, which can be prevented by using tools such as anti-malware applications, firewall and intrusion detection systems. Security in computer networks is a very important and increasingly studied subject, as for example in the work of Andreas Schilling [1], in which is implemented a framework for secure IT operations.

However, in order to recognize attack patterns, the traffic monitoring management is also very important, helping the network manager to make important decisions in critical cases, like unavailability of resources and inappropriate or unusual use of the network. In short, there is a strong need for monitoring computer networks properties in order to diagnose any problems and manage them in the best possible way, due to their expansion [2].

In order to monitor the network traffic, the Information Technology Management of our educational institute uses the Multi Router Traffic Grapher (MRTG), a free network monitoring software, which generates data charts collected from SNMP - Simple Network Management Protocol [3]. The charts provided by MRTG display the bit rate per second of the data traffic in the network at certain time intervals.

Forecasting of Internet traffic is also very important for tasks such as resource allocation, network planning and detection of network anomalies caused by attacks. An accurate prediction model can be used to detect security attacks in computer networks, by comparing predicted with actual traffic. [4] Moreover, predicting future traffic on a computer network, based on current traffic, allows the network manager to take measures before attacks, congestion, connection drops or downtime. Forecasts like that can be made by modeling the traffic of the data input and output of the network as Time Series. There are several studies in this area nowadays [2] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17].

Section 2 of this paper describes basic concepts about Time Series and forecasting methods. Section 3 presents the comparison of the experimental results of the series

https://dx.doi.org/10.22161/ijaers.6727 ISSN: 2349-6495(P) | 2456-1908(O)

adjustments corresponding to the traffic in the network of our educational institute, with different models of Exponential Smoothing, AR and ARIMA. Conclusions and final considerations are presented in Section 4.

## II. MATERIALS AND METHODS

### 2.1 Network characterization

In this paper is presented a study of forecasting of data traffic in the computer network of an educational institution. The analyzes were based on Internet traffic, input and output. In this network are constantly used various services and applications based on the Internet, available to students, employees and external community, equivalent, on average, to 300 different users per day. The connections can be wired or via Wi-Fi, in institutional or personal computers and mobile devices, with a total of 10 switches and 15 access points. The Internet link is provided by an outsourced company, with the connection being made through a single central router.

### 2.2 Time Series Forecast Models

Methods of approach by Time Series have been used by several authors in the study of forecast of data traffic in computer networks, such as Thanasis Vafeiadis, Alexandros Papanikolaou, Christos Ilioudis and Stefanos Charchalakis [2] and Sangjoon Jung, Chonggun Kim and Younky Chung [7], who proposed the use of AR, MA, ARMA and ARIMA models. Christos Katris and Sophia Daskalaki [4] studied predictions using ARIMA, FARIMA and Holt-Winters methods, and Vander Luiz Proena da Silva [8] applied Simple Exponential Smoothing, Holt and Holt-Winters methods. Other authors such as Renata Lussier Spagnol [9] and Roben C. Lunardi, Bruno L. Dalmazo, Erico M. H. do Amaral and Raul C. Nunes [10][11] wrote works where forecasting with various ARIMA models were used for the tra_c in the network. The ARIMA prediction method can also be used in many different applications, for example, many areas of Engineering and Econometrics, as studied by Ozden Ustun and Refail Kasimbeyli [18]. Landauskas, M., Navickas, Z., Vainoras, A. and Ragulskis, M. proposed an effective time series forecasting technique based on algebraic weighted moving averaging. The functionality and feasibility of the proposed short-time series forecasting technique is demonstrated by computational experiments with real-world time series [19].

A Time Series can be defined as an observation set of a particular variable ordered in time, usually equidistant [20], which have serial dependence, in other words, dependence between instants of time. A stochastic process can be classified as a model that describes the probability structure of a sequence of observations, that is, systems that evolve in time or space according to probabilistic laws. It can be said that a Time Series is the fulfillment of a stochastic process.

In order to model network traffic as a Time Series, the variable to be considered may be the rate of trafficked bytes, for example. Since this variable is random, the Time Series can be defined as a sampling of a stochastic process. However, most of the Time Series are not actually stochastic, and may present trend level variations. In this paper, the results obtained with the following prediction models were compared: Simple Exponential Smoothing, Double Exponential Smoothing (Holt Method), AR (Auto Regressive) and ARIMA (Auto Regressive Integrated Moving Average).

In this paper is presented a study of forecasting of data traffic in the computer network of an educational institution. The analyzes were based on Internet traffic, input and output. In this network are constantly used various services and applications based on the Internet, available to students, employees and external community, equivalent, on average, to 300 different users per day. The connections can be wired or via Wi-Fi, in institutional or personal computers and mobile devices, with a total of 10 switches and 15 access points. The Internet link is provided by an outsourced company, with the connection being made through a single central router.

Simple Exponential Smoothing (SES) [17], or Exponential Damping, is a well-known Time Series forecasting method. It is a simple algorithm used to predict the next value in a series, based on the current value and the current forecast. It consists of the idea of obtaining a weighted average, where the recent values have a greater weight than the older ones [8]. This technique can be used in series where no trend is observed. The adjustment P at the instant of time t + 1 is calculated as follows in (1).

$$P_{t+1} = \alpha D_t + (1 - \alpha) P_t \quad (1)$$

Where α is the smoothing coefficient ($0 \leq \alpha \leq 1$) and Dt is the actual value observed in time period t.

It can be said that the smoothing coefficient α regulates the speed with which the series fits the data [8]. For values close to 1, a rapid adjustment is achieved (when old effects have little impact on prediction, and recent effects have a high impact). And a slow adjustment (damping of old and recent effects) is obtained with values of α close to 0.

Double Exponential Smoothing (DES), or Holt Method, is a variation of the Simple Exponential Smoothing, used in series that present a linear increasing or decreasing tendency [8]. This method consists in

*https://dx.doi.org/10.22161/ijaers.6727* *ISSN: 2349-6495(P) | 2456-1908(O)*

obtaining an estimate of trend of the series, through the (2).

$$T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1} \quad (2)$$

Since β is the smoothing coefficient for the trend estimate ($0 \leq \beta \leq 1$) and N the level component, given by the (3).

$$N_t = \alpha D_t + (1-\alpha)(N_{t-1} + T_{t-1}) \quad (3)$$

Where α is the smoothing coefficient ($0 \leq \alpha \leq 1$) and Dt is the actual value observed in time period t. The prediction P at the instant of time t is calculated in Eq. (4), by seeking h periods of time ahead [8]:

$$P_{t+h} = N_t + hT_t \quad (4)$$

There are also other variations of Exponential Smoothing, such as the Holt-Winters Method, which consider the seasonal behavior of the Time Series. Nevertheless, with these methods it was not possible to obtain good adjustments for the series analyzed in this work. Therefore they will not be addressed here.

In auto regressive process AR(p) is given by (5) [9]:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t \rightarrow \phi(B)Z_t = a_t \quad (5)$$

Since $a_t$ is the white noise and $\phi_1, \phi_2, \ldots, \phi_p$ constants. This equation must satisfy certain conditions for the process to be stationary [9].

This method of approximation is based on the premise that each observation in a Time Series is related to one or more previous observations in the same series [17].

A moving average model MA (q) is a process in which the averages are adjusted according to the seasonal or cyclic components of the Time Series [17]. It is a weighted moving average, fixed-number model, in which the most recent value generally carries a weight greater than the farthest backward values. For a stationary Time Series, its average or the immediate past value can be used as a forecast for the future period. The process is given by (6).

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} = \theta(B)a_t \quad (6)$$

Where $a_t$ is the white noise and $\theta_1, \theta_2, \ldots, \theta_q$ are constants [9].

In an ARMA(p,q) model, for a stationary Time Series, both AR and MA terms are required, as follows in (7).

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \rightarrow \phi(B)Z_t = \theta(B)a_t \quad (7)$$

ARIMA Model (Auto Regressive "Integrated" Moving Average) – For series that present non-stationary behavior, it is necessary to perform a transformation to differentiate it. The model indicated to represent the series in this case is the ARIMA (p, d, q), where the parameter d indicates the number of differentiations that the series has undergone to become stationary (8).

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t \quad (8)$$

## III. RESULTS AND DISCUSSIONS

The data analyzed were generated by MRTG, and correspond to the input and output bit rates per second that trafficked through our educational institute Internet link. The period of time considered was two weeks, from June 19 to July 2, 2016. The samplings correspond to the rates measured every five minutes are shown in Fig. 1.

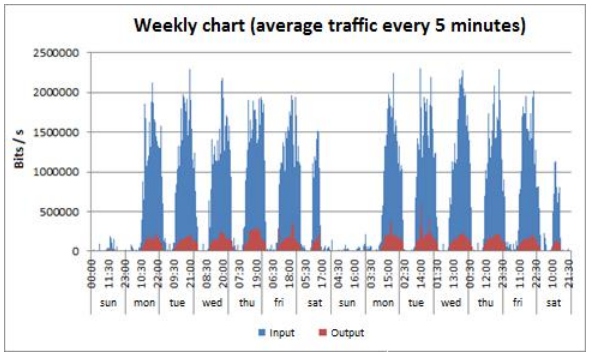


*Fig. 1: Internet traffic in the network.*

Figure 1 shows that in the period of time considered, Internet access is more intense between 9am and 11pm Monday to Friday, and Saturdays in the morning. It is noteworthy that in these two weeks no network usage were considered abnormal, nor did there occur significant drops in connection.

This information can now be used to try to predict how the Internet traffic in this network will be in the next few moments, through Time Series prediction methods. The objective is to make short-run forecasts to use them to detect possible abnormal traffic in the network in a short period of time, such as connection drops and abnormally intense traffic in the next few minutes, which have distanced themselves considerably from the forecast for that particular period.

The following are the results of the Time Series adjustments with the Simple Exponential Smoothing and Double Exponential Smoothing methods, corresponding to the bit rates per second of the input and output in the network, which are shown in Fig. 2, using Minitab1 software.

*https://dx.doi.org/10.22161/ijaers.6727* *ISSN: 2349-6495(P) | 2456-1908(O)*

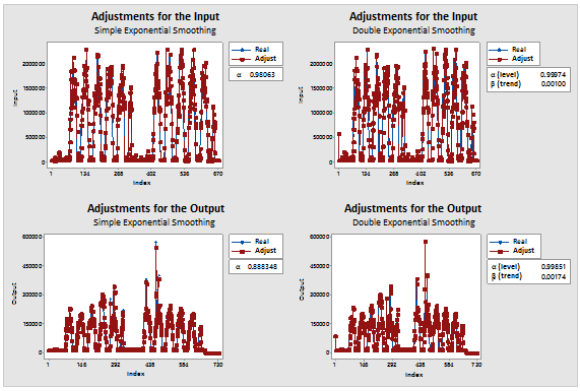


*Fig. 2: Charts of adjustment using Simple Exponential Smoothing and Double Exponential Smoothing.*

Figure 2 shows the charts that characterize the input traffic, both with the Simple Exponential Smoothing method and with the Double Exponential Smoothing. The obtained results indicate good adjustments with the real data and the models of the Time Series. But for the output traffic, better results were obtained with the Double Exponential Smoothing. The smoothing coefficients α and β were calculated automatically by the software.

To make predictions with AR and ARIMA models, it is first necessary to verify if the Time Series corresponding to the input and output traffic satisfies the stationarity condition. When a series is not stationary, it must be transformed through the differentiation process. If the differentiated series is stationary, the autocorrelation function (ACF) and partial autocorrelation function (PACF) are calculated, for then choosing the most appropriate model to perform the predictions for the series.

The procedure introduced in Fig. 3 was performed to identify the best model for series adjustments [7].

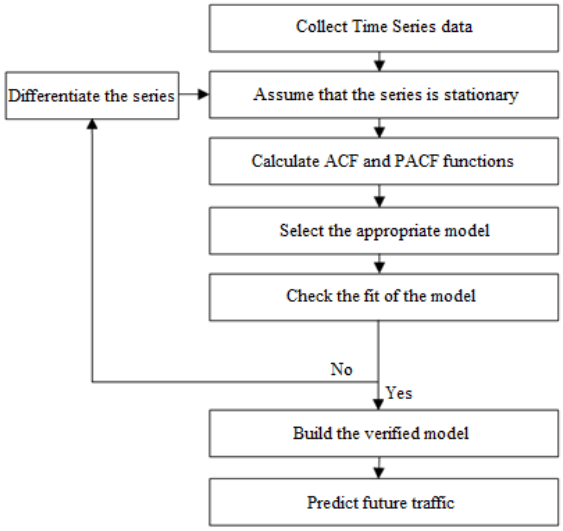


*Fig. 3: Procedure for identifying the best model for series adjustment.*

All of the results and charts in Fig. 4 were also obtained with Minitab software. Assuming initially the stationarity of the Time Series of the input and output traffic, the following autocorrelation and partial autocorrelation functions are obtained:

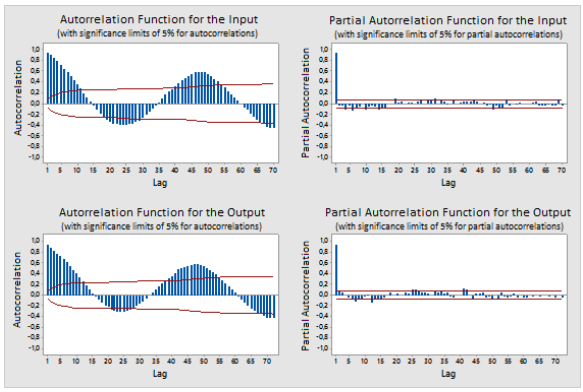


*Fig. 4: Autocorrelation and partial autocorrelation functions for the input and output series*

For both input and output traffic, the sinusoidal behavior of the autocorrelation function and a peak in the partial autocorrelation function suggest an AR(1) model, according to the adjustments obtained in Fig. 5.

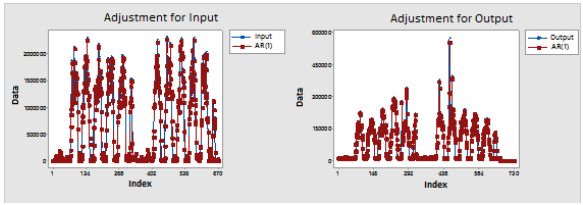


*Fig. 5: Charts of adjustments using AR(1) model*

As the AR(1) model did not _t the data so well, the Time Series analyzed were transformed through the differentiation process. The new ACF and PACF functions are in Fig. 6.

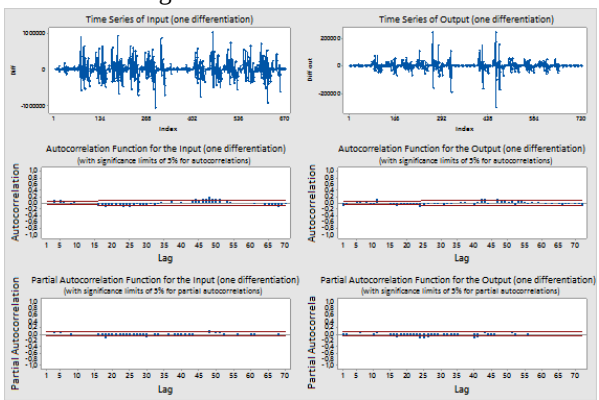


*Fig. 6: Autocorrelation and partial autocorrelation functions for the input and output series, with one differentiation*

*https://dx.doi.org/10.22161/ijaers.6727* *ISSN: 2349-6495(P) | 2456-1908(O)*

Once a single differentiation has been performed in the series, and according to the autocorrelation and partial autocorrelation functions, a possible model to be tested is ARIMA(1; 1; 1), which provides the following adjustments presented in Fig. 7.

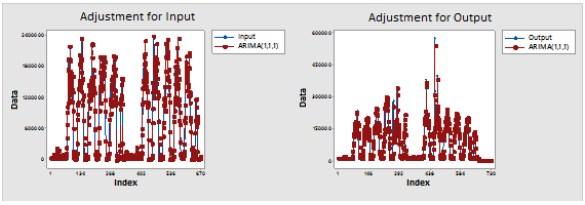


*Fig. 7: Charts of adjustments with ARIMA(1,1,1) model*

From the charts in Figure 7, it can be verified that with the ARIMA(1,1,1) model a better adjustment is obtained for the input series. However, the setting for the output series is still not appropriate. For this reason, a further differentiation was made in the two series, obtaining other ACF and PACF functions, according to Fig. 8.

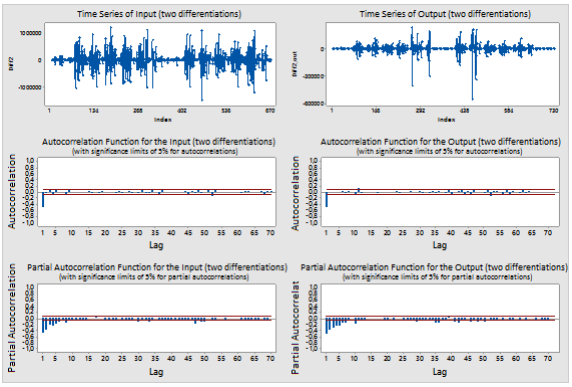


*Fig. 8: Autocorrelation and partial autocorrelation functions for the input and output series, with two differentiations*

The new autocorrelation and partial autocorrelation functions finally seem to indicate a more appropriate model. A peak in the ACF and the decreasing behavior of the PACF suggest an MA(1) model. As two differentiations were made in the series, the most appropriate model would be ARIMA(0,2,1). Fig. 9 shows the charts of this model. In this way, good adjustments are made both for input and for output traffic.

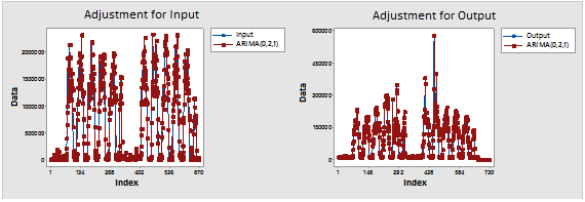


*Fig. 9: Charts of adjustments using ARIMA(0,2,1) model*

Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) were calculated in order to compare the results of the adjustments obtained with all models tested. The values obtained are shown in Table 1:

*Table 1: Comparison of the errors obtained from the adjustments obtained.*

| | | MAD | MAPE |
|---|---|---|---|
| **SES** | **Input** | 140,625.12 | 70.13% |
| | **Output** | 15,104.24 | 21.19% |
| **DES** | **Input** | 140,625.91 | 69.11% |
| | **Output** | 15,124.65 | 19.90% |
| **AR(1)** | **Input** | 146,448.44 | 144.67% |
| | **Output** | 17,064.37 | 37.53% |
| **ARIMA(1,1,1)** | **Input** | 141,240.61 | 74.82% |
| | **Output** | 15,620.86 | 22.29% |
| **ARIMA(0,2,1)** | **Input** | **140,421.06** | **69.01%** |
| | **Output** | **15,083.73** | **19.77%** |

With the results obtained, it was observed that it is possible to get good predictions with the application of the methods studied, since most of the models _t well to the traffic data in the network. The computational implementation of these methods is proven to be inexpensive and quite affordable compared to other methods used for the same purpose of this work. With the quality of these results, continuity of studies and the practical use of tools based on these methods will be possible.

## IV. CONCLUSION

This paper addressed the monitoring of network traffic in an educational institution through Time Series models to support the network manager in making decisions in critical situations such as unavailability of resources and inappropriate or unusual use of the network, besides helping in the planning and dimensioning of this network.

The Time Series approach models provided good adjustments for Internet traffic on the network. With the results achieved it is possible to carry out traffic forecasts with real-time alerts using the models studied.

According to the metrics considered, all models were adequate to be used in predictions of both input and output traffic, although AR and ARIMA models provide smaller errors. On the other hand, analyzing the charts shown previously, it can be concluded that the ARIMA(0,2,1) model is the one that provides a better fit for both Time Series studied, being therefore the most suitable model to estimate with more precision the future traffic on the network.

This paper presented a study of the best method to predict Internet traffic in a particular educational institute. As a continuation of this work, it is intended to use

*https://dx.doi.org/10.22161/ijaers.6727* *ISSN: 2349-6495(P) | 2456-1908(O)*

mechanisms and tools to compare expected traffic with current real traffic, in a dynamic and automated way, in order to provide the network manager with alerts and reports in case of discrepancies in the comparisons, which may indicate possible abnormal behavior of network traffic. Machine learning implementations will be used with the aim of eliminating false positive alerts, and it also to compare the methods studied here with other implementations, including applications with hybrid methods.

## ACKNOWLEDGEMENTS

The authors thank the Fluminense Federal University for its infrastructure and support.

## REFERENCES

[1] Schilling A. (2017). A framework for secure IT operations in an uncertain and changing environment, Elsevier - Computers and Operations Research, 85, 139-153. http://dx.doi.org/10.1016/j.cor.2017.04.008

[2] Vafeiadis T., Papanikolaou A., Ilioudis C. & Charchalakis S. (2012). Real-time network data analysis using time series models, Elsevier Simulation Modelling Practice and Theory, 29, 173-180. http://doi.org/10.1016/j.simpat.2012.07.002

[3] Kurose J.F. & Ross, K.W. (2000). Computer Networking: A Top-Down Approach, 3rd Ed. Brazil: Pearson Education.

[4] Katris C. & Daskalaki S. (2015). Comparing forecasting approaches for Internet traffic, Elsevier Expert Systems with Applications, 42(21), 8172-8183. http://doi.org/10.1016/j.eswa.2015.06.029

[5] Aragão Jr. J.B. & Barreto G.A. (2010). Novel approaches for online playout delay prediction in VoIP applications using time series models, Elsevier-Computers and Electrical Engineering, 36(3), 536-544. http://dx.doi.org/10.1016/j.compeleceng.2009.12.006

[6] Kalutarage H.K., Shaikh S.A., Wickramasinghe I.P., Zhou Q. & James A.E. (2015). Detecting stealthy attacks: Efficient monitoring of suspicious activities on computer networks, Elsevier Computers and Electrical Engineering, 47, 327-344. http://dx.doi.org/10.1016/j.compeleceng.2015.07.007

[7] Jung S., Kim C. & Chung Y. (2006). A Prediction Method of Network Traffic Using Time Series Models, Computational Science and Its Applications-ICCSA, 234-243. http://dx.doi.org/doi:10.1007/11751595_26

[8] Silva V. (2015). Identification of Network Flow Anomalies Using Time Series Predictions by the Holt-Winters Method, Masters dissertation. Postgraduate and Engineering Research - COPPE, UFRJ, Brazil.

[9] Spagnol R. (2011). Computer Networks Modeling by Statistical Methods, Master in Technology and Innovation, Campinas State University - UNICAMP, Technology College - FT, Limeira, SP, Brazil.

[10] Lunardi R. (2008). An Intrusion Analyzer based on Time Series, Graduation work 255, Computer Science Course, UFSM, Brazil.

[11] Lunardi R., Dalmazo B., Amaral E. & Nunes R. (2008). DIBSet: a Time Series Based Anomaly Intrusion Detector, VIII Brazilian Symposium on Information Security and Computational Systems, Brazil.

[12] Joshi M. & Hadi T. (2015). A Review of Network Tra_c Analysis and Prediction Techniques, School of Computer Sciences, North Maharashtra University, Jalgaon (M.S), India.

[13] Hu K., Sim A., Antoniades D. & Dovrolis C. (2013). Statistical Prediction Models for Network Traffic Performance, Scientific Data Management Research Group, Computational Research Division, Lawrence Berkeley National Laboratory; College of Computing; Georgia Institute of Technology.

[14] Hinich, M. & Molyneux R. (2003). Predicting Information Flows in Network Traffic, Journal of the American Society for Information Science and Technology, 54(2), 161-168. http://dx.doi.org/ doi:10.1002/asi.10176

[15] Santos A., Silva J., Silva L. & Sene M. (2011). Network traffic characterization based on Time Series Analysis and Computational Intelligence, Journal of Computational Interdisciplinary Sciences, Pan-American Association of Computational Interdisciplinary Sciences, 2(3), 197-205. http://dx.doi.org/doi:10.6062/jcis.2011.02.03.0046

[16] James C. (2011). Time Series Analysis of Network Traffic, IIT MADRAS

[17] Brutlag J. (2000). Aberrant Behavior Detection in Time Series for Network Monitoring, WebTV, Proceedings of the 14th USENIX conference on System administration (LISA 2000), pp. 139-146, New Orleans, Louisiana, USA.

[18] Ustun O. & Kasimbeyli R. (2012). Combined forecasts in portfolio optimization: A generalized approach, Elsevier Computers and Operations Research, 39(4), 805-819. http://dx.doi.org/10.1016/j.cor.2017.04.008

[19] Landauskas M., Navickas Z., Vainoras A. & Ragulskis M. (2017). Weighted Moving Averaging Revisited: An Algebraic Approach, Comp. Appl. Math, 36, 1545.

[20] Bowerman B.L. & O_Connel R.T. (1993). Forecasting and Time Series: an Applied Approach, Belmont. Duxbury Press.