Effects of Cross and Auto-correlation on waiting times in Markovian modulated Poisson Single Server Queueing models

D. Yoan Yomba and Shihao Song

March 3, 2018

I. Abstract

The driving point of this paper is to examine single server queueing models, preferably MAP/G/1 queues in which the arrival mechanism is dependent on some Markov Chain withholding a discrete behavior. This model permits inter-dependencies between the arrival and service processes and thus this work is primarily dedicated to observing the effects of this interdependence on the waiting process of queues in steady state. In order to arrive to a mathematical formulation describing the waiting process in steady state as well as observe the effects of all variable processes on the waiting time process, the Laplace Stieltjes Transform given by the lebesque-Stieltjes integral will be derived.

Key Terms – Stochastic Processes, Laplace Stieltjes Transform, Markovian Arrival Processes, MAP/G/1 queues, Definite Discrete Markov Chains, Steady State LST

II. Introduction

In communication systems, the dependency between two arrival processes defined by their time difference is often of great importance especially when observing the rate of packet streams of traffic over time. Additionally, due to the stochastic behaviors communication systems exhibit specifically in the varying arrival processes across variable network paths, the interdependence within consecutive service times as well as among both inter-arrival and service times is of great importance. The Markovian modulated Poisson arrival processes defined by the MAP/G/1 queue which is a close relative to the M/G/1 queueing model allows us to properly characterize these dependencies and observe their effects on the distribution of waiting times prevalent within the queue.

Such MAP/G/1 queues typically model some stochastic process lying in a state space defined by the real numbers in which the various achievable states within such a system correspond to the magnitude of values present within a queue as well as those currently in service. These queues are also defined by some Markovian Arrival Process in which successive inter-arrival times as well as service times depend on some underlying Markov Chain and withhold in sample and cross sample dependencies. This queuing model is extremely flexible especially in its ability to model both aperiodic and periodic arrival process.

Models with dependencies between inter-arrival and service times have been the focal point of many research endeavors. Some of which include the dependency between service requests and future inter-arrival times proposed by [1]. As well as others discussing the linear dependence between inter-arrival times and prior service times as proposed by [2]. Various other authors particularly [3] explore the bivariate exponential density function that service and inter-arrival times contain in relation to the M/M/1 queue. This work is primarily focused on demonstrating the effects of correlation between service and inter-arrival times on the waiting time distribution of an MAP/G/1 queue.

III. Outline

This paper is organized in the following manner. Section IV observes a queuing model derived by Adan et al. [4]. The stationary distribution of the waiting time process is then derived in section V using the Laplace Stieltjes Transform. Section VI details the experimental methodology used to observe the effects of correlated inter-arrival and service times in relation to the waiting time process. Lastly Section VII details conclusive remarks obtained from the research trial.

IV. Queuing Model

As described by Adan et al. [1], a single server queue is observed where customers are served in the order of arrival. For consistency in notation throughout this paper, our notation is as follows. t_n defines the time of the nth arrival to the present queue. $A_n = t_n - t_{n-1}$ thus defines the time difference of two arrival processes. S_n models the service time of the nth arrival process. The inter-arrival times A_n and service time S_n are regulated by some discrete Markov Chain with states lying in the real valued space $\{1,2,\dots,N\}$ defined by a transition probability matrix P. Based on the regulating property the Markov Chain expresses on S_n and A_n, probabilistically, the following trivariate process is optimal in representing the memory less behavior this system withholds.

$$P\big(A_{n+1} \le x, S_n \le \mathrm{y}, Z_{n+1} \le \mathrm{j} \mid Z_n = \mathrm{i}, \big(A_{r+1,} S_r, Z_r\big), 0 \le \mathrm{r} \le \mathrm{n}-1\big)$$

$$= P(A_1 \le x, S_0 \le \mathrm{y}, Z_1 \le \mathrm{j} \mid Z_0 = \mathrm{i})$$

$$= G_i(y) p_{i,j}(1 - e^{-\lambda_j x})$$

Where $x, y \ge 0; i, j \in \{1, 2, \dots, N\}$. The above expression demonstrates the trivariate process' independence on past states as well as its equivalent representation described by the product of the service time distribution, the probability of lying on some state space within the Markov Chain, and the cumulative density function of the exponentially distributed inter-arrival process A_n.

This system achieves true stability when the mean service time is significantly less than the mean arrival time irrespective of the manner of the distributions.

$$\sum_{i=1}^{N} \pi_i \, \gamma_i < \sum_{i=1}^{N} \pi_i \, \lambda^{-1}{}_i$$

γ represents the mean and σ^2 represents the variance of the service time distribution G_i while π models the stationary distribution of the Markov Chain Z_n. The following expression represents a matrix form representation of the above stability statement.

$$\pi(diag(\,\gamma_1,..,\,\gamma_N) - \,diag(\,\lambda^{-1}{}_1,..,\,\lambda^{-1}{}_N\,))e > 0$$

Where $\Gamma = \begin{bmatrix} \gamma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_N \end{bmatrix}$ $and\ \Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix}$ thus simplifying the above expression to the following,

$$\pi(\Lambda^{-1} - \,\Gamma)e > 0$$

Note, an assumption is made by the MAP/G/1 queueing model that there exists an auto-correlative behavior between S_0 and S_n,

$$\rho(S_0, S_{0+n}) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i(p_{i,j}^{(n)} - \,\pi_j)\,\gamma_j\,\gamma_i}{\sum_{i=1}^{N} \pi_i\, s_i^2 - (\sum_{i=1}^{N} \pi_i\,\gamma_i)^2}, \;\; n \geq 1$$

as well a cross-correlative behavior between A_n and S_n given by

$$\rho(A_n, S_n) = \frac{\sum_{i=1}^{N} \pi_i(\,\lambda^{-1}{}_i \,-\, \lambda^{-1}\,)\,(\,\gamma_i - \gamma)}{\{\sum_{i=1}^{N} \pi_i\,(\,\lambda^{-1}{}_i \,-\, \lambda^{-1}\,)^2 - \,\sum_{i=1}^{N} \pi_i\,(\,\gamma_i \,-\, \gamma)^2\}^{\frac{1}{2}}}$$

where $p_{i,j}^{(n)} = \,P(Z_n = j \mid Z_0 = i\,)$, $\lambda^{-1} = \,\sum_{i=1}^{N} \pi_i\,\lambda^{-1}{}_i$, and $\gamma = \,\sum_{i=1}^{N} \pi_i\,\gamma_i$.

This queuing model defined by Adan et al. [4] is extremely flexible in that transitions without arrivals can be modeled by allowing service times to be of magnitude zero, periodic arrivals can be described by setting the transition probability from state i to state $i+1$ to the transition probability from some state N to state 1 as demonstrated below.

$$p_{i,i+1} = \,p_{N,1}$$

V. Limiting Waiting Time Process Derivation

The waiting time process can be derived in the following manner by use of the LST as proposed by Adan et al. [4]. W_n is representative of the waiting time of the nth item in some queue. 1_A is some indicative random variable of some event A. The transform vector can be obtained from the first moment of some waiting time process relative to both the waiting time and the indicative random variable assuming the waiting time can be described by some exponential distribution and the indicative function maps to the current state of analysis in the Markov Chain.

$$\boldsymbol{\phi_i^N = E(e^{-sW_n} * Z_n = i)}$$

s is some complex number whose real valued component is a number greater than zero. Assuming a convergence is present as N approaches infinity, all computed expectation together thus make up the constituent transform probability vector.

$$\phi(s) = [\phi_1(s), \ldots, \phi_N(s)]$$

The unilateral Laplace transform of the real valued function G_i of s is represented as

$$G_i(s) = \int_0^{inf} e^{-(st)}\, dG_t$$

Which is given by the lebesque-stieltjes integral. The product of the unilateral Laplace transform of the complex variable s, transition probability of real valued states i and j as well as the service time at the arriving state is defined by

$$H(s) = G_i(s)p_{i,j}\lambda_j$$

The transform vector is applied in the following manner incorporating the waiting time to acquire a matrix representation of the model in steady state.

$$\phi(s)[H(s) + sI - \,\Lambda] = sV,$$
$$\phi(0)e = 1$$

V is a vector comprised of the individual values representing the sum of the products of the corresponding transform vector of the arrival process, arrival time distribution, and transition probability of some nth waiting time.

$$v_j = \sum_{i=1}^{N} \phi_i(\lambda_j) G_i(\lambda_j)\, p_{i,j}$$
$$v = [v_1, \ldots, v_N]$$

V in this case serves as an unknown variable and its value can be obtained by computing the determinant of the representation of the queue's steady state in matrix form.

$$v = \det([H(s) + sI - \,\Lambda]) = 0$$

The above equation yields N distinct solutions thus a unique solution for v is acquired by introducing a non-zero column vector a_i and incorporating it in the above equation therefore generating the following representation,

$$[H(s_i) + s_i I - \,\Lambda]a_i = 0$$

Which in turn allows us to acquire a unique solution for v by solving the following linear equations.

$$v = a_i \text{ , where } 2 \leq i \leq N$$
$$v\Lambda^{-1}e = \,\pi(\Lambda^{-1} - \,\Gamma)e$$

This unique solution validates the intrinsic theory of the condition of stability within Markov chains which is specified by some unique stationary distribution and thus some unique transform probability vector.

VI. Experimental Procedure

In order to explicitly observe the effects of auto-correlation and cross-correlation of the inter-arrival and service times on the waiting time within the queue, various numerical examples were defined. The first numerical example aimed to both observe movement of lag in the auto-correlation present within service and inter-arrival times as time steps approached some infinite value as well as the effect of a positive cross-correlation on waiting time. A Markov chain consisting of states lying in the real valued state space $\{1, \ldots, 4\}$ was used as our discrete regulating Markov chain.

The queuing system held arrival rates of $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] = [0.1, 10, 0.1, 10]$ and individual mean service times of $[\gamma_1, \gamma_2, \gamma_3, \gamma_4] = U * [2,1,2,1]$ where U served as a parameter used to explore the effects of augmenting the mean service time on the expected waiting times.

The following transition probability matrix was derived characterizing state transitions in a constituent Markov chain.

$$\begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{vmatrix} = p1_{i,j}$$

The cross-correlation of the inter-arrival and service times is computed by first acquiring the stationary distribution, computing the mean arrival time, and finally computing mean service time.

$$\pi = \pi * P \rightarrow [\,\pi_1,\ \pi_2,\ \pi_3,\ \pi_4] * \begin{vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{vmatrix} \rightarrow [\,\pi_3,\ \pi_0,\ \pi_1,\ \pi_2]$$

$$\begin{matrix} \pi_0 = \pi_3 \\ \pi_1 = \pi_0 \\ \pi_2 = \pi_1 \\ \pi_3 = \pi_0 \end{matrix} \rightarrow \sum_{i=0}^{N} \pi_i = 1 \rightarrow \pi_3 + \pi_2 + \pi_1 + \pi_0 = 1 \rightarrow \frac{1}{n} = \frac{1}{4} = \pi_i$$

$$\lambda^{-1} = \sum_{i=1}^{N} \pi_i\, {\lambda^{-1}}_i \rightarrow \frac{1}{4}\left(\frac{1}{0.1} + \frac{1}{10} + \frac{1}{0.1} + \frac{1}{10}\right) = 5.05$$

$$\gamma = \sum_{i=1}^{N} \pi_i\, \gamma_i \rightarrow \frac{1}{4}(2 + 1 + 2 + 1) * U \rightarrow U = 1 \rightarrow 1.5$$

The cross-correlation was then computed to be 1 through the use of the following equation.

$$\rho(A_n, S_n) = \frac{\sum_{i=1}^{N} \pi_i({\lambda^{-1}}_i - \lambda^{-1})(\gamma_i - \gamma)}{\{\sum_{i=1}^{N} \pi_i({\lambda^{-1}}_i - \lambda^{-1})^2 - \sum_{i=1}^{N} \pi_i(\gamma_i - \gamma)^2\}^{\frac{1}{2}}} = \frac{2.475}{6.125625^{\frac{1}{2}}} = 1$$

The traffic intensity was modelled as

$$p_{int.} = \frac{\sum_{i=1}^{N} \pi_i\, \gamma_i}{\sum_{i=1}^{N} \pi_i\, {\lambda^{-1}}_i} = 1.5 * \frac{U}{5.05} = 0.297U$$

Due to the system achieving stability when the traffic intensity is less than 1, U must lie in $\{0, 3.37\}$. The mean waiting time was then computed by utilizing the following system of equations.

$$W_N = \frac{(1 - p_{int.}) * s * g(s)}{s - \lambda(1 - g(s))}$$

Where s is a complex variable whose real valued component is of magnitude greater than 0 and g(s) is the LST of the service time probability density function.

$$g(s) = \int_0^\infty e^{-st}\, \lambda e^{-\lambda t} dt = \frac{\lambda}{\lambda + s}$$

An equivalent expression is achieved in the following manner,

$$E[W_q] = \frac{\lambda E[\gamma^2]}{2(1 - p_{int.})}$$

Where $E[\gamma^2]$ characterizes the first moment of the square service time, and λ defines the mean arrival time.

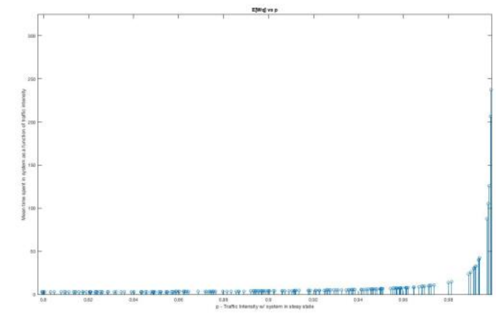

Figure 1: +1 Cross Correlation - Mean waiting time as a function of p

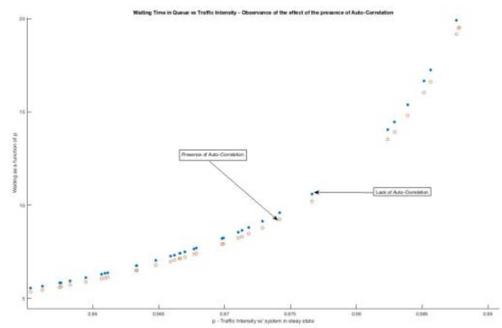

Figure 2: +1 Cross Correlation - Mean waiting time as a function of p, observance of the effects of auto-correlation

Figure 1 demonstrates the mean waiting time within a queue as it maintains steady state behavior and eventually diverges. Figure 2 models the effects of the presence of auto-correlation and allows us to conclude that when cross correlation is positive, the presence of auto-correlation allows for slighter smaller mean waiting times.

The effect of zero cross correlation was also examined. The inter-arrival and service times in this case held no inter-dependence, and were modelled as the arrival rates expressed as $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] = [0.1, 10, 0.1, 10]$ and the service times as $[\gamma_1, \gamma_2, \gamma_3, \gamma_4] = U * [2,1,2,1]$. To observe the presence of auto-correlation and identical independently distributed service times, the following transition probability matrix was generated from an underlying definite finite state Markov chain.

$$\begin{vmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{vmatrix} = p2_{i,j}$$

A second transition probability matrix was derived to provide comparative analysis of the effects of independent service times and arrival rates as well as no cross-correlation.

$$\begin{vmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{vmatrix} = p3_{i,j}$$

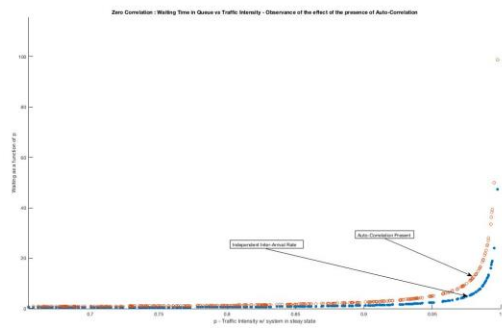

Figure 3: Zero Cross Correlation - Mean waiting times as a function of p, autocorrelation and independent inter-arrival times

From figure 3, it is easy to observe that the mean waiting time of the independent service time and arrival rates as well the lack of auto-correlation provides for substantially lower waiting times. The experiment also shows that as cross-correlation converges to 0, the effects of auto-correlation grows exponentially thus characterizing an inverse behavior between the two.

Negative Cross-correlation was another important phenomenon we aimed to analyze. This analysis was performed by arrival rates expressed as $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] = [0.1, 10, 0.1, 10]$ and the service times as $[\gamma_1, \gamma_2, \gamma_3, \gamma_4] = U * [1,2,1,2]$. Transition probability vectors $p3_{i,j}$ and $p1_{i,j}$ were used for comparative analysis and is presented below in figure 4

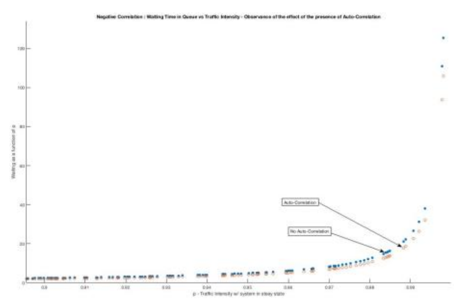

Figure 4: -1 Cross correlation - Mean waiting times as a function of p: comparative analysis of the presence of auto-correlation

From the results acquired in figure 4, when service times and arrival rates are negatively correlated, the lack of in sample auto-correlation results in lower mean waiting time compared to the case in which auto-correlation is present.

Lastly, through extensive experimentation, we were able to observe the converging behavior of auto-correlation as time approaches some infinite value. Present below in figure 5 and 6 are adequate representations of such behavior.

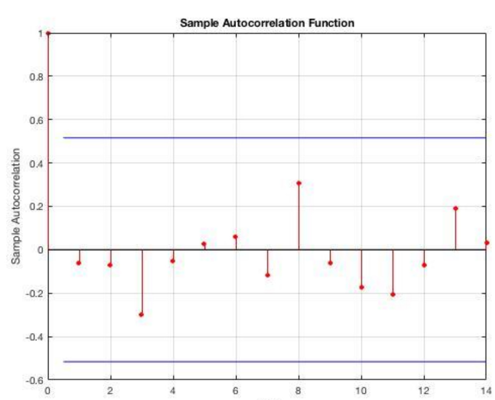

Figure 5: Sample Auto-Correlation of the exponentially distributed service time

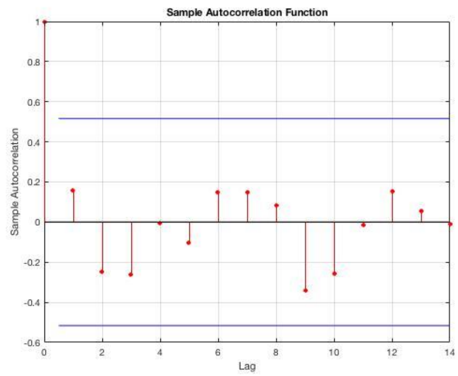

Figure 6: Sample Auto-Correlation of the Poisson distributed arrival process

From the figure 5 and 6, we conclude that as lag approaches an infinite time step, in sample auto-correlation converges to 0.

VII. Conclusion

We were able to conclude that the presence of auto-correlation induces a decrease in waiting time especially when cross-correlation is non-zero as seen in figure 2 and 4. However, the presence of auto-correlation in the absence of cross-correlation increases waiting time. From this we can concisely conclude that cross-correlation definitely has an impact on the waiting time and auto-correlation is helpful for systems only when there is a cross-correlation presence.

VIII. Appendix

Below is a list of performance parameter used to analyze and visualize the effects of variable traffic intensity on the queueing model in steady state.

$$E[W_q] = \frac{\lambda E[\,\gamma^2]}{2(1 - p_{int.})}$$

$$\boldsymbol{Mean\ time\ spent\ in\ the\ buffer}$$

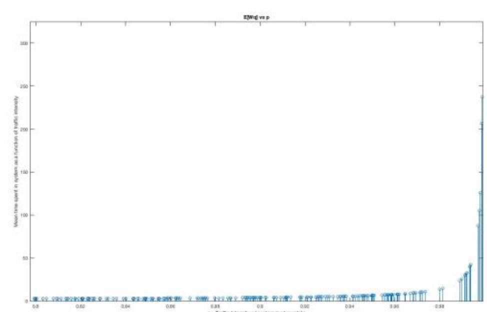

Figure 7: Mean Time Spent in the Queue

$$E[N_q] = \lambda E[W_q] = \frac{\lambda^2 E[\,\gamma^2]}{2(1 - p_{int.})}$$

$$\boldsymbol{Mean\ number\ of\ customers\ in\ the\ buffer}$$

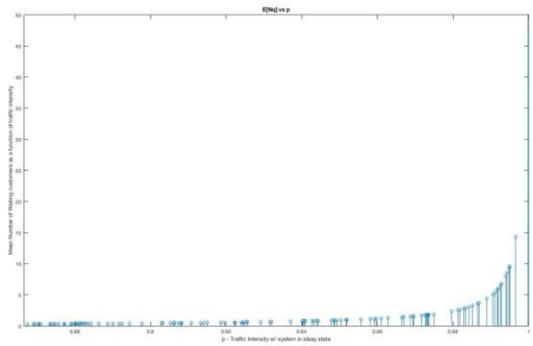

Figure 8: Mean Number of Customers In the buffer

$$E[W] = E[W_q] + E[\lambda] = \frac{\lambda E[\gamma^2]}{2(1 - p_{int.})} + E[\lambda]$$

$$\boldsymbol{Mean\ time\ spent\ in\ the\ system}$$

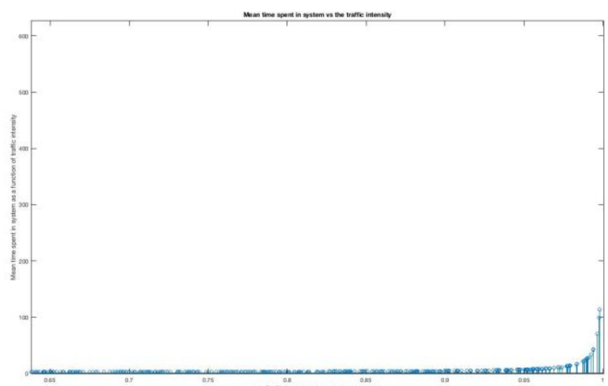

Figure 9: Mean Time Spent In the System

$$E[N] = \lambda E[W] = \frac{\lambda^2 E[\gamma^2]}{2(1 - p_{int.})} + p_{int.}$$

$$\boldsymbol{Mean\ number\ of\ customers\ in\ the\ system}$$

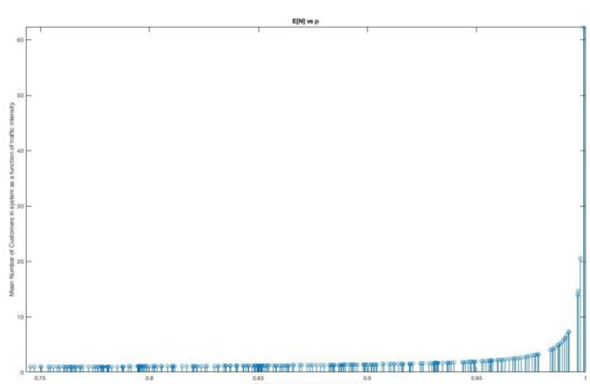

Figure 10: Mean Number of Customers In the system

References

[1] I. Cidon, R. Guerin, A. Khamisy and M. Sidi, On queues with inter-arrival times proportional to service times, Technical Report EE PUB. No. 811, Technion (1991).

[2] I. Cidon, R. Guerin, A. Khamisy and M. Sidi, Analysis of a correlated queue in communication systems, Technical Report EE PUB. No. 812, Technion (1991).

[3] B. Conolly and N. Hadidi, A correlated queue, Appl. Probab. 6 (1969) 122–136.

[4] *Adan, I.J.B.F., and V.G. Kulkarni. "Single-Server Queue with Markov-Dependent Inter-Arrival and Service Times." SpringerLink, Kluwer Academic Publishers.*