International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Network Traffic Analysis and Prediction – A Literature Review

Huma Parveen
Computer Network
BBD University, Lucknow, India
humapiu@gmail.com

Rishi Srivastava
Computer Network
BBD University, Lucknow, India
rishi.bbdu@gmail.com

***Abstract*--The analytical study and forecasting of network traffic has wide applications in many areas of technology and has led to many researches in the current field. Various experiments and analysis has been carried out in the field of computer network applications. The analysis and forecasting of network traffic is a means of reliable and secure network communication. Many techniques have been proposed for network traffic congestion analysis like soft computing, regression, etc. The present study is a survey of various works carried out in the field of network traffic analysis and prediction. The works of various authors has been investigated and their findings has been summarised.**

Keywords: Network traffic analysis, Network traffic Prediction, Regression, Soft computing techniques.

1. Introduction

It is very well known fact that network traffic analysis is vital from the point of view of attaining proper information security. This is further confirmed by the fact that many e-commerce, banking and business related confidential information are carried via these networks. Network traffic analysis and prediction resembles a proactive approach rather than reactive, where network is monitored to ensure that security breaches do not occur within network. Network traffic analysis and prediction plays a very active role in the development of a fool proof congestion avoidance scheme. This further leads to the finding of normal and abnormal packets. These network decongestion schemes help in distributing the network bandwidth with regard to the predicted traffic.

The network traffic forecasting plays a major role in areas of dynamic bandwidth allocation, network security and network planning and predictive congestion control, etc. We can identify two categories of predictions: long and short period's predictions. Traffic prediction for long period gives a detailed forecasting of traffic models to evaluate future capacity requirements, and therefore permits for more minute planning and better decisions. Short period prediction (milli-seconds to minutes) is linked to dynamic resource allotment. It can be used to improve Quality of Service (QoS) mechanisms as well as for congestion control and for optimal resource management. This can also help in packet routing. Various techniques, like soft computing, data mining, time series regression analysis, artificial neural network, etc. has been used in the network traffic analysis and forecasting. The present work is all about review of various techniques in the network traffic congestion control and prediction. The works of various authors has been summerised and analysed. The remaining paper is organized as follows. Section two deals with an extensive review of several available network analysis works, followed in the last section by conclusion.

2. Review of Work Done

Understanding the characteristics of Internet traffic has been a challenging research topic for more than a decade. A deep knowledge of the underlying dynamics of Internet traffic plays an important role in order to offer better quality of service. Studies done on earlier Ethernet data have shown that network traffic possessed properties similar to the second-order self-similar processes, alternatively defined by wide sense stationary processes. This result led to numerous study on the bursty nature of individual TCP which is a constituent of the aggregate traffic. Extensive amount of work has been done to explain the self-similar nature of Internet traffic, along with network traffic congestion analysis and prediction.

Ming Zhang and Yanhong Lu, (2015) proposed an adaptive network traffic prediction algorithm based on BP neural network. Simulation results showed that, compared with the traditional BP neural network, the present algorithm has better performance in the prediction results, and has smaller error.

Manish, P. Ganvir, Dr. S.S.Salankar, (2015) proposed a time-series prediction model for the packet loss rate (PLR). They showed that prediction of PLR is very much useful in congestion control mechanisms. They used an artificial neural network as a prediction model and it is trained with Particle swarm Optimization (PSO) as a training algorithm in order to get accurate prediction of packet loss rate. They found that the quality of real-time multimedia traffic can be improved by accurate prediction thereby reducing the congestion.

C. Narendra Babu and B. Eswara Reddy, (2015) studied about the suitability of different methods for better Internet traffic data. This suitability of hybrid ARIMA-ANN models is studied for both one-step ahead and multi-step ahead prediction cases. They found that for one-step ahead prediction, with a forecast horizon of 10 points and for three-step prediction, with a forecast horizon of 12 points, the moving average filter based hybrid ARIMA-ANN model

gave better forecast accuracy than the other compared models.

Manish R. Joshi et. Al.(2012) did the survey and analysis of many network traffic prediction techniques. The individuality and rules of preceding studies were looked into. They have also summed the earlier works done in the field on network traffic analysis and prediction. For this they surveyed the previous studies of network traffic analysis and enlisted and discussed various approaches proposed to analyze and prediction of network traffic including data mining techniques, neural network and component analysis, and linear and nonlinear time series models.

Samira Chabaa, Abdelouhab Zeroual, Jilali Antari, (2010) did the analysis of the network traffic over IP network by developing ANN model using multi layer perceptron. For this network response was evaluated by using ANN and further analyzing the time series of network data. The results so obtained led to the conclusion that the ANN model using LM algorithm can be very well used for network traffic prediction and can be applied as an excellent and fundamental tool for the management of the internet traffic at different times

Andreas Petlund, Pal Halvorsen, Pal Frogner Hansen, Torbjorn Lindgren, Rui Casais, Carsten Griwodz, (2012) presented a dataset – a real-world, server-side packet trace – from Anarchy Online. They presented statistics from the network traffic and showed that it is a representative dataset for similar games. The data so presented was very useful in giving vital informations like loss rate, packet rates, etc.

Anukool Lakhina, et.al., (2005) said that one can consider anomalies as events which leads to traffic feature distribution alteration. Thus this sort of anomaly treatment leads to considerable diagnostic power, for the detection of fresh anomalies, their structure and classification. Further showed that entropy is an effective metric to capture unusual changes induced by anomalies in traffic feature distributions.

Hyun Cheol Cho, M. Sami Fadali, and Kwon Soon Lee,(2008) led to the development of a new AQM technology using dynamic ANN for TCP congestion control. They showed that ANN acts as a controller to keep track of original queue siza very much similar to the reference target. The neural network is trained by a BP algorithm. They further applied the neural AQM to a single bottleneck network supporting multiple TCP flows. Four scenarios were examined in the simulation experiments to compare neural AQM to RED and PI-based AQM. While PI AQM resulted in queue saturation and larger overshoot, neural AQM reduced overshoot and eliminated saturation. Neural AQM was more stable with no packet loss due to congestion. Especially for the case of time-varying TCP dynamics, the neural AQM was superior. They concluded that neural AQM is an effective adaptive controller and provides higher Quality of Service (QoS) in TCP networks.

Konstantina Papagiannaki, Nina Taft, Zhi-Li Zhang, Christophe Diot, (2003) developed a method to predict when and where link upgrades/additions should take place in the hub of an IP network. For this aggregate demand was measured between any two neighbouring PoPs in the hub of a major tier-1 IP network, and was analyzed for its evolution at time scales greater than one hour. They showed that the derived time series exhibit strong periodicities at the cycle of 12, and 24 hours, as well as one week. Moreover, they experienced variability at multiple time scales, and feature distinct overall long-term trends. Using wavelet MRA, they cut off the *overall long* term drift, and analyzed inconsistency at manifold time scales. They also showed that the largest amount of variability in the signal comes from its fluctuations at the 12 hour time scale. Their analysis indicated that a parsimonious model consisting of those two identified components is capable of capturing 98% of the total energy in the original signal, while explaining 90% of its variance. The resulting model is capable of revealing the behaviour of the network traffic through time, filtering short lived events that may cause traffic perturbations beyond the overall trend.

Sun Guang, (2013) used wavelet analysis and Hopfield neural network for network traffic forecast. This formed the basis for research on network traffic prediction model. The simulated results showed that the model is able to improve the prediction precision, and has the good flexibility to the network.

Han Song, Luying Gan (2015), led to the analysis that non interference with network traffic leads to network traffic congestion and paralysis. In order to change the chaotic situation, one needs to control the network traffic. According to the traffic forecasting technology, one can get the relevant changes of the network traffic. They applied the ant colony algorithm to optimize the gray model. Then, proposed an improved IAC-Gray algorithm. Finally they proposed the improved IAC-Gray method. The experiment showed that the method has the higher forecasting accuracy and it is an effective network traffic prediction model.

Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa, (2006) analyzed the effectiveness of numerous forecasting approaches when applied to TCP/IP traffic. These include three different *Time Series Forecasting (TSF)* methods: the Holt-Winters, the ARIMA methodology and a *Neural Network Ensemble (NNE)* approach. Data from large *Internet Source Providers (ISP)* were collected and analysis was done using different forecasting varieties, may be real time (every five minutes) and short-term (hourly aggregate values). A time series forecasting comparative study showed that in general the NNE produces the lowest errors. The NNE results revealed promising performances. For the real-time forecasts, only a 1-3% error was obtained for the five minute look ahead forecasts, a value that increases to 11.17% when the forecasts are issued two hours in advance. Turning to the short-term predictions, the error goes from 3.5% (one hour ahead) to 12.23% (24 hour lookahead). Thus, the proposed approach opens room for producing

better traffic engineering tools and methods to detect anomalies in the traffic patterns.

Faculty of Computing and Informatics, Multimedia University (2012), gave evidence which suggest that the Bit Torrent network traffic can be comprehended and hence can be short term forecasted using the ARMA model. The BitTorrent data showed that Bit Torrent seed can be either cyclic or seasonal. Also showed that other network traffic activities can be arranged for different time periods when the model is predicting a low Bit Torrent network traffic. This will improve the users' application bandwidth usage with lesser occurrences of network congestion. The research analysis led to the conclusion that BityTorrent network traffic can be both cyclic as well as seasonal. ARIMA TSF can be used for short term prediction analysis of BitTorrent network. It was seen that ARMA (2, 1) is good for cyclical BitTorrent network traffic patterns and ARMA (3, 0) is suited for seasonal BitTorrent network traffic patterns

Yasir Shoaib, Olivia Das, (2015) modelled the workload intensity of FIFA World Cup website using ANN. They employed Artificial neural networks for time-series prediction of two data sets: day-requests and day 6 and day 66-part10 of epoch requests. In total, 13 cases have been studied and compared. One base network is used and subsequent cases include modifications to this network's structure, data distribution (training/validation/test), training mode and/or input delays. Based on the results when data distribution for training, validation and simulation is not varied and when batch training is employed a network with two-hidden layers and 10 hidden layer size showed the best performance. This network has shown to model the requests intensity with reasonable accuracy, as seen from the MSE and correlation coefficient.

Changsheng Xiang; Peixin Qu, Xilong Qu, (2015), showed that the modern large scale network is affected by many factors, so the network flow has periodic, chaotic changes characteristics, the traditional network traffic prediction model is difficult to accurately describe the change trends of network traffic. Thus in order to improve the prediction accuracy of network traffic, a prediction model of network traffic based on MK-SVR is proposed by them. The experimental results showed that the proposed model can accurately describe the change trend of network traffic, improve the prediction accuracy of network traffic, the prediction error is reduced, providing a new modelling tool for complex network traffic prediction..

Zhou Xian, Li Rui, Huangfu Wei, Long Keping, (2013), projected WPFNN as a method to forecast network traffic. They extended the thoughts of wavelet transform to wavelet packet transform (WPT) in order to obtain more precisely partition in the high-frequency part of the original traffic and fuzzy neural network (FNN) is also implemented here for better prediction performance on coefficients.

Nelson Piedra, Janneth Chicaiza, Jorge Lopez, Jesus García,(2014) led to the development of ANN methodology for network resource management and network traffic congestion control. Thier findings led to the conclusion that once properly trained, ANN can better handle non- linear data relationship between the input and output patterns. Thus the present methodology can better tackle the network traffic engineering problems, and be able to accomplish the objective of acting as intelligent agents for allowing data flow according to the available resources. They analyzed the opportunity and feasibility to apply Artificial Neural Networks to a number of tasks related to Traffic Engineering.

Jarosław Bernacki, et.al.(2015) proposed methods for detecting untypical situations in a network traffic and evaluated it. The proposed method of security level evaluation using modified Exponential Moving Average with subjective logic opinions showed how the anomaly detection and time series analysis can be used to detect and to classify security related problems in computer networks. Also presented results of experiments and their statistical analysis showed that forecasting Brown's exponential smoothing is efficient and can be used for detecting abnormal situations in, for example computer networks. Brown's forecasting method can be very useful in real world networks as it is a light-weight method of data analysis.

Tao Peng and Zhoujin Tang,(2015) proposed a nonlinear prediction algorithm based on a relevant local LSSVM regression model to predict small scale network traffic and Pattern Search is applied to forecasting model parameters optimization LSSVM forecasting models structure is fixed and easy to regulate. The relevant local prediction method selects training set from all the historical data to ensure that self-similarity of the training set and filters out irrelevant data of test set. Using a validation sample and PS method makes the algorithm automatically select forecasting model parameters before predicting test data. They demonstrated the great effectiveness and efficiency of both prediction technique and parameter optimization in contrast to existing methods, and showed that the prediction error mainly concentrates on the vicinity of zero.

3. Conclusion

In the last decade, the analysis and prediction of network traffic has become a subject of continuous research in various sub-fields of computer networks. Innumerable number of researchers have been implemented an effective network traffic algorithm for the analysis and the prediction of network traffic. In this paper, the previous studies of network traffic analysis has been surveyed. Various approaches proposed to analyze and prediction of network traffic including data mining techniques, neural network and component analysis, and linear and nonlinear time series models has been enlisted and analysed.

References:

[1]. Ming Zhang and Yanhong Lu, (2015), " Adaptive Network Traffic Prediction Algorithm based on BP Neural Network", International Journal of Future Generation Communication and Networking Vol. 8, No. 5 (2015), pp. 195-206.

[2]. Manish, P. Ganvir, Dr. S.S.Salankar, (2015), "Time Series Forecasting of Packet Loss Rate Using Artificial Neural Network Based on Particle Swarm Optimization", International Journal of Engineering Research and General Science Volume 3, Issue 2, pp. 466-472.

[3]. C. Narendra Babu and B. Eswara Reddy, (2015), "Performance Comparison of Four New ARIMA-ANN Prediction Models on Internet Traffic Data", Journal of Telecommunication and Information Technology, pp. 67-75.

[4]. Manish R. Joshi et. Al.,(2012) "A Review of Network Traffic Analysis and Prediction Techniques", pp. 1-22.

[5]. Samira Chabaa, Abdelouhab Zeroual, Jilali Antari, (2010), " Identification and Prediction of Internet Traffic Using Artificial Neural Networks", *J. Intelligent Learning Systems & Applications*, 2010, 2, 147-155.

[6]. Andreas Petlund, Pal Halvorsen, Pal Frogner Hansen, Torbjorn Lindgren, Rui Casais, Carsten Griwodz, (2012), "Network Traffic from Anarchy Online: Analysis, Statistics and Applications", ACM 978-1-4503-1131-1/12/02, pp. 1-6.

[7]. Anukool Lakhina, et.al., (2005), "Mining Anomalies Using Traffic Feature Distributions", SIGCOMM'05, August 21–26, 2005, Philadelphia, Pennsylvania, USA., pp.1-12.

[8]. Hyun Cheol Cho, M. Sami Fadali, and Kwon Soon Lee,(2008), "Internet Traffic Control Using Dynamic Neural Networks", Journal of Electrical Engineering & Technology, Vol. 3, No. 2, pp. 285~291, 2008 285.

[9]. Konstantina Papagiannaki, Nina Taft, Zhi-Li Zhang, Christophe Diot, (2003), "Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models", IEEE INFOCOM, pp. 1-11.

[10]. Sun Guang, (2013), "Network Traffic Prediction Based on the Wavelet Analysis and Hopfield Neural Network", International Journal of Future Computer and Communication, Vol. 2, No. 2, April 2013.

[11]. Han Song, Luying Gan (2015), "The Research on the Prediction of the Network Traffic Based on the Improved IAC-Gray Method", *CHEMICAL ENGINEERING TRANSACTIONS, VOL. 46, pp. 1297-132.*

[12]. Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa, (2006), "Internet Traf_c Forecasting using Neural Networks", 2006 International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, IEEE, pp. 4942-4949.

[13]. Faculty of Computing and Informatics, Multimedia University,,(2012), "BITTORRENT NETWORK TRAFFIC FORECASTING WITH ARMA", International Journal of Computer Networks & Communications (IJCNC) Vol.4, No.4, pp.143-156.

[14]. Yasir Shoaib, Olivia Das, (2015), "Modeling Website Workload Using Neural Networks", arXiv:1507.07204v1 [cs.DC], pp. 1-25.

[15]. Changsheng Xiang; Peixin Qu, Xilong Qu, (2015), "Network Traffic Prediction Based on MK-SVR", Journal of Information & Computational Science 12:8 (2015) 3185–3197.

[16]. Zhou Xian, Li Rui, Huangfu Wei, Long Keping, (2013), "A Novel Network Traffic Prediction Methodology with On-line and Off-line Solutions", Advances in information Sciences and Service Sciences(AISS) Volume5, Number3, pp. 419-429.

[17]. Nelson Piedra, Janneth Chicaiza, Jorge López, Jesús García,(2014), "STUDY OF THE APPLICATION OF NEURAL NETWORKS IN INTERNET TRAFFIC ENGINEERING", International Book Series "Information Science and Computing", Advanced Research in Artificial Intelligence, pp. 33-47.

[18]. Jarosław Bernacki, et.al.,(2015), " Anomaly Detection in Network Traffic Using Selected Methods of Time Series Analysis", *J. Computer Network and Information Security*, 2015, 9, 10-18.

[19]. Tao Peng and Zhoujin Tang,(2015), "A Small Scale Forecasting Algorithm for Network Traffic based on Relevant Local Least Squares Support Vector Machine Regression Model", Appl. Math. Inf. Sci. 9, No. 2L, 653-659.