© Universiti Tun Hussein Onn Malaysia Publisher's Office

JST

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/jst
ISSN : 2229-8460 e-ISSN : 2600-7924

Journal of Science and Technology

An Approximate Performance of Self-Similar Lognormal/M/1/K Internet Traffic Model

Jumoke Popoola[1*], Olusogo Popoola[2], Oyebayo Ridwan Olaniran[3]

[1]Department of Statistics, Faculty of Physical Sciences,
University of Ilorin, Ilorin, P.M.B. 1515, NIGERIA

[2]Department of Computer Engineering, Faculty of Engineering and Technology,
University of Ilorin, Ilorin, P.M.B.1515, NIGERIA

[3]Department of Statistics, Faculty of Physical Sciences,
University of Ilorin, Ilorin, P.M.B. 1515, Ilorin, NIGERIA

DOI: https://doi.org/10.30880/jst.2019.11.02.005
Received 21 October 2019; Accepted 2 December 2019; Available online 29 December 2019

Abstract: Modeling Internet traffic data with the inherent condition of concurrent arrival of packets requires the use of heavy-tailed distributions. In this paper, we present log-normal distribution as a suitable heavy-tailed distribution for modeling self-similar Internet arrival process. Specifically, we developed its approximate performance measures for large buffer size. Results using the simulated data confirm the adequacy of the proposed model.

Keywords: Internet traffic, Self-similarity, G/M/1 Model, Log-normal distribution.

1. Introduction

The understanding of Internet traffic modeling helps in finding probabilistic processes to represent the pattern or behavior of any network traffic. Following [1], self-similarity can be described mathematically as an incremental process $X_j(j = 1,2, \dots)$ whose average in non-overlapped blocks of size m is $X_j^m(j = 1,2, \dots)$ i.e;

$$X_j^m = \frac{1}{m}(X_{jm-m+1}, +X_{jm-m+2} + \dots + X_{jm}). \quad (1)$$

Then X_j is self-similar if X_j^m is approximately $m^{H-1}X_j$;$(X_j^m \sim m^{H-1}X_j)$ which is in fact equivalent to;

$$var(X_j^m) = m^{2H-2}var(X_j). \quad (2)$$

where m $(m \geq 1)$ is the scale parameter and H is the Hurst parameter. For instance, when $H = 1$, process X_j^m and process X_j have the same distribution without any decay. The Hurst parameter, H is used to measure the burstiness of Internet traffic process, where 0.5 < H >1 indicates that a process is self-similar or otherwise.

*Corresponding author: jmkbalpop@unilorin.edu.ng

2. The Internet Traffic Model

R. Zheng et. al. proposed a stochastic resource scheduling framework for real-time in mobile cloud networks to pre-allocate resources using bilayer dynamic Markov decision processes [2]; G.M. Goerg described a bi-jective transformation to simulate heavy-tailed versions of arbitrary random variables, X for modeling the Internet traffic data [3]. The simulation was based on, X, being Gaussian which reduces to Tukey's h distribution but X ,does not adequately provide a self-similar process that can be applied to any other distributions apart from the distribution generated in the transformation; [4] observed that in modeling the Internet traffic, researchers are contemplating of changing Poisson process with another distribution in form of Compound Poisson Process, Markov Modulated Poisson Process or use of Pareto model with its facility to cater for burstiness to track actual traffic characteristics. These approaches are practically inadequate for a self-similar Internet data analyses.

3. Distribution of Packets in the System in a G/M/1 Traffic

The $G/M/1$ queue is a single-server queue, where the arrivals process is general and the service process has an exponential distribution. Request-packets arrive individually and their inter-arrival times are independently and identically distributed. [5] gave the stationary distribution p_i of packets in a $G/M/1$ traffic as;

$$p_i = \rho(1-\xi)\xi^{i-1},\ i > 0. \quad (3)$$

Where ξ is interpreted as the traffic intensity and it is computed from the relation given below;

$$\xi = F_A^*(\mu - \mu\xi). \quad (4)$$

Where $F_A^*(\mu - \mu\xi)$ is the laplace transform of the arrival process evaluated at $s = \mu - \mu\xi$ and $0 < \xi < 1$.

3.1 G/M/1 where G is Lognormal

The probability density function (PDF) of Log-normal distributed variable T with mean θ and variance σ^2 as described by [6] is given as:

$$f(t;\theta,\sigma^2) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\log t-\theta)^2}{2\sigma^2}}; t > 0. \quad (5)$$

Lemma 1: If the random variable t is log-normally distributed with mean θ and variance σ^2, then $\log(t)$ has a normal distribution with mean θ and variance σ^2.

Since there exist no close form for the Laplace transform of the lognormal distribution, in this paper, we approximate it by using Laplace transform of the normal distribution. The most important thing is to transform the original inter-arrival times observed to normal by obtaining its natural logarithm before applying the model proposed in this paper.

Thus, the Laplace transform $L(t)$ of the normal distribution [7] is;

$$L(t) = e^{\frac{s^2\sigma^2}{2} - \theta s}. \quad (6)$$

Now using lemma 1, the Laplace transform of log of lognormal random variable t is (4). Thus, we derive;

$$F_A^*(\mu - \mu\xi) = e^{\frac{(\mu-\mu\xi)^2\sigma^2}{2} - \theta[\mu-\mu\xi]}. \quad (7)$$

$$\xi = e^{\frac{(\mu-\mu\xi)^2\sigma^2}{2} - \theta[\mu-\mu\xi]}$$

In this paper, using numerical approximation beginning the iterative process with $\xi^{(0)} = 0.5$ [5], the successive iteration is,

$$\xi^{(j+1)} = e^{\frac{(\mu-\mu\xi^{(j)})^2\sigma^2}{2} - \theta[\mu-\mu\xi^{(j)}]}. \qquad (8)$$

We can define the optimal traffic intensity as ϑ at final iterations in a $G/M/1$ where G is lognormal. Once ξ is obtained, in this paper, the specific derived approximations are obtain as follow.

The corresponding distribution of packets in the system is;

$$p_i = (1-\xi)\xi^i. \qquad (9)$$

Therefore, the distribution of packets in the system is;

$$p_i = \begin{cases} 1-\vartheta, & i = 0 \\ (1-\vartheta)\vartheta^i, & i > 0 \end{cases}. \qquad (10)$$

The probability of having K packets in the system often interpreted as blocking probability is;

$$p_K = (1-\vartheta)\vartheta^K. \qquad (11)$$

In the same line, the optimal buffer size K given p_K is;

$$K = \frac{log\left[\frac{p_K}{(1-\vartheta)}\right]}{log(\vartheta)} \quad . \qquad (12)$$

The above derivation indicate a close link with the $M/M/1$ traffic with intensity parameter, ρ, the mean number in the system for $M/M/1$ [8] is L; where $L = \frac{\rho}{1-\rho}$.

Thus, in this paper by analogy, the mean number in the system for $G/M/1$ where the inter-arrival time distribution is lognormal is;

$$L = \frac{\vartheta}{1-\vartheta}. \qquad (13)$$

Similarly, the derived mean number on the traffic is L_q; where $L_q = L - \vartheta$

$$L_q = \frac{\vartheta}{1-\vartheta} - \vartheta. \qquad (14)$$

Since there exist close association between $M/M/1$ and $G/M/1$ traffic, the only difference is the utilization factor, which in $G/M/1$ traffic is ϑ. The distribution of time spent in $M/M/1$ is given below as;

$$W(t) = \mu(1-\rho)e^{-[\mu(1-\rho)]t}, t > 0. \qquad (15)$$

While the distribution of time spent on the traffic is;

$$W_q(t) = \rho\mu(1-\rho)e^{-[\mu(1-\rho)]t}, t > 0. \qquad (16)$$

Now, assuming lognormal distribution as the inter-arrival distribution where it has been derived earlier in this paper that $\xi = \vartheta$. The corresponding distributions of time spent in the system and on the traffic as derived in this paper are;

$$W(t) = \mu[1-\vartheta]e^{-[\mu(1-\vartheta)]t}, t > 0. \quad (17)$$

And,

$$W_q(t) = \vartheta\mu[1-\tau]e^{-[\mu(1-\vartheta)]t}, t > 0. \quad (18)$$

It is obvious that the above distributions follow exponential distribution with parameter $\mu[1-\vartheta]$. Thus, the derived mean times in the system and on the traffic in this paper are;

$$W = \frac{1}{\mu[1-\vartheta]}. \quad (19)$$

And,

$$W_q = \frac{\vartheta}{\mu[1-\vartheta]}. \quad (20)$$

4. Simulation Study

Self-similar arrival time A_t and exponential distributed transmission time S_t were first generated using packages fArma and stat respectively in R. Then using successive random addition method, the performance (U, L, Lq, W & Wq) of the self-similar traffic were observed and saved at time t. The Empirical performance measures (U, L, Lq, W & Wq) for the traffic model G/M/1 when G is Lognormal were recorded using package arqas in R. The theoretical performance of the models for the distribution Lognormal were obtained using (8), (13), (14), (19) and (20).

The parameters used for the simulation are;

i. Self-similarity level H: 0.6, 0.7, 0.8 and 0.9 indicating low through high self-similar traffic.

ii. Traffic intensity; $U = 0.5 \,\&\, 0.9$ indicating low and heavy traffic.

4.1 Simulation Results

In this sub-section, we present the results of the empirical and theoretical performances of the proposed traffic models vis-a-vis true self-similar traffic and the standard $M/M/1$ traffic model on the simulated network data as shown in Table 1-4 and Fig. 1 and 2.

Table 1 - Performance measures of various empirical and theoretical models when the true traffic intensity is 0.5 and self-similar indices 0.6 & 0.7.

Hurst Index (H)	Model	Performance Measures				
		U	L	L_q	W	W_q
$H = 0.6$	True Self-similar Traffic	0.4899	0.8564	0.3664	0.0119	0.0051
	M/M/1 Empirical	0.4898	0.9574	0.4676	0.0122	0.0060
	M/M/1 Theoretical	0.4888	0.9562	0.4674	0.0122	0.0060
	Ln/M/1 Empirical	0.4659	0.8203	0.3543	0.0110	0.0047
	Ln/M/1 Theoretical	0.4659	0.8725	0.4066	0.0117	0.0054
$H = 0.7$	True Self-similar Traffic	0.4872	0.7724	0.2853	0.0205	0.0076
	M/M/1 Empirical	0.4882	0.9516	0.4633	0.0251	0.0122
	M/M/1 Theoretical	0.4872	0.9500	0.4628	0.0251	0.0122
	Ln/M/1 Empirical	0.4834	0.6784	0.1949	0.0181	0.0052
	Ln/M/1 Theoretical	0.4834	0.9357	0.4523	0.0250	0.0121

***Note: The proposed model is denoted by Ln/M/1.

Table 2 - Performance measures of various empirical and theoretical models when the true traffic intensity is 0.5 and self-similar indices 0.8 & 0.9.

Hurst Index (H)	Model	Performance Measures				
		U	L	L_q	W	W_q
$H = 0.8$	True Self-similar Traffic	0.4830	0.731	0.2478	0.0427	0.0145
	M/M/1 Empirical	0.4839	0.9350	0.4512	0.0545	0.0262
	M/M/1 Theoretical	0.4829	0.9342	0.4512	0.0545	0.0264
	Ln/M/1 Empirical	0.4816	0.6224	0.1407	0.0364	0.0082
	Ln/M/1 Theoretical	0.4816	0.9294	0.4477	0.0544	0.0262
$H = 0.9$	Self-similar Traffic	0.4904	0.7342	0.2438	0.1142	0.0379
	M/M/1 Empirical	0.4917	0.9647	0.4730	0.1495	0.0732
	M/M/1 Theoretical	0.4904	0.9622	0.4719	0.1496	0.0733
	Ga/M/1 Empirical	0.4906	0.6181	0.1275	0.0961	0.0198
	Ga/M/1 Theoretical	0.4904	0.9622	0.4719	0.1497	0.0734
	Ln/M/1 Empirical	0.4901	0.6169	0.1267	0.0959	0.0197
	Ln/M/1 Theoretical	0.4901	0.9611	0.4710	0.1495	0.0733

***Note: The proposed model is denoted by Ln/M/1.

Table 3 - Performance measures of various empirical and theoretical models when the true traffic intensity is 0.9 and self-similar index 0.6.

Hurst Index (H)	Model	Performance Measures				
		U	L	L_q	W	W_q
$H = 0.6$	True Self-similar Traffic	0.8799	6.1694	5.2895	0.0438	0.0376
	M/M/1 Empirical	0.8800	7.1521	6.2720	0.0504	0.0442
	M/M/1 Theoretical	0.8798	7.3229	6.4430	0.0520	0.0458
	Ln/M/1 Empirical	0.9223	11.7449	10.8226	0.0826	0.0760
	Ln/M/1 Theoretical	0.9247	12.9748	12.0501	0.0921	0.0856

***Note: The proposed model is denoted by Ln/M/1.

Table 4 - Performance measures of various empirical and theoretical models when the true traffic intensity is 0.9 and self-similar index 0.7, 0.8 and 0.9.

Hurst Index (H)	Model	Performance Measures				
		U	L	L_q	W	W_q
$H = 0.7$	True Self-similar Traffic	0.8769	4.7282	3.8512	0.0697	0.0567
	M/M/1 Empirical	0.8774	6.9854	6.1080	0.1021	0.0892
	M/M/1 Theoretical	0.8769	7.1253	6.2484	0.1050	0.0921
	Ln/M/1 Empirical	0.8832	5.0394	4.1562	0.0736	0.0606
	Ln/M/1 Theoretical	0.8813	5.0306	4.1493	0.0741	0.0611
$H = 0.8$	True Self-similar Traffic	0.8870	4.7297	3.8427	0.1479	0.1201
	M/M/1 Empirical	0.8873	7.7009	6.8135	0.2386	0.2109
	M/M/1 Theoretical	0.8870	7.8477	6.9607	0.2454	0.2176
	Ln/M/1 Empirical	0.8895	4.8611	3.9715	0.1507	0.1230
	Ln/M/1 Theoretical	0.8883	4.8209	3.9326	0.1507	0.1230
	True Self-similar	0.8888	4.5723	3.6836	0.3869	0.3117

	Traffic					
$H = 0.9$	M/M/1 Empirical	0.8888	7.7629	6.8740	0.6511	0.5760
	M/M/1 Theoretical	0.8888	7.9890	7.1002	0.6761	0.6009
	Ln/M/1 Empirical	0.8908	4.5910	3.7002	0.3851	0.3099
	Ln/M/1 Theoretical	0.8890	4.5901	3.7010	0.3884	0.3132

***Note: The proposed model is denoted by Ln/M/1.

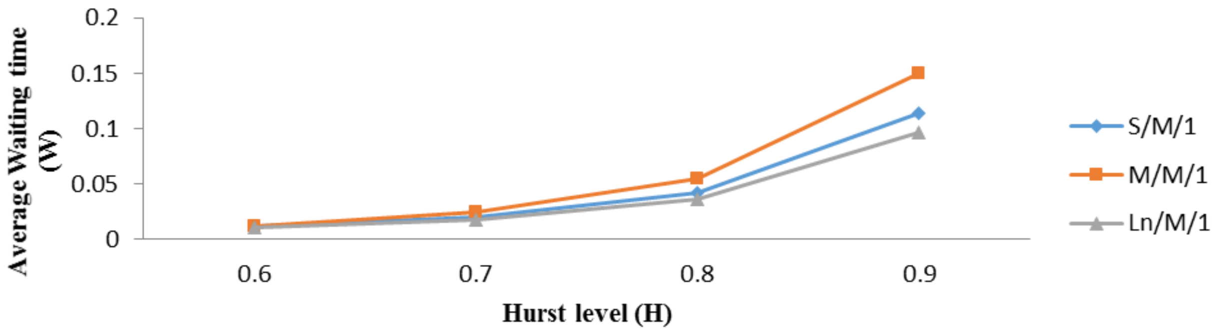

Fig. 1 - Average Waiting Time in the system (W) at various Hurst level for the G/M/1 traffic models when traffic intensity is 0.5.

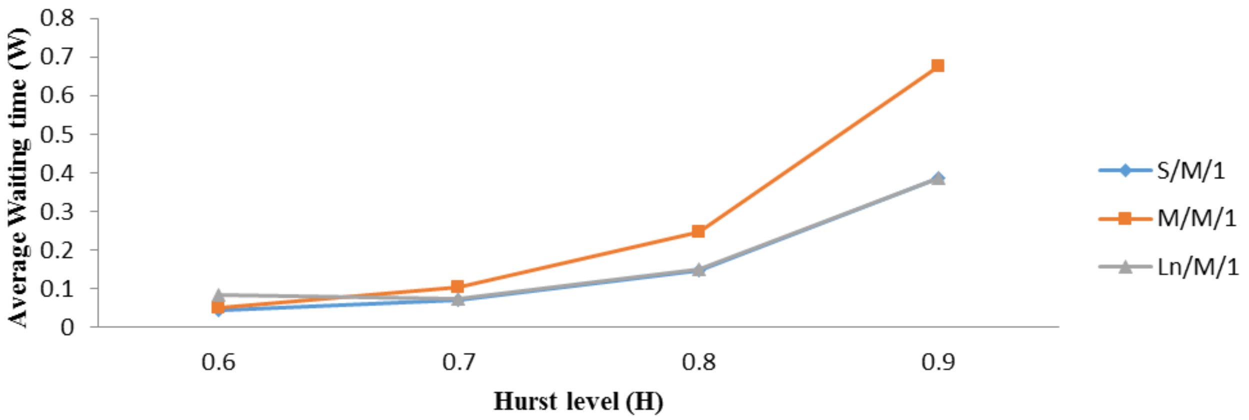

Fig. 2 - Average Waiting Time in the system (W) at various Hurst level for the G/M/1 traffic models when traffic intensity is 0.9.

5. Conclusion and Future work

We have derived the approximate performance measures for a *G/M/1/K* class of Internet traffic model where G is lognormal for large buffer size (waiting room or bandwidth) K. Expression of various performances were given and their relationships to the *M/M/1* traffic model were also provided. Adequacy of the proposed model was observed for modeling high intensity self-similar traffic induced by self-similar arrivals based on the results of simulated data used. The current work can be extended by comparing with other classes of traffic models and heavy tailed distributions other than lognormal to provide a final stand on modeling self-similar Internet traffic.

Acknowledgement

The authors would like to thank to the University of Ilorin Nigeria for the financial and facilities support throughout this research.

References

[1] X. Bai and A. Shami. (2013) Modeling Self-Similar Traffic for Network Simulation," *arXiv preprint,* p. 1308.3842.

[2] R. Zheng, K. Liu, J. Zhu, M. Zhang and Q. Wu. (2019) Stochastic resource scheduling via bilayer dynamic Markov. Computer Communications. 234-242.

[3] G. M. Goerg. (2015) The lambert way to gaussianize heavy-tailed data with the inverse of tukey'sh transformation as a special case.," *The Scientific World Journal.,* p. https://www.hindawi.com/journals/tswj/2015/909231/.

[4] O. H. Alakiri, A. Oladeji, C. B. Benjamin, C. C. Okolie and M. F. Okikiola. (2014) The desirability of pareto distribution for modeling modern internet traffic characteristics.," *International Journal of Novel Research in Engineering and Applied Sciences (IJNREAS)*.

[5] W. J. Stewart. (2019) Probability, Markov chains, queues, and simulation: the mathematical basis of performance modeling. Princeton, New Jersey: Princeton University Press.

[6] M. Fras, J. Mohorko and Ž. Čučej. (2013) Limitations of a Mapping Algorithm with Fragmentation Mimics (MAFM) when modeling statistical data sources based on measured packet network traffic. Computer Networks, 57(17), 3686-3700.

[7] A. G. Rossberg. (2008) Laplace transforms of probability distributions and their inversions are easy on logarithmic scales. Journal of Applied Probability. 531-541.

[8] Z. Rosberg, Y. Peng, J. Fu, J. Guo, E. W. Wong and M. Zukerman. (2014) Insensitive job assignment with throughput and energy criteria for processor-sharing server farms. IEEE/ACM Transactions on Networking (TON), 22, 4, 1257-1270.