A Packet Distribution Traffic Model for Industrial Application: A Case of BIUST Network

Thato Solomon, Adamu Murtala Zungeru, Rajalakshmi Selvaraj, and Mmoloki Mangwala

Abstract—**Mathematical modelling is by and large comprehended as the way toward applying mathematics to real world problems with a perspective comprehension of the last mentioned. This paper presents mathematical model used for modelling real world problems using Botswana International University of Science and technology (BIUST) network traffic as a case. Sophisticated analysis of data is done to model the BIUST network with the succor of statistics, as it implies the collection and interpretation of data through mathematical processes called stochastic processes. From the attained results, the model and estimation of packet traffic distribution for BIUST Network based on Pareto distribution, it was perceived that about 20% of the users had about 80% of the bandwidth consumed.**

Index Terms—**Mathematical modelling, mathematics, mathematical models, mathematical processes, stochastic processes, packet traffic distribution, Pareto distribution.**

I. INTRODUCTION

Network traffic modelling is used as the basic for the network applications and for capacity planning of network systems. Given the impact of poor choices in this arena, it is clear that the validity of the underlying models is of critical importance [1], [2]. They are a wide number of mathematical models that could be used to model network traffic depending on the type of network to be modelled. The factors used to evaluate a network are taken directly from the underlying traffic model.

Statistics is concerned with making inferences about the way the world is, based upon things we observe happening. Nature is complex, so the things we see hardly ever conform exactly to simple or elegant mathematical idealizations as the world is full of unpredictability, uncertainty, randomness. Probability is defined as the language of uncertainty, and so to understand statistics, we must understand uncertainty, as probability and statistics work hand in hand [3]. Sophisticated analysis of data will be done to model the BIUST network with the aid of statistics, as it involves the collection and interpretation of data through mathematical processes called stochastic processes.

T. Solomon, R. Selvaraj, and M. Mangwala are with the Department of Computer Science and Information Systems, Botswana International University of Science and Technology, Palapye, Botswana (e-mail: solomont@ biust.ac.bw, selvarajr@ biust.ac.bw, mangwalam@biust.ac.bw).

A. M. Zungeru is with the Department of Electrical, Computer and Telecommunication Engineering, Botswana International University of Science and Technology, Palapye, Botswana (e-mail: zungerum@biust.ac.bw).

A stochastic process is simply a probability process that is, any process in nature whose evolution we can analyze successfully in terms of probability. On the empirical side, a discussion of the nature of probability would take us too far afield (and might sidetrack us into philosophy), and on the mathematical side the definitions would require too much high-powered mathematics [3].

According to [3], random variable associates a unique numerical value with each outcome in the sample space. Usually a random variable it is explained as a real valued function defined over a sample space. Random variable is denoted by a capital letter (e.g. Y or X) and a particular value taken by a random variable is denoted by corresponding lower case letters (y or x).Random variable can either be discrete or continuous.

The remainder of this paper is organized as follows. Section II gives a brief explanation on mathematical modelling. It will also include the modelling process as well as the classification of mathematical models. Section III gives the related work to this paper. In Section IV, we provided data sets that will be used to model BIUST network and a packet distribution model for BIUST network will also be presented. Finally, the paper is concluded in Section V.

II. MATHEMATICAL MODELLING

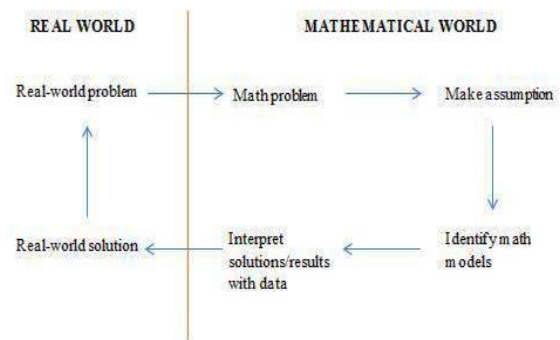### A. Overview of Mathematical Modelling

Fig. 1. Modelling process [4].

Mathematical modelling is by and large comprehended as the way toward applying mathematics to a real world problem with a perspective of comprehension the last mentioned [4]-[8]. One can contend that mathematical modelling is the same as applying mathematics where we additionally begin with a real world problem; we apply the essential mathematics, yet after having found the result we no more consider the underlying issue aside from maybe to check if our answer makes sense. This is not the situation with mathematical modelling where the utilization of mathematics is more to understand this real world problem.

The modeling process might possibly result to solving the problem altogether however it will reveal insight to the circumstance being studied [4]. Fig. 1 shows the key phases in modelling process. The adopted model that is used to model the network is based in BIUST Network.

III. MODELLING OF TRAFFIC NETWORKS

Traffic models reflect our best knowledge of traffic behavior. Latest studies of real telecommunications network traffic data have revealed that tele-traffic exposition self-similar (or fractal)properties over a wide range of time scales [9]-[14].The properties of self-similar telecommunications network traffic are very distinct from properties of traditional models based on Poisson, Markov-modulated Poisson, and related processes. Usage of traditional models in networks characterized by self-similar processes can result in to biased conclusions about the performance of analyzed networks [13], [15]. Traditional models can result in over-estimation of the network performance [16], [17], lack of allocation of communication and data processing resources, and hence problems in ensuring the QoS. Then, full understanding is that the self-similar nature in tele-traffic is a vital issue.

Self-similar tele-traffic is seen in LAN and WAN, where superposition of strictly independent alternating ON/OFF traffic models whose ON- or OFF-periods have heavy-tailed distributions with infinite variance can be used to model aggregate network traffic that shows self-similar (or long-range dependent) behavior typical for measured Ethernet LAN traffic over a wide range of time scales [12], [18].

In ATM network traffic self-similar traffic arriving at an ATM buffer results in a heavy-tailed buffer occupancy distribution, and buffer cell loss probability reduces with the buffer size, not exponentially as in traditional Markovian models, but hyperbolically.

One more implementation of traffic self-similarity is in Internet traffic, where many characteristics of WWW can be modeled using heavy-tailed distributions, including the distribution of traffic times, the distribution of user requests for documents and the distribution of WWW document sizes [15].

In TCP/IP network traffic the transfer of files or messages shows that the reliable transmission and flow control mechanisms of TCP serves to maintain long range dependent structure which include heavy-tailed file size distributions [19]. The relationship between self-similar traffic and network performance is defined, as captured by performance measures such as packet loss rate, retransmission rate and queueing delay, where increased self-similarity results in degradation of performance, and queueing delay exhibits a dramatic increase as self-similarity increases.

The self-similarity observed in video traffic provides possibility for developing models for Variable Bit Rate (VBR) video traffic using heavy-tailed distributions [20]. The autocorrelation of the VBR video sequence decay hyperbolically and can be model educing Fractional Autoregressive Integrated Moving-Average (F-ARIMA) and Fractional Gaussian Noise (FGN) self-similar processes [13], [21].

The impact of self-similar models on queuing performance is important and the main trends in such findings are connected with (a) permission traffic modeling for high speed networks, (b) efficient simulation of actual network traffic and (c) analyzing queuing models and protocols under realistic traffic scenarios [22]. The traditional models of tele-traffic that assume independent arrivals, based on Poisson processes, Markov-modulated Poisson processes and other related processes are not able to capture the self-similar nature of tele-traffic [23].

In another work presented by [13], the time series of self-similar processes exhibit burstiness over a wide range of timescales. Self-similarity can statistically detail wireless IP network traffic that is bursty on numerous time scales. Modeling and simulation of self-similar telecommunications network traffic can be done with the generators of synthetic self-similar sequences, which are separated into two practical classes: the sequential generators and the fixed-length sequence generators. The fixed-length sequence generators for simulation of self-similar wireless IP network traffic are contemplated in this work [13].

IV. DATA SETS

A. BIUST Network Packet Intervals

Live Packet Caption in Morning Time: The data collection was from the live BIUST Network in the morning between 10:00:23 to 10:15:23 on the 06-07-2015. The data is presented in Table I.

Live Packet Caption in Afternoon Time: The data collection was from the live BIUST Network in the afternoon between 14:59:44 to 15:14:44 on the 22-06-2015.

Day Live Packet Caption for a Working hours: The data collection was from the live BIUST Network for the whole day between for a period of 8 hours on the 10-09-2015. The data is presented in Table II.

B. Sizes and Frequency of Occurrence of the Packets

The first two columns of Table I show the values in the log file obtained from Wireshark and the remaining columns are derived from it. The first line shows 0 packets with sizes ranging from 0-19 (bytes). Using the concept of intervals, class limits and the midpoint of a class from statistical theory, the fourth column of Table I, Average Packet Length (APL_i) , is obtained. APL for the interval i is given by

$$APL_i = \frac{x_{mi} + x_{Mi}}{2} \quad 1 \le i \le 10$$

In which, x_{mi} and x_{Mi} are the lower and higher values of the i-th interval, respectively. Column "PL" in Table I. PL shows the packet size (in *bytes*) and FO is the frequency of occurrence of the packets.

APL Standard value APL_s shown in Table I is obtained from equation (20) dividing the value of APL_i by MTU (1500), standard for Ethernet networks. In the last column of the same table, the values FD or (FD_s) are obtained dividing the number of occurrence of the packet size by the total number of packets, given by Equation (21).

$$APL_s = \frac{APL_i}{1500} \quad 1 \le s \le 10,$$

$$FD_s = \frac{FO_i}{\sum_{i=1}^{10} FO_i} \quad 1 \le s \le 10$$

C. Mathematical Model

The mathematical model is based on the analysis of Table I, data input and data output of the system and being modelled through the use of Matlab software. The mathematical model used is Pareto distribution which its probability density function is given by Equation (10). The parameter a was varied such that $1 \le a \le 3$. This is presented in Fig. 2, Fig. 3, and Fig. 4 respectively. It was observed that the PDF curve of BIUST network with a parameter of 1 reaches 0 at an x value of 5, with a parameter of 2 it reaches 0 at an x value of 4, and a parameter of 3 at an x value of 3. The cumulative distribution function curve of BIUST network is then presented in Fig. 5.

TABLE I: LIVE PACKET CAPTION IN MORNING TIME

i	PL	FO	APL_i	APL_s	FD_s
1	0-19	0	9.5	0.00633	0
2	20-39	0	29.5	0.1967	0
3	40-79	27801	59.5	0.03967	0.4510659701
4	80-159	16063	119.5	0.07967	0.2606191388
5	160-319	2034	239.5	0.15967	0.0330012655
6	320-639	2072	49.5	0.033	0.0336178084
7	640-1279	927	959.5	0.63967	0.0150403998
8	1280-2559	11711	1919.5	1.27967	0.1900087614
9	2560-5119	990	3839.5	2.55967	0.0160625629
10	5120-10239	36	7679.5	5.11967	0.0005840932
	Total	**14905**			

TABLE II: LIVE PACKET CAPTION IN WORKING HOURS

| i	PL	FO	APL_i	APL_s	FD_s
1	0-19	0	9.5	0.00633	0
2	20-39	0	29.5	0.1967	0
3	40-79	46783	59.5	0.03967	0.0261327321
4	80-159	1563211	119.5	0.07967	0.8732012555
5	160-319	115409	239.5	0.15967	0.0644668466
6	320-639	46708	49.5	0.033	0.0260908375
7	640-1279	2654	959.5	0.63967	0.0014825101
8	1280-2559	11976	1919.5	1.27967	0.0033897292
9	2560-5119	3264	3839.5	2.55967	0.0018232528
10	5120-10239	202	7679.5	5.11967	0.0001128361
	Total	**1790207**			
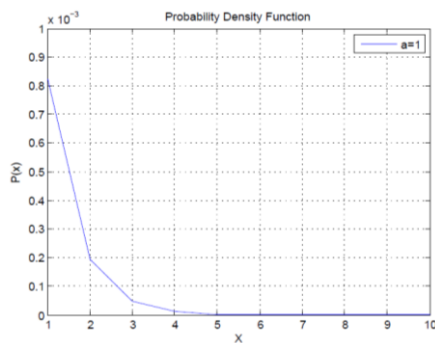
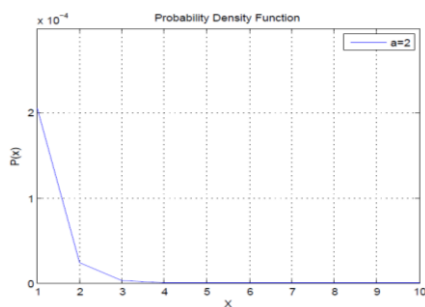Fig. 2. PDF with parameter=1.
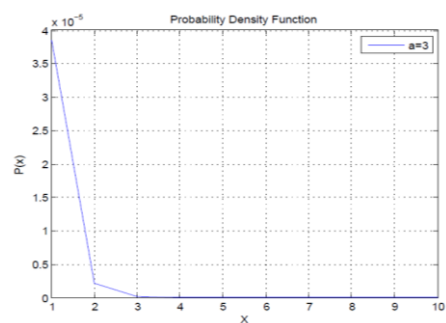
Fig. 4. PDF with parameter=3.

Fig. 3. PDF with parameter=2.

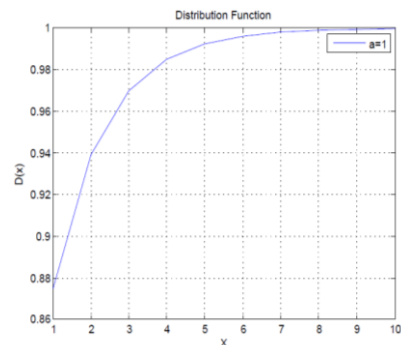Fig. 5. CDF with parameter=1.

As Pareto distribution uses the 80-20 rule corresponds to a particular value of parameter (α), it was observed that about 20% of the users had about 80% of the bandwidth consumed. The PDF graphs shows that the probability or fraction is rather high at a small value of *PL*, and then decreases steadily as value of *PL* increases. Rather the CDF curve is high at higher values of *PL* then decreases steadily as value of *PL* decreases. Fig. 6 shows the relationship between packet length (PL) and frequency of occurrence (FO).

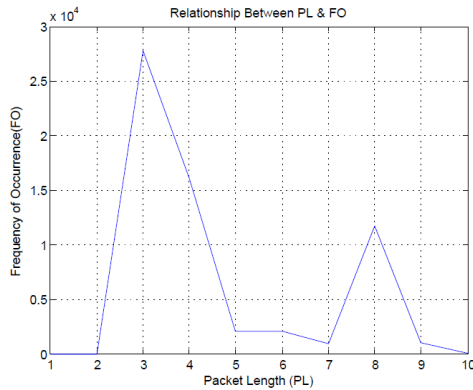Fig. 6. Relationship between PL & FO.

V. CONCLUSION

Models are used in traffic engineering to predict network performance and to evaluate congestion control schemes. Traffic models must have a manageable number of parameters and the estimation of these parameters must be simple. This paper presents a packet distribution model for BIUST network. The model used was Pareto distribution and it was observed that the PDF curve changes when there is a change in parameter. This research was conducted within the boundaries of BIUST network and the attained results might be difficult to generalize to other campus networks due to the difference in network setups and policies implemented. As such, future studies may focus on applying the research on different campus networks so that the results can be generated across different network setups.

ACKNOWLEDGMENT

REFERENCES

[1] T. Solomon, A. M. Zungeru, and R. Sevaraj, "Network traffic monitoring in an industrial environment," in *Proc. 2016 3rd International Conference on Electrical, Electronics, Computer Engineering and their Applications, EECEA 2016*, pp. 133–139, 2016.
[2] M. Wilson. A historical view of network traffic models. [Online]. Available: http://www.cse.wustl.edu/jain/cse56706/ftp/traffic_models2/index.html
[3] A. C. Tamhane and D. D. Dunlop, "Statistics and data analysis "for elementary to intermediate," USA: Prentice-Hall, 2000.
[4] Erlina. (April 2011). What is mathematical modeling? [Online]. Available: http://math4teaching.com/2011/04/16/what-is-mathematical-modeling/
[5] A. Alzaatreh and F. Famoye, "Gamma-pareto distribution and its applications," *Journal of Modern Applied Statistical Methods*, vol. 11, no. 1, pp. 78-94, 2012.
[6] V.S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine*, vol. 32, issue 3, pp. 70-81, 1994.
[7] F. J. Ascombe, "The statistical analysis of insect counts based on the negative binomial distribution," *Biometrics*, vol. 5, pp. 165-173, 1949.
[8] X. Yang and A. P. Petropalu, "The extended alternating fractal renewal process for modeling traffic in high-speed communication networks," *IEEE Trans. Sig. Proc.*, vol. 49, no. 7, July 2001.
[9] O. Osterbo, "Models for calculating end-to-end delay in packet networks," *ITC-18*, Berlin, Germany, August 31-September 5, 2003, pp. 1231-1240.
[10] O. Boxma and I Cohen, *Self-Similar Network Traffic and Performance*, New York: John Wiley & Sons, 2000.
[11] D. Radev and I. Lokshina. (2009). Modeling and simulation of self-similar wireless IP network traffic. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5068931&tag=1
[12] H. Bevrani and S. Sharififar, "An approximation to binomial distribution," *Journal of Statistical Theory and Practice*, vol. 14, no. 2, pp. 1-8, 2014.
[13] B. Chandrasekaran, "Survey of network traffic models," *IEEE Communications Magazine*, March, 1994.
[14] H. D. Jeong, "Modelling of self-similar teletraffic for simulation," PhD Thesis, University of Canterbury, USA, 2002.
[15] R. Faraj, "Modeling and analysis of self-similar traffic in ATM networks," PhD thesis, 2000.
[16] X. Zou. (2016). Source traffic modeling and generation. Computer Science Department, KSU. [Online]. Available: http://www.cs.kent.edu/~xzou/NetProj/TrafficModel.htm
[17] H. J. Kushner, *Heavy Traffic Analysis of Controlled Queueing and Communications Networks*, New York: Springer, 2001.
[18] A. Bobbio, A. Horvath, M. Scarpa, and M. Telek, "Acyclic discrete phase type distributions: properties and a parameter estimation algorithm," *Performance Evaluation,* vol. 54, no. 1, pp. 1-32, 2003.
[19] D. Radev, "Fluid flow analysis for long-range dependent traffic," in *Proc. the Computer Science'2005, Chalkidiki*, Greece, 2005.
[20] D. Ravid and I. Lokshina, "Modeling and simulation of self-similar teletraffic," in *Proc. Industrial Simulation Conference ISC'2007, Delft*, The Netherlands, 2007.
[21] J. Mehdi, *Stochastic Models in Queueing Theory*, Academic Press, 2003.
[22] B. J. Hayes and T. V. J. Ganesh, *Modeling and Analysis of Telecommunications Networks*, New York: John Wiley and Sons, 2004.
[23] G. Giovanni, *Queuing Theory and Telecommunications: Networks and Applications*, US: Springer, 2005.

Thato Solomon received the MSc and BSc from Botswana International University of Science & Technology and Limkokwing University of Creative Technology, Botswana in 2017 and 2012 respectively. He is currently a research/teaching assistant in the Department of Computer Science and Information Systems, Botswana International University of Science and Technology (BIUST) Botswana. This paper "A packet distribution traffic model for industrial application: A case of BIUST network" forms part of his MSc research work. His research interests are bandwidth management, network analysis, planning and mathematical modelling of networks. He has published four papers in journals and international refereed conferences which were related to his MSc research work.

Adamu Murtala Zungeru received the PhD, MSc and BEng from Nottingham University, Ahmadu Bello University Zaria Nigeria and Federal University of Technology Minna Nigeria respectively. He was a research fellow in the Electrical Engineering and Computer Science Department at Massachusetts Institute of Technology (MIT) USA, where he also obtained a Postgraduate Teaching Certificate in 2014. He is currently a senior lecturer at Botswana International University of Science and Technology (BIUST). His research interests are in swarm intelligence, wireless sensor networks, embedded systems, and design of analog and digital electronic circuits. He has authored three academic books and has published over forty papers in journals and international refereed conferences. He has also filed 2 international patents application for his innovative work under the Patent Cooperation Treaty (PCT). He is a member of the IEEE, ACM and AP-S.

Rajalakshmi Selvaraj is currently working as a senior lecturer at Department of CS & IS in BIUST, Botswana. She has received the PhD in computer science and also the M.Phil. master's (MCA) degree and bachelors (BSc) degree in computer science. She has 15 years' experience of teaching undergraduate and postgraduate students of computer science, a wide range of subjects such as computer networks, sensor networks, image processing, data mining and ware housing, software engineering, Java, object oriented programming, unified modelling language, data base management systems, data base design, web technologies etc. She got CCNA certification from cisco as well.

Mmoloki Mangwala is a lecturer of computer engineering in the College of Engineering at Botswana International University of Science and Technology (BIUST). He received his BSc degree in computer engineering from University of KwaZulu Natal, South Africa, the MSc degree in computer science and the PhD in computer science from North West University in South Africa specializing in all-optical next generation networks. Prior to joining BIUST, he was a computer systems engineering lecturer at Botswana Accountancy College.