Predicting Computer Network Traffic: A Time Series Forecasting Approach using DWT, ARIMA and RNN

Rishabh Madan
Amity School of Engineering and Technology
Amity University
NOIDA, INDIA
Email: rishabh.madan@student.amity.edu

Partha Sarathi Mangipudi
Amity School of Engineering and Technology
Amity University
NOIDA, INDIA
Email: psmangipudi@amity.edu

Abstract: **This paper proposes the Discrete Wavelet Transform (DWT), Auto Regressive Integrated Moving Averages (ARIMA) model and Recurrent Neural Network (RNN) based technique for forecasting the computer network traffic. Computer network traffic is sampled on computer networking device connected to the internet. At first, discrete wavelet transform is used to decompose the traffic data into non-linear (approximate) and linear (detailed) components. After that, detailed and approximate components are reconstructed using inverse DWT and predictions are made using Auto Regressive Moving Average (ARIMA) and Recurrent Neural Networks (RNN), respectively. Internet traffic is a time series which can be used to predict the future traffic trends in a computer network. Numerous computer network management tasks depend heavily on the information about the network traffic. This forecasting is very useful for numerous applications, such as congestion control, anomaly detection, and bandwidth allocation. Our method is easy to implement and computationally less expensive which can be easily applied at the data centers, improving the quality of service (QoS) and reducing the cost.**

***Keywords*—ARIMA; DWT; internet traffic; neural network; prediction; RNN; time series**

I. INTRODUCTION

Forecasting network traffic based on the past trends is a crucial step in controlling and optimally using a computer network. Traffic prediction is really useful in certain domains like, controlling the admissibility, routing decisions, allocating the bandwidth and other applications which are adaptive in nature. Information about the network traffic is really important configuration input for network providers for planning and managing a computer network, especially at the data centers [1, 2].

Network traffic can be best defined as the amount of data bits, collectively called packets, moving across devices at a given point of time. Due to the temporal nature, when it is sampled at uniform time intervals, data collected is a time series. The main aim in analyzing any time series and predicting the future values based on the old observations is mathematical model that is used. Numerous forecasting mathematical models have been developed over the years, Artificial Neural Networks (ANN) and Autoregressive Integrated Moving Average (ARIMA) are well-known and widely used in literature [3, 4]. ARIMA models showcase an extremely accurate forecasting ability for a few types of time series. Only drawback in using ARIMA model is that, the data needs to show the evidence of stationarity, that is, they are not suitable for non-linear time series modelling [3]. RNNs as compared to ARIMA are much better when, predicting the nonlinear time series.

Many methods have been proposed in literature for the forecasting of the network traffic. Based on the models utilized, these can be broadly classified into two classes: nonlinear prediction and linear prediction. The most commonly used models for linear prediction are: i) HoltWinters Algorithm [5] and ii) ARIMA model [5, 6]. The neural networks based forecasting is used widely for nonlinear prediction [7, 8].

Discrete wavelet transform (DWT) scoops out both location and frequency information from a signal. Wavelet transform when used in forecasting models can improve the accuracy of the forecast [9].

This paper proposes a time series forecasting technique that utilizes RNN and ARIMA for predicting the future computer network traffic. It models both linear and nonlinear structures of the network traffic time series. The linear and nonlinear structures are obtained after the decomposition using discrete wavelet transform (DWT). The real-world time series, like, the data traffic comprises both nonlinear and the linear patterns [3]. Proposed approach utilizes the decomposition power of DWT, and uses the prediction power of ARIMA and RNN for both linear and nonlinear structures, respectively. The test was made on samples of time series of internet traffic, obtained on DataMarket database.

II. ARIMA MODEL

The Autoregressive Integrated moving average is an efficient statistical model used for time series forecasting. It was pioneered by Box and Jenkins [10]. This model is derived from one fundamental principle that the future values can be predicted using the white noise characteristics and the past values. We can express an ARIMA (p, d, q) model mathematically as follows:

$$\Phi(L)\,(1-L)^d y^t = \theta(L)\,\varepsilon_t \qquad (1)$$

- Here, p, d and q ϵ Z^+, and can be referred to as the order of AR, I and MA parts of the ARIMA model.

- y_t refer to the input data, or, the points of observation at time t.
- ε_t refers to the white noise at any given time t.
- φ (L) are the lag polynomials, and L is the lag operator.
- Term'd' refers to the degree of ordinary differencing, it is used to make the time series stationary.

III. Recurrent Neural Networks

Given a set of data points with pre-defined outputs, the neural networks have a capability to quickly self-learn and self-adapt, this enables them in modelling and predicting the complex patterns of nonlinear nature. RNNs are one such set of neural networks which due to one or more connection between the neurons forms cycles. It is these cycles in RNN that store the data and pass one neuron's feedback to another, this mechanism builds an internal memory and this facilitates the learning of data which is sequential in nature.

While reading input is being read, a RNN allows information to be carried because it has loops. This makes them different than the other neural networks, the memory that they have helps them in finding correlations between events. These loops can be used in either direction and theoretically anywhere in the network [11].

The process of carrying the memory forward can be expressed mathematically as follows:

$$h_t = \Phi(W_{xt} + U\, h_{t-1}) \tag{2}$$

- h_t here refers to hidden state at any given time step t
- x_t represents the input at the same time step t
- W is a variable representing the weight matrix
- U refers to hidden-state-to-hidden state matrix

From eq. (2) it is clear that hidden state is function of weight matrix modification of the input added to product of hidden-state-to-state-matrix and hidden state at time t-1.

IV. The Proposed Network Traffic prediction technique based on DWT, ARIMA and RNN

Discrete wavelet transform (DWT) decomposes a given discrete signal into orthogonal wavelet functions. In case of 1D signals like a time series, results in a transformed vector of equal length. The vector is first filtered with a low pass filter and then a high pass filter. Mathematically, we can represent a DWT as follows:

$$\Psi_{l,m}(t) = 2^{-l/2}\, \psi(2^l t - m) \tag{3}$$

The terms l and m in above equation for DWT represent the scale factor and translation index, respectively.

In this proposed approach, DWT (db2 wavelet) has been used to decompose the network traffic time series into low frequency and high frequency components. In this approach two phases are utilized (i) decomposition and (ii) reconstructions [12]. Initially, DWT is applied on the time series data, this leads to generation of low frequency and high frequency components of the data, called Detailed (D) and Approximate (A). After this step, these two components are reconstructed by passing D and A through inverse discrete wavelet transform (iDWT). Once, the reconstructed components are generated, ARIMA model is applied to the D part to generate the forecasts. Then, RNN is fitted to the A part along with the residuals to obtain other set of forecasts. Finally, these forecasts are added component-wise to obtain the final forecasts. The approach is presented in Fig. 1. , elaborated in Alg.1.

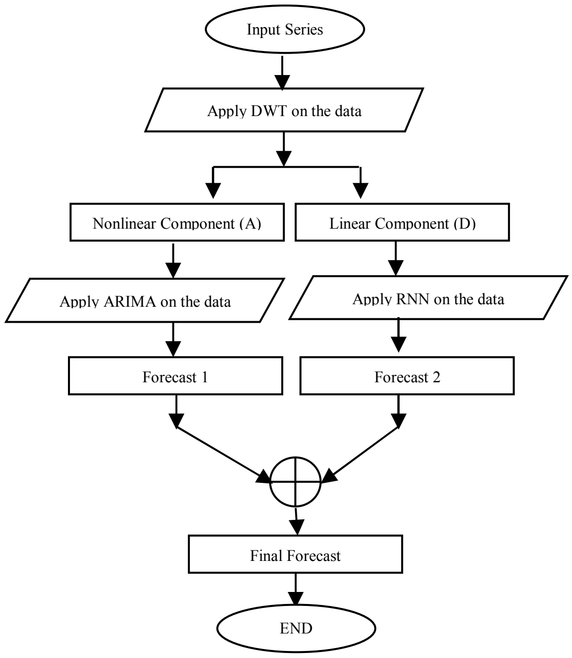


Fig. 1. Flow Chart showing the proposed technique for network traffic prediction

---

Algorithm 1. The proposed network traffic prediction technique based on DWT, ARIMA and RNN

Inputs: The network traffic time series dataset $\mathbf{X_{in}} = [x_1, x_2, \ldots, x_n]^T$ and size n of the testing dataset

Outputs: A forecast vector consisting outputs from both ARIMA and RNN $\mathbf{Y_{out}} = Y_{ARIMA} + Y_{RNN}$

Steps:

1. Load the network traffic time series into a vector $\mathbf{X_{in}}$
2. Apply DWT on the input variable and decompose it as follows: **cA**, **cD** = DWT ($\mathbf{X_{in}}$, 'filter')
3. Apply iDWT on the cA and cD to obtain the reconstructions: $\mathbf{X_a}$ = idwt(**cA**, 'filter') , $\mathbf{X_d}$ = idwt(**cD**, 'filter')
   //cD contains the details part and
   //cA contains the approximation parts

4. By comparing the Residual Sum of Squares (RSS) of AR, MA and ARIMA determine the appropriate **ARIMA (p, d, q)** model for $\mathbf{X_d}$.
5. **//Using ARIMA to estimate the Linear part**
6. $\mathbf{Y_{for1}}$ = forecast1('ARIMA model', $\mathbf{X_{d(eff)}}$) // $X_{d(eff)}$ is obtained by subtracting (p+d+q) from the training set
7. **//Using RNN to estimate the nonlinear part**
8. Set, $\mathbf{X_{NN}}$**(1: ($\mathbf{X_{in}}$ - p+d+q)) = U + (Xa(p+d+q+1:end))**
9. **//Loop to determine the appropriate RNN model for //$\mathbf{X_{NN}}$**
10. **i = 0**
11. **While(True):**
12. $\mathbf{Y_{for2}}$ = RNN_forecast(X_{NN})
13. **i++**
14. **if i == size($\mathbf{X_{in}}$)**
15. **Break**
16. **//Combining the forecasts from linear and the nonlinear part to obtain the final forecast for the data**
17. Obtain $\mathbf{Y_{out}} = \mathbf{Y_{for1}} + \mathbf{Y_{for2}}$

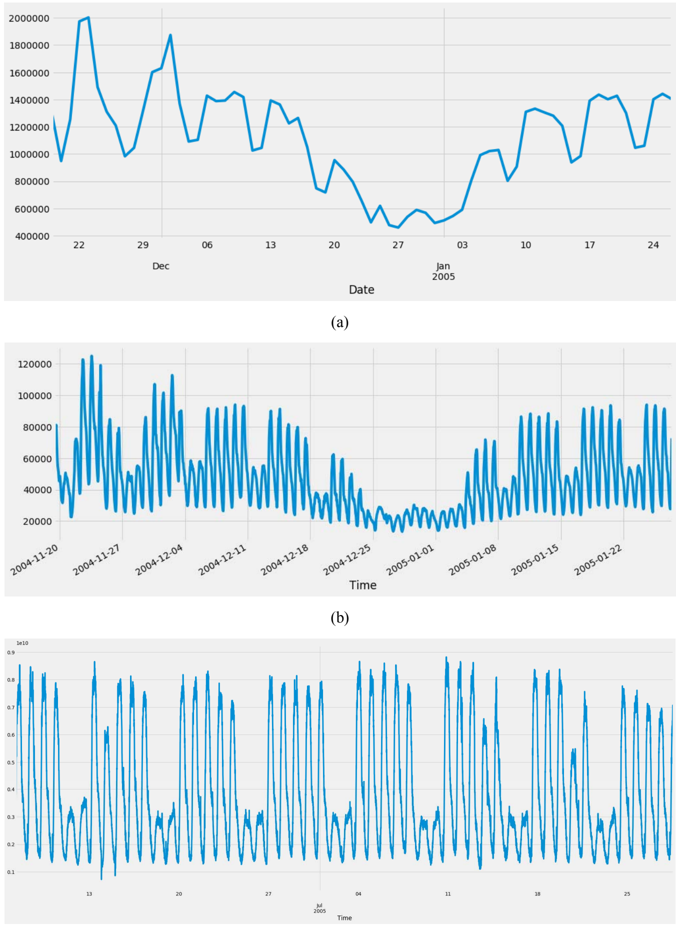
V. RESULTS

The experiments were conducted with 6 time series, created by R.J Hyndman [13]. The time series consisted of data recorded per hour, per day and per 5 minutes. The details about datasets have been described in Table 1 and depicted in Fig 2.

The time series that have been used, consists of internet traffic (in bits) as recorded by an ISP and it corresponds to transatlantic link. The experiments were conducted using Anaconda-navigator which were installed with various python libraries. The software package is open source.

TABLE I. DETAILS OF TIME SEIRES DATASETS

	Time Series Datasets		
	Name of the time series	*Total Size*	*Time Interval*
1.	Daily-1	51	1 day
2.	Daily-2	69	1 day
3.	Hourly-1	1231	1 hour
4.	Hourly-2	1657	1 hour
5.	5min-1	14772	5 minutes
6.	5min-2	19888	5 minutes



(c)

Fig. 2. Time plots of the time series: (a) Daily-2, data recorded per day, (b) Hourly-1, data recorded per hour, (c) 5min-1, data recorded per 5 min

After decomposing the time series using DWT, the linear and the nonlinear part were modelled using ARIMA and RNN respectively. Forecast obtained for each method was then added to get the final forecast. The forecasts obtained in the end through 'db2' wavelet. The results have been tabulated in Table II, it contains the Normalized Root Mean Squared Error (NRMSE) for each dataset. Mathematically NRMSE can be represented as:

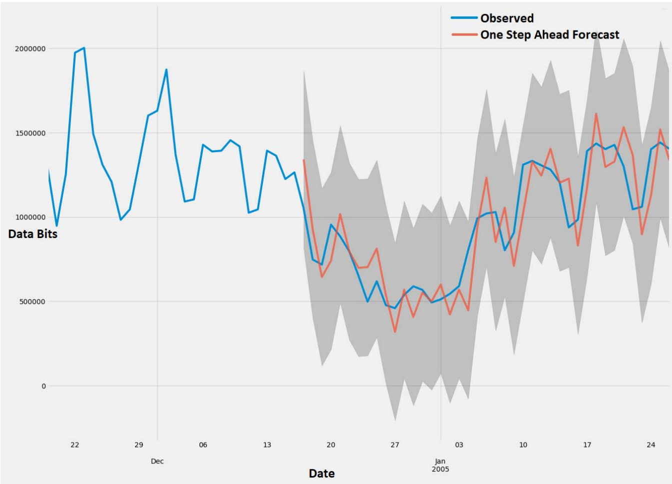
$$NRMSE = (MSE)^{1/2} / (y_{max} - y_{min}) \tag{4a}$$

where,
$$MSE = \sum_{i=1}^{n} e_i^2 / n \tag{4b}$$

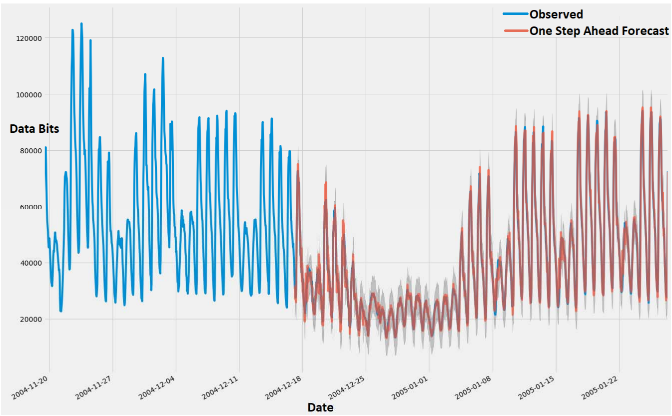
- Here, MSE stands for Mean Squared Error
- e is the error, calculated by taking the difference of corresponding true value and the forecasted value
- n is the total data points in the dataset.
- y_{max} and y_{min} refer to the maximum and the minimum values in the dataset respectively

TABLE II. NRMSE FOR THE SIX TIME SERIES

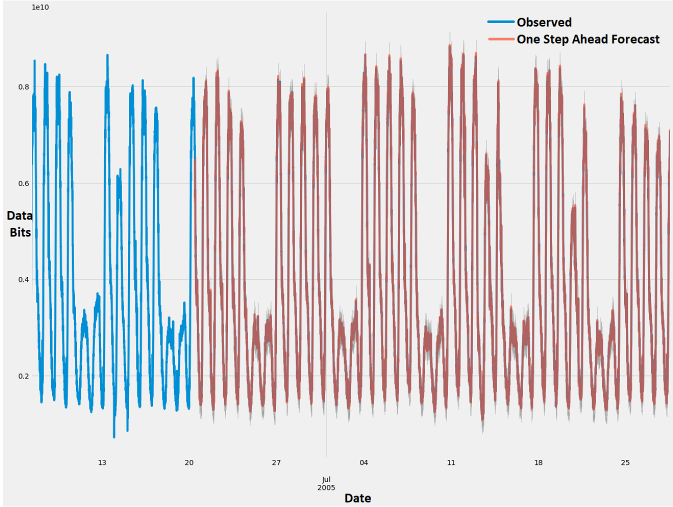
	NRMSE for different models			
			Proposed	
	Name of the time series	*RNN [14]*	*ARIMA*	*ARIMA + RNN*
1.	Daily-1	0.197	0.130	0.115
2.	Daily-2	0.116	0.240	0.191
3.	Hourly-1	0.041	0.030	0.022
4.	Hourly-2	0.027	0.050	0.24
5.	5min-1	0.016	0.020	0.012
6.	5min-2	0.009	0.010	0.008



(a)



(b)



(c)

Fig. 3. Time Series and its forecast using the proposed method: (a) Daily-2, data recorded per day, (b) Hourly-1, data recorded per hour, (c) 5min-1, data recorded per 5 min

VI. Conclusion

Internet traffic prediction is an important task when it comes to managing a computer network and avoiding traffic congestion. In this paper we proposed a time series forecasting technique to predict internet traffic based on past values using past values. Many forecasting techniques like ARIMA are used extensively in literature for making forecasts, but, it is useful mostly for a time series which is linear in nature. On the other hand, neural networks like RNN are very useful in forecasting time series which are nonlinear in nature. Proposed technique uses Discrete Wavelet Transform and using a high pass filter and a low pass filter producing linear and nonlinear parts for the time series. The proposed technique clearly outperforms ARIMA and RNN [14] (see Table II). And because of the simplicity of the technique it can be easily employed at data centers.

References

[1] Z. Hu, Y. Qiao, and J. Luo, "Coarse-grained traffic matrix estimation for data center networks," Computer Communications, vol. 56, no. 2, pp. 25-34, 2016

[2] T. Benson, A. Akella, and D.A. Maltz, "Network traffic characteristics of data centers in the wild," in Proceedings of ACM IMC, 2010, pp. 267-280

[3] Zhang, G.,P. "Time series forecasting using a hybrid ARIMA and neural network model". Neurocomputing vol . 50, p.p 159-175, 2003

[4] Zhang, G. P., Qi, M. "Neural network forecasting for seasonal and trend time series". European Journal of Operational Research 160 2005, pp 501-514.

[5] P. Cortez, M. Rio, M. Rocha, P. Sousa, Internet Traffic Forecasting using Neural Networks, International Joint Conference on Neural Networks, 2006, pp. 2635-2642.

[6] Felix A. Gers, Nicol N. Schraudolph, and Jurgen Schmidhuber, Learning precise timing with LSTM recurrent networks, Journal of Machine Learning Research , vol. 3, , Mar. 2003 pp. 115-143.

[7] V. B. Dharmadhikari, J. D. Gavade, An NN Approach for MPEG Video Traffic Prediction, 2nd International Conference on Software Technology and Engineering, pp. V1-57-V1-61. San Juan, USA, 2010.

[8] A. Abdennour, Evaluation of neural network architectures for MPEG-4 video traffic prediction, IEEE Transactions on Broadcasting, Volume 52, No. 2, . ISSN 0018-9316, 2006. pp. 184-192

[9] Al Wadia, M. T. I. S., Tahir Ismail, M. (2011). Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model. Applied Mathematical Sciences 5 (7), p.p 315-326.

[10] Box, G. E. P, Jenkins, G. M. Time series analysis, forecasting and control. 3rd ed. Holden-Day, California, 1970.

[11] Haykin, S. Neural Networks: A Comprehensive Foundation, 2nd ed., Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.

[12] Niu, C., Ji, L, "A hybrid method based on wavelet analysis for short term load forecasting." Journal of Convergence Information Technology[Online]. Available: https://pdfs.semanticscholar.org/3a4a/0d5824deeb28d1870fb6b26c7b8d7c13119e.pdf

[13] DataMarket.[Online] Available: https://datamarket.com/data/list/?q=provider:tsdl

[14] Tiago Prado Oliveira, Jamil Salem Barbar, Alexsandro Santos Soares Computer network traffic prediction: a comparison between traditional and deep learning neural networks, Int. J. Big Data intelligence, 2006, p.p 28-37