Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Brief Review of Recent Advances in Internet Traffic Modelling

Siddhesh S. Kulkarni
Independent Researcher, Pune,
Maharashtra, India

Abstract—Various characteristics of Internet traffic traces can be given as, dynamism, flexibility, emergence of new technology and applications generating varied amount of traffic.Traffic characterization and modelling constitute important steps towards understanding and solving performance-related problems in future IP networks Classification, analysis and modeling of this traffic with such volatility in its characteristics is research challenge to analysts, technocrats. This paper is an attempt to give brief overview of recent advances in the field of internet traffic modeling.

Keywords— Aggregate Internet Traffic, Self Similarity, Heavy Tailed Distribution, Long Range Dependence, Quality of Service

I. INTRODUCTION

Authors of [1] point out that current internet traffic can be viewed to have two key constituents, namely Web+ and P2P+; Web+ traffic consists of traffic from both Web 1.0 and Web 2.0 applications; P2P+ traffic consists largely of traffic from P2P applications and other non-Web applications excluding applications on well-known ports such as FTP and SMTP. Further they argue that the immense popularity of new-age "Web 2.0" applications such as YouTube, Flicker, and Facebook, and non Web applications such as Peer-to-Peer (P2P) file sharing, Voice over IP, online games, and media streaming have significantly altered the composition of internet traffic with respect to what it was a few years ago. Classification, analysis and modeling of this traffic with such volatility in its characteristics are some of research challenges to analysts, technocrats.

Quality of Service is the key term in any business. Traffic characterization and modelling constitute important steps towards understanding and solving performance-related problems in future IP networks. The central idea of traffic modelling lie in constructing the models that capture most important statistical properties of underlying measured trace[2], [6]. This analysis in turn helps to improve the quality of service.

Analysis of IP traffic shows that the data set exhibits characteristic properties: Heavy tail [10], Long Range Dependence (LRD), self similarity. This evidence on existence of LRD in traffic measurement can be found in Crovella and Bestavros [3] , Crovella et al.[4] ,Leland et al.[5] .The article [1] shows that recent traces still exhibits these traffic characteristics. Further this paper shows that that both traffic of Web+ and P2P+ components exhibit self-similar behavior and can be approximated by Poisson Process at smaller time scales.

Authors of [5] point out that the fractal behaviour of internet traffic is different from conventional telephone traffic. In turn this cannot be modelled using conventional formal models for packet traffic for e.g. Poisson or Poisson related models such as Poisson Batch Model, Markov Modulated Poisson processes, Packet train models, fluid flow models etc.

The problem of accurately capturing these characteristic properties in aggregate traffic models has been solved for non analytically tractable models for example fractional Gaussian Noise (fGN) and fractional autoregressive integrated moving average model (FARIMA) which captures burstiness as well as self similarity[7]. As referenced in [6], Ledsma and Liu reported effective construction of FGN in [11]. Construction and Fitting of analytically tractable models is still a subject of interests of many researchers.

There are extensive reviews are available on modelling of self similar data and analysis [8], [9]. But looking at the rapid development in this fertile field, periodic review is necessary. This is an attempt to give a brief review of recent advances in modelling aggregate internet traffic data. It does not include all the papers on the given topic but may give an idea about recent developments in the field.

II. LITERATURE REVIEW

In this section some papers in the recent past has been reviewed, which are helpful to give insight of developments in the field of IP, internet traffic modelling.

As mentioned above internet traffic traces follow heavy tail distribution. Markov models due to their efficiency to give precise and numerically stable computation of many result measures, are used in performance and dependability analysis. Base of Markov modeling is the exponential distribution. According to authors of [12], members of the exponential family of distributions like Weibull , log normal and Pareto distributions that can also capture a possibly heavy tailed behavior but cannot be analyzed analytically or numerically. Considering this as a lacuna, authors of [12] have fitted phase type distributions to internet traces. In [12] it is pointed out that Phase type (PH) distributions have some

advantages: models using phase type distribution can be analyzed analytically and numerically which is not the case in most other distributions. Various efficient algorithms are available for fitting of PH distributions. Approaches on fitting the parameters to match the coefficient of correlation are presented in the paper. As correlation in traces is hard to analyze, authors have observed that there is a need of more research to match correlations by Markov model. Limitation of the paper is, it only considers parameterization of Phase type distribution and not analysis of resulting model.

Generating synthetic data traffic, which statistically resembles its recorded counterpart, is one of the main goals of network traffic modeling. In [13] authors have addressed the problem of designing a generative model for arbitrary network source traffic, with focus on multivariate stationary random processes. By proposing Transformed Gaussian Autoregressive Moving Average Model (TARMA), authors have presented a framework enabling joint representation of distributions, autocorrelation and cross correlation of multiple processes. For each random process three statistical measures are considered, namely, the distribution, the auto-correlation function and the cross-correlation function with other processes. TARMA approach facilitate for decoupling of problem into three sub problems one for each statistical measure. More advantages, pointed out by authors that are: its complete analytical tractability, parsimoniousness in the number of model parameters. Generation of samples also exhibits low complexity. Authors have fitted this model on various traces and have obtained satisfactory results.

As mentioned in introduction Quality of Service (QoS) is now backbone of every service. Delay and jitter are the two crucial parameters, of QoS which are if not met can degrade quality of service at user level to unacceptable level. [14] Addresses problem of QoS. They analyze delay and jitter in networks with traffic as Interrupted Poisson Process (IPP). Authors have considered delay and jitter analysis in two cases like one node and N-nodes. In N-node case, tandem network is analyzed and a model for end to end jitter is proposed. It also shows that end to end jitter depends on phase probabilities of the incoming process and determines nature of that correlation. Results derived in the paper lead to fast computations for limit cases which are according to author of [14], useful for estimation of parameters of QoS in real life network.

Article [15] studies parameters of QoS for Voice over Internet Protocol (VoIP) traffic. This paper shows that for VoIP traffic jitter can be modeled by short range dependence as well as long range dependence. Authors have presented a methodology for simulating packet loss. Also the authors have found relationship between Hurst Parameter with packet loss rate. These relationships are modeled using power law function characterized by three fitted parameters. A methodology for simulating packet loss on VoIP is presented. According to the authors these models can be used in future for (i) design a de-jitter buffer, (ii) to implement a synthetic generator of VoIP jitter data traces, where the synthetic jitter data traces can be used as test vectors to carry out the performance evaluation of a de-jitter buffer of VoIP system, and (iii) design effective schemes for packets.

Self Similarity is a characteristic of recent traffic traces. Hurst parameter is a good summarizer of important self similar scaling property. The article [16] compares various methods of estimation of Hurst parameter. This study is conducted on laboratory generated trace data. Authors have highlighted important features as well as differences of these methods. On the basis of this recommendation, the authors have given some recommendations for practical applications of these methods.

Usage of Hidden Markov Model (HMM) in simulation of internet traffic is done by the authors in [21]. The authors have used the Bellcore trace data, which are divided into training and testing data set. Training data set is used for the training of HMM model parameters. This model is tested on testing data set. Later this model is used as a source of Self similar data.

As technologies have developed and internet has become a part of each cell phone, various studies have beenperformed to study internet traffic on cell phones. Below is a review of couple of such papers is presented.

In [17], spatial and temporal dynamics of internet traffic are studied. For this study, a week-long aggregated flow level mobile device traffic data collected from a major cellular operator's core network data is used. Further the authors have studied traces at two different dimensions: device type and application that generates such traffic patterns. After the analysis of this traffic traces the authors have suggested two models for the data: a Zipflike model to capture the volume distribution of application traffic and a Markov model to capture the volume dynamic of aggregated internet traffic. One of the key contributions of this paper is `Modeling Dynamics of Network Traffic'**:** Since due to vast technological advancements, different types of devices show different traffic behavior. Hence to accommodate all such heterogeneous changes, authors have extended aggregated model by customizing it for different types of devices to improve prediction theory. For that authors have used unsupervised clustering algorithm.

Paper [18] also addresses the cellular data traffic modeling. Authors of the paper have addressed modeling aspects of the 3G cellular data. They have tried to incorporate vast data of a million subscribers over thousand base stations into their study. An individual specific study of each subscriber has been carried out, by observing their behaviors and usage. Base station specific study is also conducted for significant temporal and spatial variation in different parts of the network. This study revealed how the network wise resources have been used. According to the authors, this study is the first major study in measurement analysis of subscriber and network behavior in a large scale 3G data network.

Article [19] addresses the internet traffic modelling problem on WIFI traffic in Public Hotspot. Server side measurements from several Wi-Fi hotspots deployed in a nationwide network over different types of venues from small coffee shops to large enterprises are used to highlight differences in traffic volumes and patterns. The authors have examined characteristics of the traffic observed at a large number of public Wi-Fi hot-spots deployed in two large metropolitan cities in terms of arrival counts and temporal variations, connection durations, byte counts, etc. The different venues were categorized into some business types, such as coffee shops and fast food chains, book stores and hotels, enterprises, and

examined the salient differences among these venue types. One of the key contributions of this paper is the usage of statistical clustering algorithms and Poisson regression to fit non-stationary Poisson models to arrival counts. The models were validated against test data. The authors have obtained impressive results. This paper also suggests an innovative method to model heavy tail distributions. It suggests that modeling the logarithm of a heavy tailed distribution with phase type distribution gives better match for size and shape. Distribution of the number of simultaneously present customers from an Mt/G/1 queuing model using a novel regenerative argument is obtained which is transparent and avoids the customarily made assumption of the queue starting empty at an infinite past. The authors have validated all the results and the predictions that are obtained by fitting models on training data with testing data.

III. CONCLUSION AND FUTURE WORK

.Authors of [12] suggest that, would be interesting to compare the results of phase type distributions if they are used as parts of larger dependability models. Furthermore, the distributions can also be used in simple queuing models to analyze the performance under failure or the performability. In [18]; the authors have given a nice insight of 3G network usage. This data can be modeled using some advance probabilistic models. Results of [19] have generated the possibility of embedding statistical methods of higher level in the context of stochastic process. Also providing theoretical support for the method developed to model heavy tail distribution can be a future work as suggested by the authors.

Advancement of technology is going to pose new questions for modeling, managing and predication of network data. Due to the dynamic nature of internet traces, these models should be tested and updated regularly. Hence this is certainly challenging and a fertile research field for multidisciplinary scientists. This paper is an attempt to provide a brief review of the advancements in the past 5 years. Due to limitations, I cannot include all the papers. Researchers may refer [20], For more a detailed reviews.

REFERENCES

[1] Himanshu Gupta, AnirbanMahanti, Vinay J. Ribeiro(2010)**,***Revisiting coexistence of Poissonity and self-similarity in Internet traffic Modelling*, Analysis & Simulation of Computer and Telecommunication Systems, 2009. MASCOTS '09. IEEE International Symposium on

[2] G.R Ash, Traffic Engineering and QOS method for IP-,ATM- & TDM- Based Multiservice Networks, internet Draft draft –ietf –tewg-qos-routin-01.txt 2001

[3] M.Crovella, ABestavros, *Self similarity in World Wide Web traffic: evidence and possible causes*. In Proceedings of the 1996 ACM SIGMETRIC international conference on measurement and Modeling of computer systems (1996) vol.24 pp.160-169

[4] M.Crovella ,A.Bestavros and M.S. Taqqu , *Heavy Tailed Probability Distribution in the World Wide Web*. A practical guide to Heavy Tailes: Statistical Techniques for Analysis of Heavy Tailed Distributions, R. Adler , R. Feldman and M.S. Taqqu (eds) (Bikhauser, Boston 1998) pp. 3-76

[5] W.E. Leland, M.S.Taqqu, W.E Willinger and D.V. Wilson, *On Self Similar Nature of Ethernet traffic*, ACM/SIGCOMM Computer communication Review (1993) 183-193

[6] Alexander Klemm, Christopher Lindemmann, Marco Lohmann, *Modeling IP traffic using the batch markovianarraival process*, Performance Evaluation 54(2003) , 149-173

[7] D.L Jagerman, B. Melamed, W.Willinger, *Stochastic Modeling of traffic process in: Frontiers in queuing: Modeling method and problems*, CRC press, Boca Raton , 1996

[8] W. Willinger,M.S.Taqqu and A.Erramili, *A bibliographical guide to self similar traffic and performance modeling for modern high speed network in: F.P Kelly, S.Zacharyand I.Ziedins(Eds.)*,Stochastic Network: Theory and Applications, Clarendon Press(Oxford University Press) ,Oxford (1996) 336-366

[9] Abdelnaser Adas, *Traffic Models in Broadband Networks*, IEEE communication magazine (1997), 82-88

[10] A Feldmann and W. Whitt. *Fitting mixtures of exponential to long tail distributions to analyze network performance models*, In proc. IEEE, infocom, Kobe, Japan, April 1997

[11] S.Ledesema, D. Liu, *Synthesis of fractional Gaussian Noise using linear approximation for generating self similar network traffic*, Comput. Commun. Rev 30(2000) 4-17.

[12] Peter Buchholz, Jan Kriege, *Markov Modeling of Availability and Unavailability Data*, Proc. of the Tenth European Dependable Computing Conference - EDCC 2014, IEEE, 2014

[13] Markus Laner, Philip Svoboda and Markus Rupp Parsimonious, *Network Traffic Modeling By Transformed ARMA Models*, IEEE Access, Volume 2 2014

[14] Adnan Huremovic, Mesud Hadzialic, *Delay and Jitter Network with IPP Traffic: Theoretical Model*, Journal of Networks, VOL. 9, NO. 12, December 2014

[15] HomeroToral-Cruz, Al-Sakib Khan Patha, Julio C. Ramírez Pacheco, *Accurate Modeling of VoIP Traffic QoS Parameters in Current and Future Networks with Multifractal and Markov Models*, Algorithms and Architectures for Parallel ProcessingVolume 7017 of the series Lecture Notes in Computer Science pp 255-265

[16] Cheolwoo Park, Felix Hernandez Campos , Long Le, J. S. Marron, Juhyun Park, VladasPipiras, F. D. Smith, Richard L. Smith, Michele Trovero, Zhengyuan Zhu, *Long Range Dependence Analysis of Internet Traffic*, Journal of Applied Statistics 38 (7), 1407-1433

[17] M. Zubair,ShafiLusheng,JiAlex X., LiuJia Wang, *Characterizing and Modeling Internet Traffic Dynamics of Cellular Devices* , SIGMETRICS '11 Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems , Pages 305-316

[18] Utpal Paul, AnandPrabhu Subramanian, Milind Madhav Buddhikot, Samir R. Das ,*Understanding Traffic Dynamics in Cellular Data Networks*, CONFERENCE PAPER in PROCEEDINGS - IEEE INFOCOM · MAY 2011

[19] AmitabhaGhosh, Rittwik Jana, V. Ramaswami, Jim Rowlandand N. K. Shankaranarayanan , *Modeling and Characterization of Large-Scale Wi-Fi Traffic in Public Hot-Spots*, INFOCOM, 2011 Proceedings IEEE, pp.2921 – 2929

[20] Carla Di Cairano-Gilfedder, Richard G. Clegg, *Adecade of internet research: advances in models and practices*, BT Technology Journal, 2005 - Springer

[21] Joanna Domańska, Adam Domański, Tadeusz Czachórski, *Internet trace source based on Hidden Markov Model*, Smart Spaces and Next Generation Wired/Wireless Networking Volume 6869 of the series Lecture Notes in Computer Science pp 395-40

[22] Salvo Rossi, P, Romano, G. ; Palmieri, F. ; Iannello, G. , *Joint end-to-end loss-delay hidden Markov model for periodic UDP traffic over the Internet* , Signal Processing, IEEE Transactions on (Volume:54 , Issue: 2)

