An Exact Method for a Class of Self-Similar M/Ga/1/K Queueing System of Internet Stream

*[1]Jumoke Popoola, [2]Olusogo J. Popoola and [1]Oyebayo R. Olaniran
[1]Department of Statistics, University of Ilorin, Nigeria
[2]Department of Computer Engineering, University of Ilorin, Nigeria
{**jmkbalpop|olusogo|olaniran.or}@unilorin.edu.ng**

ORIGINAL RESEARCH ARTICLE
Received: 09-DEC-2021; Reviewed: 27-JAN-2022; Accepted: 10-FEB-2022
http://dx.doi.org/10.46792/fuoyejet.v7i1.747

Abstract- In Internet stream monitoring, computing the expected parameters and distributions for the stream process under self-similarity and long-range dependency requires appropriate and exact methods queueing analyses to obtain performance of the Internet. Our method avoids fitting a non-heavy tailed distribution to Internet stream data by specifically fitting Gamma distribution to its service time process, which enhances the features of correlated events that explain self-similarity and long-range dependency compared to the classical memoryless assumptions of standard Poisson arrival and Exponential service. We propose the specific expectations of the parameters for a class of M/Ga/1/k queueing system ('M' represents a Markovian arrival process that follows Poisson distribution, 'Ga' represents a Gamma distribution service process, '1' represents a single server, 'k' represents the buffer size) in an exact closed-form derivations using the Pollaczek Khintchine formulae alongside the Laplace transform to obtain the performance of queue. The adequacy of the proposed model was confirmed using simulated data of Internet traffic.

Keywords- Gamma distribution, heavy-tail distribution, Internet stream, M/Ga/1/K model, Self-similarity

1 Introduction

Cloud computing is going through evolutions that make Internet architecture undergo notable restructuring. The restructuring will improve users' requests on the network and bring the network closer to its end users (Alberro, et al.,2022). Consequently, our study examines the property of the Transmission Control Protocol (TCP) of the Internet architecture in order to propose a queueing model that supports its non-memoryless property. Internet traffic requires adequate understanding of the structure for measuring traffic that generate data on the Internet and statistical models for the data (Raftery, et al., 2001). Researchers in most recent studies track actual Internet streams characteristics and performance with its self-similar nature using other distributions aside the inadequate Poisson distribution used in the pasts (Alakiri, et al., 2014).

The classical memoryless queueing modelling in a class of M/G/1/K queueing system with Markovian Poisson arrival process, general distribution for the service process and one server is used for exact analyses of Internet traffic data using methods such as Residual Life Approach and Imbedded Markov Chain (Bose, 2013). Consequently, the method of analyses in Bose (2013) is not adequate in a M/G/1 class to capture the often-concurrent Internet stream process due to its non-memoryless concurrent arrivals, which characterizes the TCP of the Internet suite. This occurs particularly in a bulk-service queue (a queue with large requests) and first-come-first-serve (FCFS) queue.

*Corresponding Author

Section B- ELECTRICAL/ COMPUTER ENGINEERING & RELATED SCIENCES
Can be cited as:
Popoola J., Popoola O., and Olaniran O.R. (2022): An Exact Method for a Class of Self-Similar M/Ga/1/K Queueing System of Internet Stream, FUOYE Journal of Engineering and Technology (FUOYEJET), 7(1), 11-16. http://dx.doi.org/10.46792/fuoyejet.v7i1.747

FCFS is a queue that requests are serviced in the order they arrive the server, which may turn out to be a heavy traffic in most of the realistic cases and often lead to self-similar service process. Therefore, our present study derives specific performance indicators in a M/Ga/1/K self-similar Internet traffic model that considers the non-memoryless property of the Internet. The queueing model is a basic model in Internet traffic management to keep track of network performance, equipment designing, quality of service (QoS), security and the management of Internet communication technologies (Cleveland, et al., 2000; Xie, et al., 2018).

2 Review of Distributions for Modelling Internet Traffic

In Internet traffic data analysis, log-normal distribution is a better predictor than Gaussian or Weibull distributions based on likelihood ratio (LLR) test by following the Clauset's methodology, and fitting the log-normal and Gaussian distributions using the correlation coefficient (Alasmar, et al., 2019). Our present work identifies Gamma distribution as another appropriate distribution for fitting data of Internet stream and also gave specific derivations of the expectations of parameters for a sub-class M/Ga/1/k of M/G/1 queueing system in an exact closed-form. This is done using the Pollaczek Khintchine formulae alongside the Laplace transform to derive indicators that measure the performance of M/G/1/K of queue.

Considering the importance of an efficient performing Internet, (Sun, et al., 2018) applied information saved in cloud server based on an automatic and efficient performing Internet stream for building an auxiliary system for patients with memory obstacles. (Popoola, et al., 2019) affirms lognormal distribution as an approximate heavy tail distribution for modelling Internet traffic and derived the specific performance of self-similar lognormal/M/1/K for Internet traffic monitoring. (Popoola, et al., 2020) worked on a real-life application of data on Internet stream, using a general form of Weibull Internet traffic model to monitor its process and performance. The results from the analyses establish Weibull distribution as a possible and suitable distribution for fitting Internet traffic data.

3 Pollaczek Khintchine Formula for Solving *M/G/1* Queue Model

The probability distributions of the performance indicators within the M/Ga/1 queueing system model may be obtained using the Pollaczek Khintchine formulae alongside Laplace transform (Stewart, 2009). These performance indicators at the server end are denoted as average: number of request-packets (L); waiting time of request-packets (W); number of request-packets queueing (L_q); waiting time of request-packets queueing at the server end (Wq). A continuous function, $f(t)$, has Laplace transform expressed as:

$$L(t) = \int_{-\infty}^{\infty} e^{-st} f(t)dt \qquad (1)$$

Similarly, the z transforms of a discrete function , $N(k)$, is;

$$P(z) = \sum_{k=0}^{\infty} z^k N(k) \qquad (2)$$

and the desired distribution may be obtained using inverse z transform.

3.1 Our Proposed Specific Derivations for M/Ga/1/K where G is Gamma

Our present work proposes Gamma distribution among other reviewed distributions for modelling self-similar service time under M/G/1 queue;

The probability density function of Gamma distribution is given as;

$$f(t,\alpha,\mu) = \frac{t^{\alpha-1}\mu^{\alpha}e^{-\mu x}}{\Gamma(\alpha)} \quad t>0 \qquad (3)$$

The Laplace transform of a service time that has Gamma distribution for M/G/1/K as derived in this study is;

$$L(t) = \left[\frac{\mu}{s+\mu}\right]^{\alpha} \qquad (4)$$

where in M/G/1/K queue, according to Pollaczek Khintchine formula, $\rho = \lambda E(S)$ and $(S) = \frac{\alpha}{\mu}$ for a Gamma distribution. Therefore, for specific derivation for M/G/1/K when G is Gamma; $\rho = \frac{\alpha\lambda}{\mu}$.

Using the Pollaczek Khintchine transform, the distribution of M/G/1/K traffic model request packet has the z transform expressed as:

$$P(z) = \frac{(1-\rho)(z-1)B^*[\lambda(1-z)]}{z - B^*[\lambda(1-z)]} \qquad (5)$$

Therefore, specific derivation in this study for $P(z)$, in M/G/1/K when G is Gamma is

$$P(z) = \frac{(1-\rho)(z-1)\left[\frac{\mu}{\lambda(1-z)+\mu}\right]^{\alpha}}{z - \left[\frac{\mu}{\lambda(1-z)+\mu}\right]^{\alpha}} \qquad (6)$$

$$P(z) = \frac{(1-\rho)(z-1)}{z[(\rho(1-z)+1)^{\alpha}]-1} \qquad (7)$$

By using Pollaczek Khintchine transform for M/G/1/K; having no connection of multiple request-packets at the server end has probability denoted by, p_0, and the transform denoted as $P(z)$ when $z = 0$ .Similarly, having i connection of multiple request-packets at the server end probability is expressed as:

$$p_i = \frac{d^i P(z)}{dz^i} \text{ evaluated at } z = 0.$$

Thus, for specific derivations in this study for M/G/1/K when G is Gamma distribution;

$$p_0 = (1-\rho) \qquad (8)$$

$$p_1 = \frac{dP(z)}{dz} \qquad (9)$$

$$= \frac{(1-\rho)\{(z[(\rho(1-z)+1)^{\alpha}]-1)-(z-1)[-\rho z\alpha(\rho(1-z)+1)^{\alpha-1}]+(\rho(1-z)+1)^{\alpha}\}}{[z[(\rho(1-z)+1)^{\alpha}]-1]^2} \qquad (10)$$

at z=0.

By mathematical induction,

$$p_i = \begin{cases} (1-\rho), & i = 0 \\ (1-\rho)[(1+\rho)^{\alpha}-1], & i = 1 \\ (1+\rho)^{\alpha}\left[p_{i-1} - \frac{\alpha\rho^{i-1}[(1-\rho)]}{1+\rho}\right], & i \geq 2 \end{cases} \qquad (11)$$

Using Pollaczek Khintchine transform equation, in this study, the specific derived probability of reaching the maximum blocking bandwidth available to the server for concurrent connections of request-packets, often interpreted as the probability of blocking of the bandwidth, K, available to the server for M/G/1/K when G is Gamma is;

$$p_K = (1+\rho)^{\alpha}\left[p_{K-1} - \frac{\alpha\rho^{K-1}[(1-\rho)]}{1+\rho}\right] \qquad (12)$$

The optimal buffer size, K, given p_K and p_{K-1} derived is;

$$K = 1 + \left[\frac{log\left[\frac{1+\rho}{\alpha(1-\rho)}\left[p_{K-1} - \frac{p_K}{(1+\rho)^{\alpha}}\right]\right]}{\log(\rho)}\right] \qquad (13)$$

The verification of the relationship between Gamma and Exponential distributions for G in M/G/1 by comparing

http://dx.doi.org/10.46792/fuoyejet.v7i1.747 http://journal.engineering.fuoye.edu.ng/

their distributions of request-packets at the server end is thus:

Recall for $M/M/1$, where $\alpha = 1$,

$$p_i = (1-\rho)\rho^i \quad (14)$$

where, $p_0 = (1-\rho)$; $p_1 = (1-\rho)\rho$; $p_2 = (1-\rho)\rho^2$, for the distributions derived in our work, substituting $\alpha = 1$ takes us to M/M/1, which is the parent class model.

Furthermore, using the Pollaczek Khintchine formula, $L = L_q + \rho$, where L is the number of request-packets at the server end, L_q is the number of clients' request-packets queueing at the server end. Thus, request-packet queueing distribution at the server end is given as;

$$p_i(L_q) = p_i(L-\rho) \quad (15)$$

Thus, using Pollaczek Khintchine formula for $p_i(L_q)$, the specific derivation in this study for M/G/1/K when G is Gamma for $p_i(L_q)$ is;

$$p_i(L_q) = \begin{cases} 1-2\rho, & i=0 \\ (1-\rho)[(1+\rho)^\alpha - 1] - \rho, & i=1 \\ (1+\rho)^\alpha \left[p_{i-1} - \frac{\alpha\rho^{i-1}[(1-\rho)]}{1+\rho}\right] - \rho, & i \geq 2 \end{cases} \quad (16)$$

Using the Pollaczek Khintchine mean value formulae for M/G/1/K, the request-packets at the server end and request-packets queueing at the server end, respectively are;

$$L = \rho + \frac{\lambda^2 E(S^2)}{2(1-\rho)} \quad (17)$$

$$L_q = L - \rho = \frac{\lambda^2 E(S^2)}{2(1-\rho)} \quad (18)$$

The distribution of service time process, S, in M/G/1/K for Gamma distribution is;

$$E[S] = \frac{\alpha}{\mu} \quad (19)$$

and,

$$Var(S) = \frac{\alpha}{\mu^2} \quad (20)$$

$$Var(S) = E[S^2] - [E[S]]^2 \quad (21)$$

$$E[S^2] = \frac{\alpha}{\mu^2} + \left[\frac{\alpha}{\mu}\right]^2 \quad (22)$$

thus, for specific derivation in this study for M/G/1/K when G is Gamma;

$$L = \rho + \frac{\lambda^2 \left[\frac{\alpha}{\mu^2} + \left[\frac{\alpha}{\mu}\right]^2\right]}{2(1-\rho)} \quad (23)$$

and,

$$L_q = \frac{\rho\lambda[1+\alpha]}{2\mu(1-\rho)} \quad (24)$$

Using the Pollaczek Khintchine transform equation, time spent by request-packets at the sever end for M/G/1/K distribution has Laplace transform expressed as:

$$W^*(s) = B^*(s)\frac{s(1-\rho)}{s-\lambda+\lambda B^*(s)} \quad (25)$$

Thus, for specific derivation in this study for M/G/1/K when G is Gamma is

$$W^*(s) = \left[\frac{\mu}{s+\mu}\right]^\alpha \frac{s(1-\rho)}{s-\lambda+\lambda\left[\frac{\mu}{s+\mu}\right]^\alpha} \quad (26)$$

$$W^*(s) = \frac{\mu^\alpha s(1-\rho)}{(s+\mu)^\alpha(s-\lambda)+\lambda\mu^\alpha} \quad (27)$$

The corresponding inverse of the derived $W^*(s)$ is:

$$[W^*(s)]^- = \frac{x^{\alpha-1}[\mu(1-\rho)]^\alpha e^{-[\mu(1-\rho)]x}}{\Gamma(\alpha)} \quad (28)$$

Thus, the specific derivation in this study for the time spent by request-packets at the server end for M/G/1/K when G is Gamma has distribution expressed as:

$$W(t) = \frac{t^{\alpha-1}[\mu(1-\rho)]^\alpha e^{-[\mu(1-\rho)]t}}{\Gamma(\alpha)}, t > 0 \quad (29)$$

The derivation in Equation (29) explains that the time spent by request-packets at the server end distribution is also Gamma with parameters, α and μ(1-ϱ). Also, the Laplace transform of time spent by request- packets queueing at the server end, $W_q^*(s)$, can be derived in a similar manner.

Using Pollaczek Khintchine transform, $W_q^*(s)$ for M/G/1/K when G is Gamma is:

$$W_q^*(s) = \frac{s(1-\rho)}{s-\lambda+\lambda B^*(s)} \quad (30)$$

Thus, for specific derivation in this study for the time spent by request-packets queueing at the server end for M/G1/K when G is Gamma has Laplace transform expressed is:

$$[W_q^*(s)]^- = \frac{x^{\alpha-1}\rho[\mu(1-\rho)]^\alpha e^{-[\mu(1-\rho)]x}}{\Gamma(\alpha)} \quad (31)$$

Thus, the corresponding specific derivation in this study of time spent by request-packets queueing at the server

http://dx.doi.org/10.46792/fuoyejet.v7i1.747 http://journal.engineering.fuoye.edu.ng/

end for M/G/1/K when G is Gamma has distribution expressed as:

$$W_q(t) = \frac{t^{\alpha-1}\rho[\mu(1-\rho)]^{\alpha}e^{-[\mu(1-\rho)]t}}{\Gamma(\alpha)}, t > 0 \quad (32)$$

The derivation in equation (32) connotes that the time spent by request-packets queueing at the server end distribution is a product of ρ and a Gamma random variable with parameters α and $\mu(1-\rho)$.

Using the Pollaczek Khintchine formula, the mean time spent by request-packets at the server end for M/G/1/K is expressed as:

$$W = \frac{\alpha}{\mu} + \frac{\lambda E(S^2)}{2(1-\rho)} \quad (33)$$

Thus, for specific derivation in this study for, W, for M/G/1/K when G is Gamma:

$$W = \frac{\alpha}{\mu} + \frac{\lambda\left[\frac{\alpha}{\mu^2} + \left[\frac{\alpha}{\mu}\right]^2\right]}{2(1-\rho)} \quad (34)$$

$$W = \frac{1}{\mu}\left[\alpha + \frac{\rho[1+\alpha]}{2(1-\rho)}\right] \quad (35)$$

Using Pollaczek Khintchine formula, the specific derived formula in this study for the mean time spent by request-packets queueing at the server end for M/G/1/K when G is Gamma is:

$$W_q = \frac{\rho[1+\alpha]}{2\mu(1-\rho)} \quad (36)$$

Table 1. Summary of specific derived performance indicators in this study for M/G/1 where G is Gamma Distribution

Performance indicators	Results
Distribution of request-packets at the server end	$p_i = \begin{cases} (1-\rho), & i=0 \\ (1-\rho)[(1+\rho)^{\alpha}-1], & i=1 \\ (1+\rho)^{\alpha}\left[p_{i-1} - \frac{\alpha\rho^{i-1}[(1-\rho)]}{1+\rho}\right], & i \geq 2 \end{cases}$
Distribution of request- packets queueing at the server end	$p_i(L_q) = \begin{cases} 1-2\rho, & i=0 \\ (1-\rho)[(1+\rho)^{\alpha}-1]-\rho, & i=1 \\ (1+\rho)^{\alpha}\left[p_{i-1} - \frac{\alpha\rho^{i-1}[(1-\rho)]}{1+\rho}\right]-\rho, & i \geq 2 \end{cases}$
Packet loss probability	$p_K = (1+\rho)^{\alpha}\left[p_{K-1} - \frac{\alpha\rho^{K-1}[(1-\rho)]}{1+\rho}\right]$
Optimal Buffer size	$K = 1 + \left[\frac{log\left[\frac{1+\rho}{\alpha(1-\rho)}\right]\left[p_{K-1} - \frac{p_K}{(1+\rho)^{\alpha}}\right]}{\log(\rho)}\right]$
Mean number of request-packets at the server end	$L = \rho + \frac{\rho\lambda[1+\alpha]}{2\mu(1-\rho)}$
Mean number of request-packets queueing at the server end	$L_q = \frac{\rho\lambda[1+\alpha]}{2\mu(1-\rho)}$
Distribution of time spent by request-packets at the server end	$W(t) = \frac{t^{\alpha-1}[\mu(1-\rho)]^{\alpha}e^{-[\mu(1-\rho)]t}}{\Gamma(\alpha)}, t > 0$
Distribution of time spent by request-packets queueing at the server end	$W_q(t) = \frac{t^{\alpha-1}\rho[\mu(1-\rho)]^{\alpha}e^{-[\mu(1-\rho)]t}}{\Gamma(\alpha)}, t > 0$
Mean time spent by request packets at the server end	$W = \frac{1}{\mu}\left[\alpha + \frac{\rho[1+\alpha]}{2(1-\rho)}\right]$
Mean time spent by request-packets queueing at the server end	$W_q = \frac{\rho[1+\alpha]}{2\mu(1-\rho)}$

4 Simulation Study and Results

Self-similar Internet traffic arrival time denoted by A_t and exponential distributed service (transmission) time S_t were initially simulated using packages fArma and stat respectively in R. Next, we used a method of successive random addition to measure performance of S/S/1/K ("S" in the notation represents self-similar arrival process and self-similar service process) traffic model using parameters U, L, Lq, W & Wq that were recorded at time t.

Alongside, empirical indicators of Internet traffic for a M/Ga/1/K queue model performance measurement (i.e., U, L, Lq, W & Wq) were recorded using arqas in R software package. The performance for M/Ga/1/K queue system using the derived theoretical parameters listed in Table 1 of our study were also recorded. The Hurst parameter, H for H values greater than 0.5 which indicates a self-similar process, that is, 0.6, 0.7, 0.8, 0.9 were also used in the simulation. This is to measure the

http://dx.doi.org/10.46792/fuoyejet.v7i1.747 http://journal.engineering.fuoye.edu.ng/

severity of self-similarity at traffic intensities 0.5 and 0.9 for both low and heavy traffics respectively. The results of empirical and theoretical indicators of performance for S/S/1/K, M/M1/K, M/Ga/1/K and; M/Ln/1/K (i.e., M/lognormal/1/K) queue systems were recorded. Specifically, M/Ga/1/K results were compared with those of M/M/1 and S/S/1/K models to examine the behaviour of M/Ga/1/K at congested and non-congested Internet traffic times respectively.

This will help to determine the suitability of M/Ga/1/K as an exact method for modelling the Internet stream.
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 respectively.
ii. Traffic intensity; $U = 0.5$ & 0.9 indicating low and heavy traffic respectively.

Table 2 and Table 3 show empirical and theoretical results of S/S/1/K, M/M/1/K, M/Ga/1/K, and M/Ln/1/K. Specifically, the link between the derived parameters of the M/Ga/1/K class model and M/M/1/K can be verified when we substitute $\alpha = 1$ in the M/Ga/1/K model, which takes us back to M/M/1/K model.

Table 2 shows values of the empirical and theoretical performance indicators of Internet streams of the queueing systems considered at low traffic intensity of 0.5 and at moderately low H value of 0.6. It was observed that the empirical and theoretical performance indicators of each of the considered queueing systems have close values that are also close to the values of the S/S/1/K performance indicator at low self-similarity, H = 0.6. It is observed specifically that the values of the performance indicators for the M/Ga/1/K are close to that of M/M/1/K at moderately low H value, which affirms that the M/Ga/1/K as derived and summarized in Table 1 when there is low traffic intensity and moderately low H value, exhibits low self-similarity. This may be explained by non-concurrent arrivals and service processes leading to its close estimates to a M/M/1/K queue model.

Table 3 shows on the contrary what happens to the queueing systems at high traffic intensity of 0.9 and high H value of 0.9 on the Internet stream. This affirms specifically that the M/Ga/1/K queue class during a heavy Internet traffic with high traffic intensity can be used to estimate a heavy and self-similar traffic since its empirical; and theoretical values as derived and summarized in Table 1 have close values, which are also close to that of the S/S/1/K performance indicator values. This may be explained by concurrent arrivals and service processes leading to its close estimates to a S/S/1/K queue model. Figs. 1 and 2 further show close values of the empirical results with those of the theoretical parameters derived and summarized in Table 1 for a M/Ga/1/K queue through moderately low H value at low traffic intensity; and at high H values in Internet traffic with high intensity respectively. This affirms the adequacy of the M/Ga/1/K queue system as an exact method for modelling a self-similar Internet stream.

Table 2. Performance indicators of various empirical and the derived theoretical models in this study when the true traffic intensity is 0.5 and self-similar index is 0.6 for M/G/1/K queuing models

Hurst Index (H)	Model	Performance Indicators				
		U	L	L_q	W	W_q
H=0.6	Self-similar Traffic(S/S/1/K)	0.4888	0.8725	0.3837	0.0112	0.0049
	M/M/1/K Empirical	0.4898	0.9574	0.4676	0.0122	0.0060
	M/M/1/K Theoretical	0.4888	0.9562	0.4674	0.0122	0.0060
	M/Ga/1/K Empirical	0.4894	0.8641	0.3747	0.0110	0.0048
	M/Ga/1/K Theoretical	0.4888	0.8618	0.3730	0.0110	0.0048
	M/Ln/1/K Empirical	0.5154	1.0983	0.5829	0.0140	0.0074
	M/Ln/1/K Theoretical	0.5137	1.0897	0.5760	0.0139	0.0074

Where: U: utilization factor (traffic intensity); L: Mean number of request-packets in the system;
L_q: Mean number of request-packets queueing; W: Mean time spent in the system; W_q: Mean time spent queueing;

Table 3. Performance indicators of various empirical and the derived theoretical models in this study when the true traffic intensity is 0.9 and self-similar index is 0.9 for M/G/1 queuing models.

Hurst Index (H)	Model	Performance Indicators				
		U	L	L_q	W	W_q
H=0.9	Self-similar Traffic(S/S/1/K)	0.8888	4.5723	3.6836	0.3869	0.3117
	M/M/1/K Empirical	0.8888	7.7629	6.8740	0.6511	0.5760
	M/M/1/K Theoretical	0.8888	7.9890	7.1002	0.6761	0.6009
	M/Ga/1/K Empirical	0.8878	4.5343	3.6465	0.3811	0.3059
	M/Ga/1/K Theoretical	0.8887	4.5709	3.6821	0.3868	0.3116
	M/Ln/1K Empirical	0.8908	4.5910	3.7002	0.3851	0.3009
	M/Ln/1/K Theoretical	0.8890	4.5901	3.7010	0.3884	0.3132

Where: U: utilization factor (traffic intensity); L: Mean number of request-packets in the system;
L_q: Mean number of request-packets queueing; W: Mean time spent in the system; W_q: Mean time spent queueing;

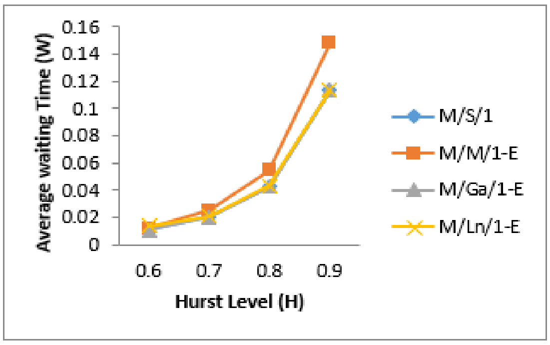

Fig. 1: Average Waiting Time (W) spent in the system at various Hurst levels for the derived M/G/1/K traffic models when traffic intensity is 0.5.

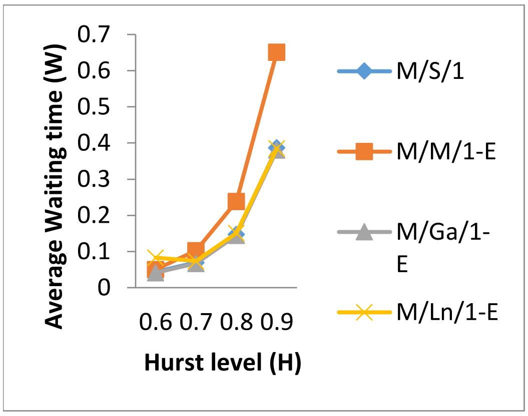

Fig. 2: Average Waiting Time in the system (W) at various Hurst levels for the derived M/G/1traffic models when traffic intensity is 0.9.

5 Conclusion

Our present work has established the appropriateness of M/Ga/1/K traffic model for the performance evaluation of Internet concurrent arrival and service times of request-packets at the server end with specific derivations of the indicators that measure performance within the class. The results of the theoretical derived parameters of M/Ga/1/K queueing system as summarized in Table 1 of our study will help to adequately predict the often-concurrent arrival and service processes of request-packets on the server, specifically for a self-similar M/Ga/1/K Internet traffic.

It is recommended that, the proposed theoretical parameters derived for a M/Ga/1/K queueing system of Internet traffic can be considered for modelling a self-similar Internet traffic with its non-memoryless and concurrent arrival and service processes. This will aid monitoring of Internet streams; help find out performance bottlenecks in existing systems and develop improvements; and management of its limited resources particularly under a self-similar traffic. Our work only investigates the request-packets at the server end for a corporate organization but did not consider the individual properties of devices of users that generated the requests on the server.

References

Alakiri, O. H., Oladeji, A., Benjamin, C. B., Okolie, C. C., & Okikiola, M. F. (2014). The desirability of pareto distribution for modelling modern internet traffic characteristics. *International Journal of Novel Research in Engineering and Applied Sciences (IJNREAS)*, *1*(1), 2-9.

Alasmar, M., Parisis, G., Clegg, R., & Zakhleniu, N. (2019). On the Distribution of Traffic Volumes in the Internet and its Implications. *In IEEE INFOCOM 2019-IEEE Conference on Computer Communications* (pp. 955-963). Paris, France: IEEE. doi:10.1109/INFOCOM.2019.8737483

Alberro, L., Castro, A., & Grampin, E. (2022). Experimentation Environments for Data Center Routing Protocols: A Comprehensive Review. *Future Internet.*, *14*(1), 29.

Bose, S. K. (2013). *An introduction to queueing systems.* New York: Springer Science & Business Media, LLC.

Cleveland, W. S., & Sun, D. X. (2000). Internet traffic data. *Journal of the American Statistical Association.*, *95*(451), 979-985.

Popoola, J., Popoola, O., & Olaniran, O. R. (2019). An Approximate Performance of Self-Similar Lognormal M 1 K Internet Traffic Model. *Journal of Science and Technology.*, *11*(2).

Popoola, J., Yahya, W. B., Popoola, O., & Olaniran, O. R. (2020). Generalized Self-Similar First Order Autoregressive Generator (GSFO-ARG) for Internet Traffic. *Statistics, Optimization & Information Computing.*, *8*(4), 810-821.

Raftery, A. E., Tanner, M. A., & Wells, M. T. (2001). *Statistics in the 21st Century.* CRC Press.

Stewart, W. J. (2009). *Probability, Markov chains, queues, and simulation. The Mathematical Basis of Performance Modeling.* Princeton: Princeton University Press. Doi: https://doi.org/10.1515/9781400832811

Sun, Q., Guo, P., & Zhang, C. (2018). An Auxiliary Notification System for Amnesiac Patients Based on Face Recognition and Voice Synthesis Using Cloud Computing. *In 3rd International Conference on Automation, Control and Robotics Engineering (CACRE 2018) IOP Conference Series: Materials Science and Engineering. 421*, p. 012024. Chengdu, China: IOP Publishing Ltd.

Xie, K., Peng, C., Wang, X., Xie, G., Wen, J., Cao, J., . . . Qin, Z. (2018). Accurate recovery of internet traffic data under variable rate measurements. *EEE/ACM transactions on networking.*, *26*(3), 1137-1150.

