

Network Congestion Analysis using ARIMA Model and Bandwidth Cost Prediction

Ashutosh Shankhdhar¹, Shani Kumar Gupta², Mudit Mangal³, Varsha Varshney⁴, Sanjana Gupta⁵

GLA University, Mathura, Uttar Pradesh^(1,2,3,4,5)

Abstract

Analysis and prediction of telecom network traffic as well as computer network traffic has a good scope in extensive set of areas and possess significant number of studies. Congestion in the network is one of the major causes of performance degradation which effect the whole communication network. Network Congestion analysis and its prediction is a better approach to ensure safe and authentic network communication. Different techniques have been proposed and experimented for analysing the traffic in network. This paper applies Time Series analysis and used Arima model to analysis the congestion in the network and based on the analysis the model predicts the cost based on the required bandwidth which is the need of different telecom company. By using Arima model the system proposed weather the congestion in the network is sessional or any trend in the network congestion which helps us to find when the congestion in the network is high and when it is low so that the bandwidth will be provide those telecom company so that they increase the performance of the network.

Keywords— *Network congestion, Time Series Analysis, Arima model, Network traffic, Cost Prediction, Stationarity, Seasonality*

INTRODUCTION

Congestion in the network is one of the major causes of performance degradation, security attacks, adherence variance which effect the day by day communication problem [1]. Due to the complex nature of telecommunication network, congestion in the network and its main cause for network resources and hardware components has not been understood well. This leads to the problem of congestion in the network. Therefore, we want some fresh approaches and algorithm to understand and model congestion in the network as well as bandwidth cost prediction in order to improve the functioning of big-scale communication system.

When the people want to communicate with telecom devices sometime the data packet is lost in the communication medium due to the cause of hardware failure which generate the congestion in the network so due to the failure of hardware or any other components of network the data packet lost and the people phase many problems which effect the telecom company [2]. Our approach is based on using Time Series Analysis, analyze the real time data to prevent the congestion in the network using Arima model and based on that analysis we use different model to predict the bandwidth cost. So, using Time Series Algorithm we can find whether the network follow the sessional pattern or any other trends which help us to analysis the network traffic.

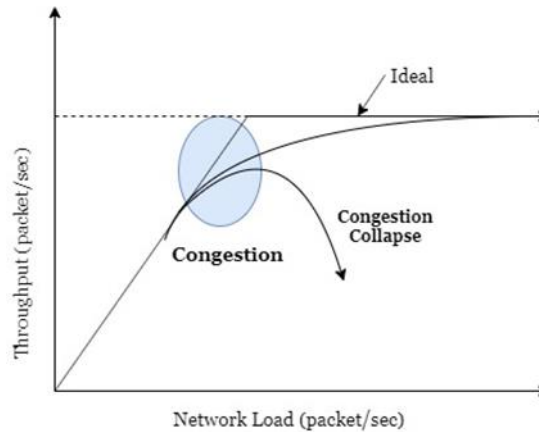


Fig 1. With too much traffic, performance drop sharply

RELATED WORK

In this digital era, communication is one of the medium to connect others but congestion in the network drastically create the problem in the communication so there are also many systems is already proposed to decrease or optimize the congestion in the network so that there will be less congestion. Prediction of network traffic has become an important and a debatable topic that has to be recently received much interest from computer network society. The prediction in the network traffic is one of the difficult issues but can be useful for monitoring network, can be useful for network security, to avoid congestion and enhance the speed of networks. Different techniques have been used by researchers for predicting the network traffic but we have categorized these techniques under four broad categories namely time series model which is further categorized by two types linear and non-linear, hybrid model and decompose model. Figure.2 gives a basic idea of four categories under which differential algorithms of predicting the network traffic has been categorized.

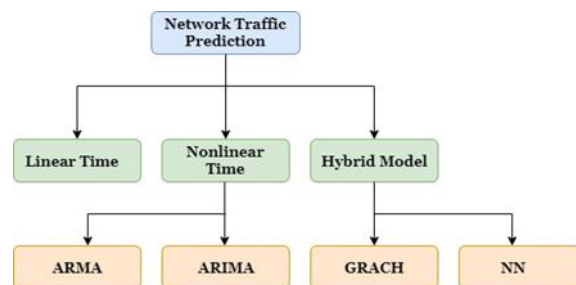


Fig-2. Techniques to determine network traffic

There are so many researches done in previous years. Because Network Traffic and congestion is the biggest problem so they applied the different forecasting method like Arima and Time series to find the congestion which helps us to identify the culprit in the network so by using this type of algorithm helps us to analyse the network on the day to day or time to time. Network congestion basically the problem generated in the time basics and most of the researches based upon to find and analyse the telecom as well as the computer network congestion analysis which helps to understand the behaviour of the network on the particular or specific time which identify the culprit behind the network congestion in the future.

Another paper in which for congestion detection is identifying through Artificial Neural Network to find the linear and non-linear relationships in data transmission so that to identify the culprit behind the network and identify the network congestion minimization factor.

PROPOSED SYSTEM

Congestion in network is the main cause of this era which increase the load on the network which effect the network speed, take so much time to upload and download the files from the network and increase the possibility of packet loss which create the problem for the user because user didn't get their message on time and so time the packet will collapse or lost so people phase the problem so we want this type of system which is used to analysis the network to get the information about the congestion most of the time the congestion in the network will increase in the special occasion because people will talk together with their relative. The load on the network will decide the traffic in the network so if we will denote load on the network as L_d and congestion in the network as C_n so we get the relation as between them is given below-

$$L_d \propto C_n$$

Which generally show if load increase in the network then the congestion will also increase.

In this paper basically we use the dataset which consists of the date and on the particular date the traffic raises through which firstly analyse the trend so that we can identify which type of trend basically followed by the dataset after analysis the trend we got the seasonal graph which means traffic depends upon the seasons or follow the particular patter which is shown in the below graph.



Fig.3. Seasonal pattern after analysis

So, in this paper basically we discussed the ARIMA model to analysis of the congestion in the network which is based upon Time series analysis because congestion in the network basically depend upon time. So, in the series of analysis of congestion in the network by the help of Time series analysis using ARIMA model after the analysis of trend we check the data is stationarity or not because Time Series has a particular behaviour, there is a high probability that it follows the same pattern in the future. To check stationarity, we use the Rolling mean and standard deviation method and after that analysis we got this graph:

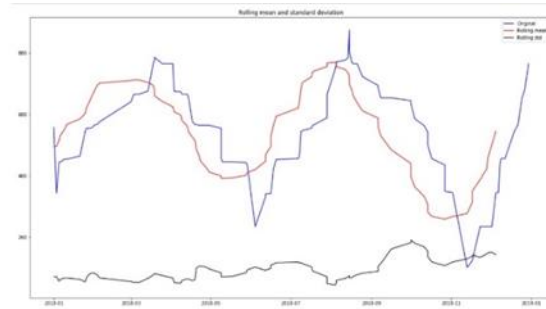


Fig.4. Stationarity using Rolling mean

If the time series forecasting is not stationary, we can transform it to stationarity data with one of the following techniques.

We can take the difference of the data. That is, given the series S_t , we create the new series

$$T_t = S_t - S_{t-1}.$$

The difference of the data will contain one min point than the main data. Although we can difference the data more than one, one difference is usually enough for the analysis.

Given a sequence $\{a_i\}_{i=1}^N$, an n -moving average is a new sequence $\{s_i\}_{i=1}^{N-n+1}$ defined from the a_i by taking the arithmetic mean of subsequences of n terms,

$$s_i = \frac{1}{n} \sum_{j=i}^{i+n-1} a_j.$$

So, the sequences S_n giving n -moving averages are

$$S_2 = \frac{1}{2} (a_1 + a_2, a_2 + a_3, \dots, a_{n-1} + a_n) \quad (2)$$

$$S_3 = \frac{1}{3} (a_1 + a_2 + a_3, a_2 + a_3 + a_4, \dots, a_{n-2} + a_{n-1} + a_n).$$

Through this analysis we got that our data follow seasonality trend that means congestion in network depends upon the different season which basically specify that congestion in the network vary time to time. After that analysis we analyse that is there any outlier which means that there is also the possibility that on the particular day there are more congestion in the network and we plot the AR model:

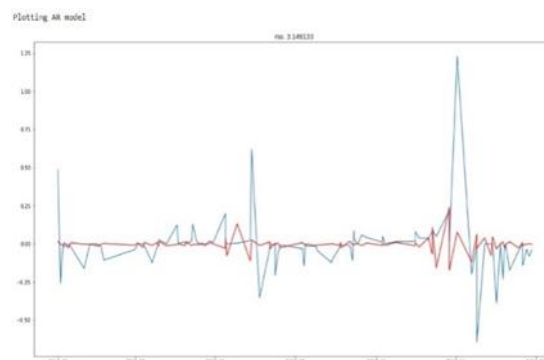


Fig.5. Finding Traffic using ARIMA Model

The forecasting equation is constructed as. Firstly, let t denote the n^{th} difference of T , which means:

$$\text{If } d1=0: t_t = T_t$$

$$\text{If } d1=1: t_t = T_t - T_{t-1}$$

$$\text{If } d1=2: t_t = (T_t - T_{t-1}) - (T_{t-1} - T_{t-2}) = T_t - 2T_{t-1} + T_{t-2}$$

In terms of t , the general equation of forecasting is:

$$M_t = \mu + \phi_1 M_{t-1} + \dots + \phi_p M_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

So, After this analysis using ARIMA Model we got that the congestion in the network follow the seasonality trend which help us to reduce the congestion in the network on the particular season so after that we predict the cost based on the required bandwidth so that we can reduce the congestion in the particular telecom service provider.

Cost Prediction based on the required Bandwidth

So, after analysing the data we get the network congestion depend on the season so if we get in which particular month the congestion in the network is high so we can provide the extra bandwidth to the telecom service provider so that they can reduce the congestion in the network.

In cost prediction we use the different variable like traffic in particular month, per bandwidth cost and bandwidth required by the service provider so we can formulate it by If the traffic in particular month will be T and per bandwidth cost will be P_B Then the cost C will be calculated as

$$C = T * P_B$$

So, through this formula we can calculate the bandwidth cost.

CONCLUSION

Network Congestion is one of the cause of network performance degradation due to the heavy load on network so in the series of analysis we use different type of algorithm and finally we Arima model which is related to Time series forecasting in which firstly check the stationarity of the data then using Arima model to identify the how much traffic we generated in particular time span so that the service provider will manage the extra bandwidth to reduce the congestion in the network so based on the traffic will identify the required bandwidth cost will be predicted so that the service provider can take the bandwidth through the other service provider and pay the required money so in this analysis traffic in the network will be identify and cost will be predicted.

ACKNOWLEDGEMENT

We want to thank by heart to our mentor Asst. Ashutosh Shankhdhar and GLA University, Mathura, India; for their encouragement, believe and support towards us to completion of this paper.

REFERENCES

1. C. Hyatt and D. P. Agrawal, "Congestion control in the wormhole-routed torus with clustering and delayed deflection," in *Parallel Computer Routing and Communication*, ser. Lecture Notes in Computer Science, S. Yalamanchili and J. Duato, Eds. Springer Berlin Heidelberg, 1998, vol. 1417, pp. 33–38.
2. A. Bhatele and L. V. Kale, "Quantifying network contention on large parallel machines," *Parallel Processing Letters*, vol. 19, no. 04, pp. 553–572, Dec. 2009.
3. J. Escudero-Sahuquillo, E. Gran, P. Garcia, J. Flich, T. Skeie, O. Lysne, F. Quiles, and J. Duato, "Combining congested-flow isolation and injection throttling in hpc interconnection networks," in *2011 International Conference on Parallel Processing (ICPP)*, Sept 2011, pp. 662–672.
4. Sachin Kumar Saxena, Dhaneshwar Kumar, Astha Sharma, Nikita Gupta, Radhika Kochhar, "Congestion Control in Wired Network for Heterogeneous Resource Using Neural Network", Uttarakhand India, Vol 3, 2013.
5. Tran Xuan Truong, Le Hung Lan, Nguyen Duy Vie and Mai Vinh Du, "Congestion Control in TCP/IP Differentiated Service Network Using Neural Network", *IEEE Seventh International Conference on Natural Computation* 978-1-4244-9953-3, 2011.
6. Anup Kumar Sharma, Ashok Kumar Behra "A Survey on Active Queue Management Techniques", *International Journal of Engineering and Computer Science (IJECS)* ISSN:2319-7242, <https://www.ijeecs.in>, Volume 5 Issue 11, Nov 2016, 18993-18997, DOI:10.18535/ijeecs/v5i11.47
7. H. M. A. El Hag, S. M. Sharif, An Adjusted ARIMA Model For Internet Traffic (IEEE, 2007) 978-1-4244-0987-7.
8. S. Fu-ke, Z. Wei, C. Pan, An Engineering Approach to Prediction of Network Traffic Based on Time Series Model (IEEE, 2009) 978-0-7695-3615-6.
9. T. Qin, X. Guan, W. Li, P. Wang, Monitoring Malicious Traffic Flow Based on Independent Component Analysis, (IEEE, 2009) 978-1-4244-3434-3.
10. Y. Yu, J. Wang, M. Song, J. Song, Network Traffic prediction and result analysis based seasonal and ARIMA and Correlation Coefficient (IEEE, 2010) 978-1-4244-8333-4.
11. S. Chabaa1, A. Zeroual1, J. Antari1, Identification and Prediction of Internet Traffic Using Artificial Neural Networks (scientific research, 2010) 2, 147-155.
12. A.R Syed, A.S.M Burney, B. Sami, Traffic Forecasting Network Loading Using Wavelet Filter and seasonal Autoregressive Moving Average Model (International Journal of Computer and Electrical Engineering, 2010) Vol.2, No.6.
13. Y. Shu, M. Yu, J. Liu, O.W.W. Yang, Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models (IEEE, 2003) 0-7803-7802-4.
14. M. Vijayakumar, R.M.S. Parvathi, Concept Mining of High Volume Data Streams in Network Traffic Using Hierarchical Clustering (Academic Journal, 2010) Vol. 39 Issue 2, p234.