Fleet Sizing for Transporters with Seasonal Demands

Debjit Roy
Joseph Sebastian

W.P. No. 2012-12-06
January 2013

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD – 380015
INDIA

Fleet Sizing for Transporters with Seasonal Demands

Debjit Roy
Joseph Sebastian

Abstract:

We develop a quick and approximate way to address the fleet sizing problem for transporters with non-stationary demands and uncertain trip times. Non-stationary demands are modeled using Markov-modulated Poisson process (MMPP) and the variable fleet size is modeled using multi-server queuing construct ($MMPP(2)/M/C$ queue). The customer service level is expressed by the probability that a customer on arrival waits for availability of trucks. The waiting time probabilities are estimated from the steady-state probabilities, which are obtained from the solution of the continuous-time Markov chain. We demonstrate the fleet-sizing approach for a transporter with two level of demand rates and exponentially distributed truck unavailable times.

Keywords: Fleet sizing, non-stationary demand, Markov Modulated Poisson Process

1. Introduction

The road transport sector in India is dominant mode for freight transport, carrying over 60% of domestic freight ([8]). Low startup capital, ease in truck purchase finance, and less effort to obtain driver license and permits allow for an easy entry to this sector. Due to relatively easy entry, this sector is highly fragmented and disorganized with a large number of players. The business practices such as driver management, route management, and costing models are also not standardized. Transporters carry a fleet of about 1-2 trucks to 5000 trucks. However, about 75% of the transporters have a fleet size of five. Transporter margins are about 2%-3% and they are under immense financial pressure with increasing fuel costs, toll rates, tyre costs, interest rates. Recently, the increase in interest rates has also triggered rise in vehicle prices; however, the transportation rates are stagnant, resulting in lower margins.

The transporters often need to decide the appropriate fleet size when the load pattern is seasonal and the truck unavailable time is uncertain. Figure 1 includes the order profile for a liquid bulk transporter company that transports consignment from a port to the manufacturer's facility. The loading quantity in a month varies between 5 metric tonnes to 100 metric tonnes, which can be translated to number to truck requirements based on loading capacity per truck. The one-way travel distance is about 3100 kilometers and the corresponding trip time including delays varies between 7-10 days. The uncertainty in the truck unavailable times are developed due to inadequate transport infrastructure and additional delays due to the regulatory procedures, which take place at

[1]Debjit Roy, Indian Institute of Management Ahmedabad, Gujarat-380015, India, Email: debjit@iimahd.ernet.in
[2]Joseph Sebastian, Indian Institute of Management Ahmedabad, Gujarat-380015, India, Email: jsf.seb@gmail.com

the inter-state borders/check-posts. Under such economic conditions, determining an optimal fleet size is not an easy task, because the demand is not constant in time, and the vehicle unavailable times that includes the loading/ unloading, and travel times are also uncertain, among the other factors. In this research, we attempt to estimate the number of trucks required to provide a given level of service subject to a given load pattern. In particular, we determine the optimal number of trucks (for a particular tonnage capacity and a specific load movement) that the owner should maintain at its depot to minimize the probability of truck unavailability.
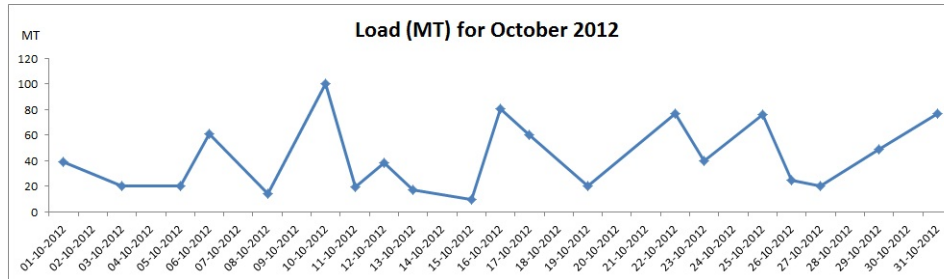

Figure 1: A liquid bulk transporter's order profile for a month

We model this decision problem using a multi-server queue where the servers denotes the vehicles. Using Markov Modulated Poisson Process (MMPP) and Markovian arrival based on Continuous Time Markov Chain (CTMC), we allow for different levels of demand. Further, we assume exponential unavailable times for the vehicles. The main idea of this study is to find the steady state probabilities for the number of customers in the queue. These steady state probabilities help to determine if it is necessary to invest in additional trucks for decrease the average customer waiting time in the system. We could also find out what should be the optimum number of servers that not only satisfy transporter's budget but also minimize customer's waiting time probabilities for truck availability. The rest of this paper is organized as follows. We review the literature on fleet sizing as well as applications in Markov Modulated Poisson Process in Section 2. Section 3 describes the fleet sizing model which includes modeling the arrival process, service times, and the solution approach. Section 4 presents the insights from the numerical experiments whereas in Section 5, we provide concluding remarks and possible course of future research.

2. Literature Review

In this section, we first review literature related to the fleet-sizing problem and then introduce models related to the Markov Modulated Poisson Process (MMPP).

- *Fleet sizing*: Gould [3] showed that a simple fleet size problem with fluctuating demands can be formulated as a linear program. However, he considers certain demand. Our work aligns to the study by Parikh [10] where a multi-server queuing model is developed for the fleet-sizing problem. However, we consider non-stationary customer demands. Vis et al. [12] develop a integer linear programming model to solve the problem of determining vehicle requirements under time-window constraints. The vehicles that operate in a container terminal must start to move the containers from the quay-side to the stack-side within a time-window[release time, due time]. Ronen [11] constructed a cost-based tradeoff model (between speed reduction and number of additional vessels) to determine the number of vessels required to maintain a liner

service. They suggest that "changing prices of bunker fuel open the door for substantial cost savings by adjusting the sailing speed of ships. A large ship may be burning up to 100,000 USD of bunker fuel per day, which may constitute more than 75% of its operating costs. Reducing the cruising speed by 20% reduces daily bunker consumption by 50%. However, in order to maintain liner service frequency and capacity, reducing the cruising speed may require additional ships to operate a route."

● *MMPP*: In the case of Internet traffic, similar looking traffic bursts can be seen at every time scale ranging from few milliseconds to hours. Thus processes having this property tend to exhibit long range dependence, i.e. the values at instances tend to positively correlate with those at future instants. Markov Modulated Poisson Processes are adopted for long range internet traffic measurement studies (see [5], and [2]). MMPP is also adopted to analyze inventory control policies such as Abhyankar and Graves [1] examine an instance of production planning problem with nonstationary demand. In particular, they assume that that aggregate demand follows a cyclic pattern where a high-demand period follows the low-demand period in a cycle.They model cyclical demand as a two-state Markov-modulated Poisson process. The Markov-modulated Poisson process is analyzed via the matrix-analytic method proposed by Neuts[9]. The Markovian arrival process (MAP) was originally explored by Lucantoni et al. [7] (also see [6]). MAP is actually a Markov process where the arrivals are governed by an underlying m-state Markov chain. MAP includes phase-type renewal process and the Markov-modulated Poisson process.

3. Modeling the Fleet Sizing Problem

We model the fleet sizing problem for a transporter using a multi-server queue where the order arrival process is modeled using a Markov-Modulated Poisson Process (MMPP) and the truck unavailable times are modeled as *iid* exponential random variable. Our analysis is to find the optimal number of trucks/servers required so that the waiting probability is below a threshold level. We first discuss the components of the queuing model: modeling of an order arrival process and the modeling of service times. We then present the queuing model for fleet sizing decision.

3.1 Modeling the Order Arrival Process: Markov Modulated Poisson Process

The Markov-modulated Poisson process (MMPP) is a doubly stochastic Poisson process whose rate varies according to a continuous-time Markov process. As with Markovian Arrival process (MAP) and Batch Markovian Arrival process (BMAP), the use of MMPP permits modelling of time-varying systems while keeping the analytical solution tractable. All transitions that are associated do not change the phase of a state in a Markov chain that defines an MMPP. The state-transition-rate diagram for MMPP(4) is shown in Figure 1 where the state of the CTMC is defined by the two-tuple state vector (x, y), [4]. x denotes the number of customers in the system and y denotes the phase of arrival. Note that the transitions in the state-transition-rate diagram denoted the rate at which the process moves from state to state and not the probability of moving from one state to another. MMPP(4) stands for a Markov Modulated Poisson Process with four states in each level. We denote the transition-rate between states of the same levels using α_{ij}, while λ_i is used to denote the arrival rates for the corresponding states.

The generator matrix Q_{MMPP} for the MMPP(4) process is expressed as follows.
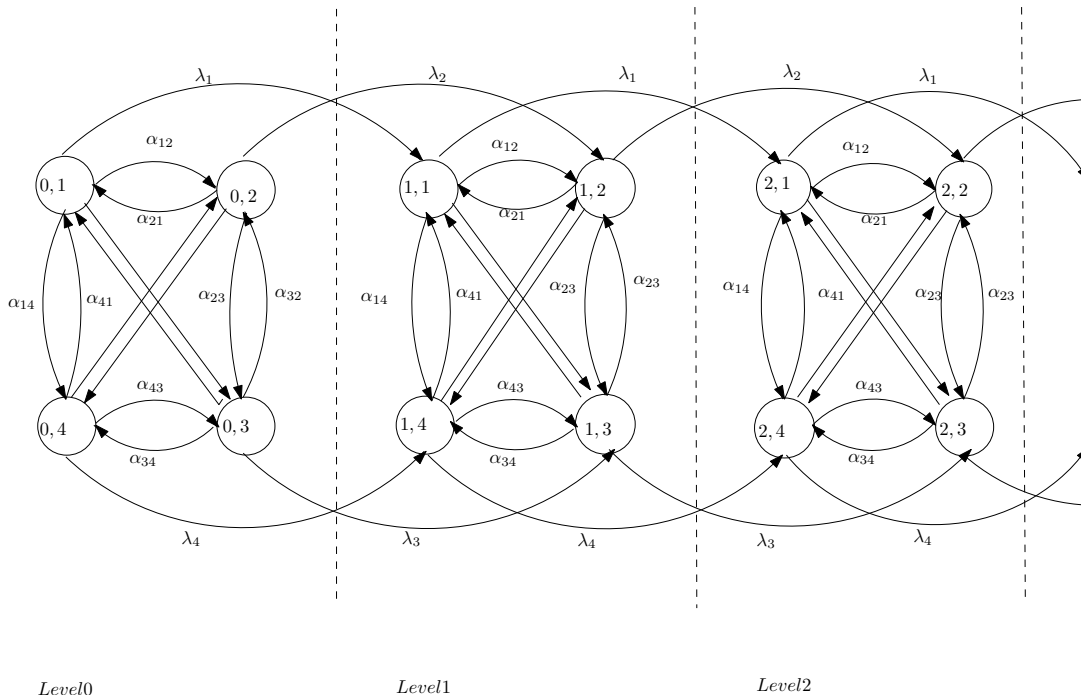
Figure 2: State-transition-rate diagram for MMPP(4) (Adapted from [4])

$$
D_0 = \left\{
\begin{array}{ccccc}
-(\lambda_1 + \alpha_1) & \alpha_{12} & \alpha_{13} & ... & \alpha_{1m} \\
\alpha_{21} & -(\lambda_2 + \alpha_2) & \alpha_{23} & ... & \alpha_{2m} \\
\alpha_{31} & \alpha_{32} & -(\lambda_3 + \alpha_3) & ... & \alpha_{3m} \\
... & ... & ... & ... & \\
\alpha_{m1} & \alpha_{m2} & \alpha_{m3} & ... & -(\lambda_m + \alpha_m)
\end{array}
\right\}
$$

$$
D_1 = \left\{
\begin{array}{ccccc}
\lambda_1 & 0 & 0 & ... & 0 \\
0 & \lambda_2 & 0 & ... & 0 \\
0 & 0 & \lambda_3 & ... & 0 \\
... & ... & ... & ... & ... \\
0 & 0 & 0 & ... & \lambda_m
\end{array}
\right\} = \Lambda
$$

$$
Q_{MMPP} = \left\{
\begin{array}{cccccc}
D_0 & D_1 & 0 & 0 & 0 & ... \\
0 & D_0 & D_1 & 0 & 0 & ... \\
0 & 0 & D_0 & D_1 & 0 & ... \\
... & ... & ... & ... & ... & ...
\end{array}
\right\}
$$

where Λ is the m×m diagonal matrix whose elements are the arrival rates λ_i, i=1,2,...,m; i.e. Λ=diag($\lambda_1, \lambda_2, ... \lambda_m$) and α_{ij} are the transition rates from state i to state j of the same level.

3.2 Modeling the Service Time

The components of a truck unavailable times for a liquid bulk transporter (as referred in Figure 1) are described in Figure 3. The components include wait for truck, travel to port, wait at port and load, travel to plant, wait at plant and unload, and wait for departure. In this research, we

aggregate all the time components and use an aggregate truck unavailable time as the service time in the queue. The service time is assumed to be *iid* exponential random variable.
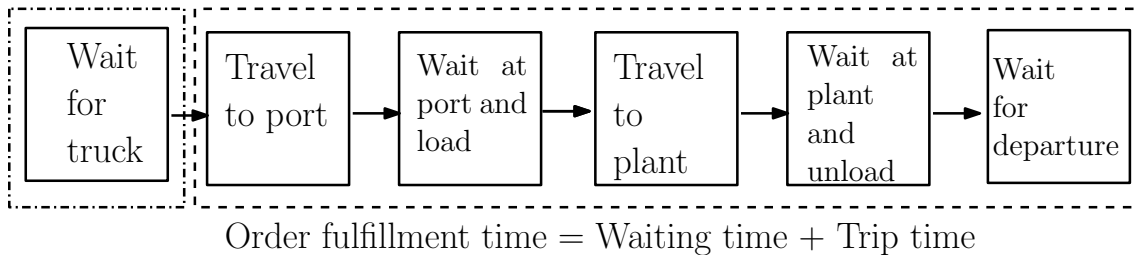

Order fulfillment time = Waiting time + Trip time

Figure 3: Components of customer order fulfillment time

3.3 The $MMPP(2)/M/C$ Queue

MMPP(2) is a two state Poisson process with parameter λ_1, for a time that is exponentially distributed with mean $1/\alpha$ and then switches over to the next state with parameter λ_2 that is exponentially distributed with mean $1/\beta$. The process switches between the two states.

Consider a situation where the customer demand is unpredictable. It can switch between being high or low. The duration of the high and low demand conditions are exponentially distributed with parameters λ_1 and λ_2 respectively. The customer arrival rate has no effect on the customer service rate and the service times are exponentially distributes with a mean μ^{-1}. We can show this example by the following state-transition-rate diagram.
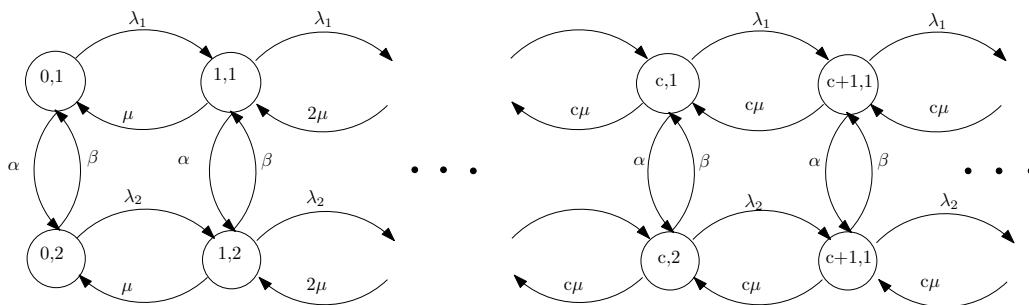

Figure 4: State-transition-rate diagram for $MMPP(2)/M/C$

If we write the lexicographical ordering of the states used, then the infinitesimal generator, Q, is given by

$$
Q = \left\{
\begin{matrix}
-(\lambda_1 + \alpha) & \alpha & \lambda_1 & 0 & 0 & 0 & 0 & 0 & \ldots \\
\beta & -(\lambda_2 + \beta) & 0 & \lambda_2 & 0 & 0 & 0 & 0 & \ldots \\
\mu & 0 & (\lambda_1 + \alpha + \mu) & \alpha & \lambda_1 & 0 & 0 & 0 & \ldots \\
0 & \mu & \beta & -(\lambda_2 + \beta + \mu) & 0 & \lambda_2 & 0 & 0 & \ldots \\
0 & 0 & \mu & 0 & -(\lambda_1 + \alpha + \mu) & \alpha & \lambda_1 & 0 & \ldots \\
0 & 0 & 0 & \mu & \beta & -(\lambda_2 + \beta + \mu) & 0 & \lambda_2 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{matrix}
\right\}
$$

$$
= \left\{
\begin{array}{ccccccc}
D_0 & D_1 & 0 & 0 & 0 & 0 & ... \\
A_2 & A_0 & A_1 & 0 & 0 & 0 & ... \\
0 & A_2 & A_0 & A_1 & 0 & 0 & ... \\
0 & 0 & A_2 & A_0 & A_1 & 0 & ... \\
0 & 0 & 0 & A_2 & A_0 & A_1 & ... \\
0 & 0 & 0 & 0 & A_2 & A_0 & ... \\
... & ... & ... & ... & ... & ... & ...
\end{array}
\right\}
$$

where

$$
D_0 = \left\{
\begin{array}{cc}
-(\lambda_1 + \alpha) & \alpha \\
\beta & -(\lambda_2 + \beta)
\end{array}
\right\}
$$

$$
D_1 = \left\{
\begin{array}{cc}
-\lambda_1 & 0 \\
0 & \lambda_2
\end{array}
\right\}
$$

$$
A_0 = \left\{
\begin{array}{cc}
-(\lambda_1 + \alpha + \mu) & \alpha \\
\beta & -(\lambda_2 + \beta + \mu)
\end{array}
\right\}
$$

$$
A_1 = \left\{
\begin{array}{cc}
-\lambda_1 & 0 \\
0 & \lambda_2
\end{array}
\right\}
$$

$$
A_2 = \left\{
\begin{array}{cc}
-\mu & 0 \\
0 & \mu
\end{array}
\right\}
$$

3.4 Steady State Probabilities

In a queueing system, customers arrive from a specified population to a service facility to receive service. The facility can have more than one server to cater the customers. If when a customer arrives, all the servers are busy, it joins the queue till a server becomes free. we do not allow service with feedback. We also consider that the systems follow the work conservation rule: a server cannot sit idle when there are customers in the queue.

Steady state probability is the probability that the system will be in a particular state for a particular number of customers in the system. The steady state probability of the system for n customers is denoted by P_n. We first identify the system states and the parameters associated. In our case, we are taking 2 states, State 1 and State 2. The other parameters include the arrival rates, service time, number of servers, the duration of the distribution of the two states and the number of customers. Here in our case there is no limitation on the number of customers in the system. We limit the number of customers in the system only for the final analysis part where the steady state is being found out where specific number of customers arrive in the system. We can use this data to draw a state-transition-rate diagram to represent the states as in a Markov chain. We can write the steady state relations in a linear form to be represented as a Matrix. The matrix that contains these entries is known as the *generator matrix*. Now the resultant matrix is solved using Gaussian elimination to obtain the steady state probabilities.

3.5   Estimating Waiting Time Probability

Our analysis is to find the optimal number of servers required so that the waiting probability is below a certain level. We are able to get the steady state probabilities at each state of the MMPP(2)/M/C. Each state is characterized by the ordered pair (i,j) where it denotes that there are i customers in j^{th} state. Therefore $P(ij)$ denotes the probability that there are i customers in j^{th} state. Customer Waiting Probability can be calculated using the formula

$$P_{wait} = \sum_{i=c+1,j=1}^{i=\infty,j=2} P_{ij}$$

4.   Numerical Experiment

We assume that customer require one truck each time (without loss of generality). Consider an example in which that on a on-season day, an average of ten orders are are placed by incoming customers according to a Poisson process while on an off-season day, the number goes down to six. It takes a mean of three days for completion of an order and return back to the service station. We also make an assumption that in a year, three months are off-seasonal. We want to calculate the number of trucks that are required at any moment in available for a service, so that the customer waiting probability is less than 5 %. Here we take the value of n, i.e. the maximum trucks available by the company as 50.

In this problem, we construct a generator matrix with the said values and apply Gaussian elimination to obtain the steady state probabilities. The steady state probabilities are then used to show the relation between the number of trucks required, which is plotted on the x-axis and the customer waiting probability which is plotted on the y-axis.
In any transportation company, one of the prime motto of the company to improve customer service will be to have very low customer waiting time. There will be a certain number of trucks, say n, which are used to provide services to an inbound customer. We should find out the optimum number of trucks that are required for services by the company.

Therefore, for the above example, the various parameters are λ_1=10, λ_2=6, μ=1/3, α=1/270, β=1/90 and n=100. From the results obtained, it may be inferred that the customer waiting probability is less than 5 %, when the number of trucks available are 40 or higher. So as a solution to the problem we have discussed here, a minimum of 40 trucks are required so that the waiting time probability of a customer is less than 5 %, which was our objective (see Table 1 and Figure 5).

5.   Conclusions and Future Work

In this research, we present an approximate way to model the fleet sizing problem in a stochastic setting with non-stationary demands and exponential truck unavailable times. Using an $MMPP(2)/M/C$ queue, we were able to obtain the optimal number of trucks to provide a minimum service level. Our work can be extended to consider other dimensions such as orders requiring multiple trucks at an instant, truck unavailable times modeled with general distribution, and order arrival times modeled with general distribution.

Table 1: Number of trucks vs. Waiting Probability

Number of Trucks	Waiting time Probability ($\times 100\%$)
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	0.9999
18	0.9974
19	0.9839
20	0.9572
21	0.9254
22	0.8948
23	0.8677
24	0.8445
25	0.8247
26	0.8072
27	0.7903
28	0.7701
29	0.7362
30	0.6695
31	0.5676
32	0.4572
33	0.3583
34	0.2756
35	0.2086
36	0.1553
37	0.1138
38	0.0820
39	0.0581
40	0.0405
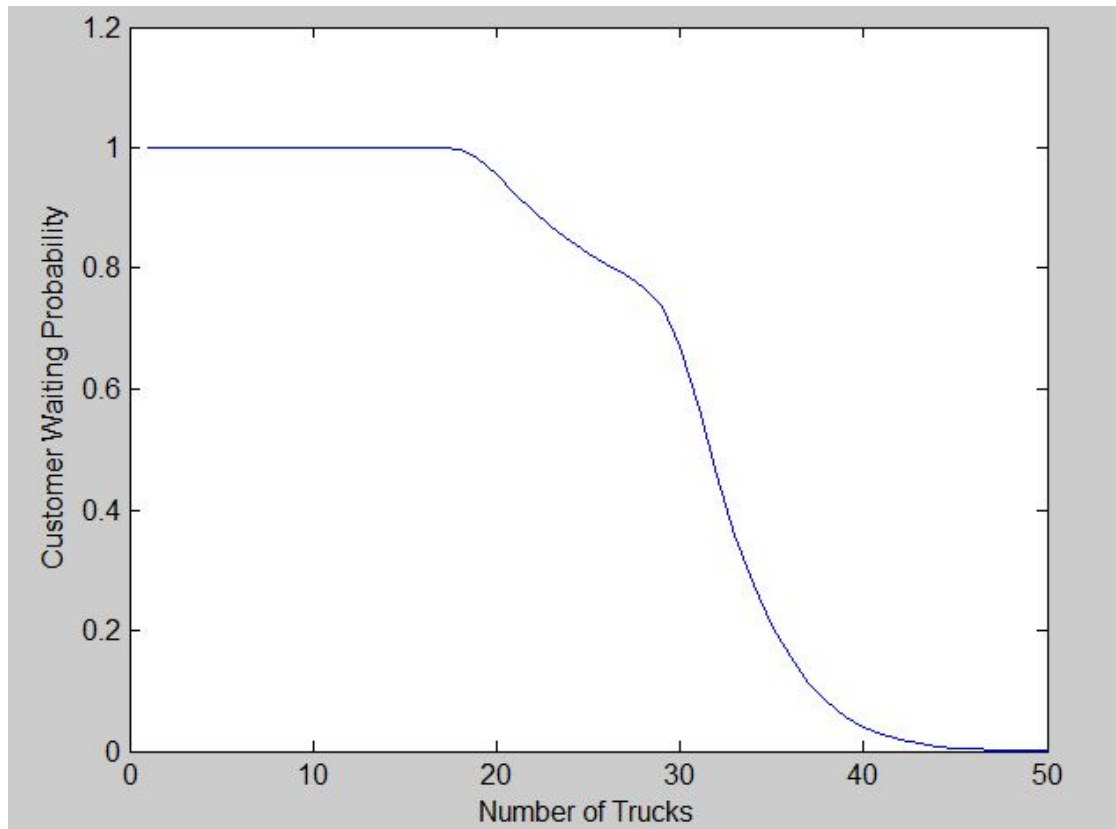
Figure 5: Waiting Probability

Acknowledgements

References

[1] H. S. Abhyankar and S. C. Graves. Creating an inventory hedge for markov-modulated poisson demand: An application and model. *Manufacturing & Service Operations Management*, 3(4):306–320, 2001.

[2] Mark E. Crovella and Azer Bestavros. Self-similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Transactions on Networking*, 5(6):835–846, 1997.

[3] J. Gould. The size and composition of a road transport fleet. *OR*, 20(1):81–92.

[4] O. C. Ibe. *Markov Processes for Stochastic Modeling*. Academic Press, Burlington, MA, 2009.

[5] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the self-similar nature of ethernet traffic. In *ACM SIGCOMM Computer Communication Review*, volume 23, pages 183–193. ACM, 1993.

[6] D. M. Lucantoni. New results for the single server queue with a batch markovian arrival process. *Stoch. Mod.*, 7:1–46, 1991.

[7] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts. A single-server queue with server vacations and a class of non-renewal arrival processes. *Advances in Applied Probability*, 22(3):676–705, 1990.

[8] n.a. Logistics in india 2012. Technical report, India Infrastructure Publishing Pvt. Ltd.

[9] M. F. Neuts. *Matrix-geometric solutions in stochastic models : an algorithmic approach.* Johns Hopkins University Press, Baltimore, MD, 1981.

[10] S. C. Parikh. On a fleet sizing and allocation problem. *Management Science*, 23(9):972–977, 1977.

[11] D. Ronen. The effect of oil price on containership speed and fleet size. *JORS*, 62(1):211–216, 2011.

[12] Iris F. A. Vis, M.B.M. de Koster, and Martin W. P. Savelsbergh. Minimum vehicle fleet size under time-window constraints at a container terminal. *Transportation Science*, 39(2):249–260, 2005.