Modeling Optimal Lane Configuration at the Toll Plaza by Nonlinear Integer Programming Incorporated with an M/G/1 Queueing Process

Seongmoon Kim
School of Business, Yonsei University
Shinchon-dong 134 Seodaemun-gu, Seoul, Korea
(phone) 82-2-2123-5485, (fax) 82-2-364-7828, kimsm@yonsei.ac.kr

Abstract

This paper provides an $M/G/1$ queueing model for the operations management problem at the toll plaza. This queueing process is incorporated with two nonlinear integer programming models - the user cost minimization model during the peak times and the operating cost minimization model during the off-peak hours.

1 INTRODUCTION

Ever increasing traffic volume and corresponding congestion necessitate efficient design and management of toll plaza operations. Optimal lane configuration against the non-stationary and stochastic traffic volume can help prevent long waits during the peak times at the toll plaza. Most of queueing studies, which are considered to best describe the traffic situations at the toll plaza, have not been applied to toll plaza operations. Some previous studies use $M/M/1$ or $M/M/s$ queueing systems to model the toll plaza problem, but the assumption on the exponential service time is unrealistic in practice. Several studies employ heuristics or simulation studies to find *good* sub-optimal solutions. This paper is differentiated from the previous works by providing an $M/G/1$ queueing model, which is realistic and still mathematically tractable, and it is then integrated with nonlinear integer programming models to determine *optimal* lane configurations over time at the toll plaza.

The rest of the paper is organized as follows. An $M/G/1$ queueing model is presented in Section 2. The queueing model is then incorporated with two nonlinear integer programming models in Section 3 - the user cost minimization model during the peak times and the operating cost minimization model during the off-peak hours. Lastly, concluding remarks and future research directions are provided in Section 4.

2 An M/G/1 Queueing Model

The author provides in this section framework of the toll plaza configuration problem which can be best explained by the queueing theory. The main components of the queueing model are arrival and service processes, number of servers, queue discipline, and so on. We begin this section with the following variables and parameters which are used in the model.

2.1 Notation

- N: total number of lanes at the toll plaza,
- K: number of lane types,
- n_i: planned capacity for lane type i, that is, the number of type i lanes to open, where $\sum_{i=1}^{K} n_i \leq N$,
- l_i: lower bound for the number of type i lanes to open,
- u_i: upper bound for the number of type i lanes to open,
- λ: mean arrival rate for service at the toll plaza,
- λ_i: mean arrival rate for lane type i, where $\sum_{i=1}^{K} \lambda_i = \lambda$,
- p_i: mean percentage of drivers using lane type i, where $p_i = \lambda_i/\lambda$,
- μ_i: mean service rate for lane type i,
- σ_i: standard deviation of service time for lane type i,
- L_i: mean number of vehicles waiting in the queue at a type i lane,

- W_i: mean waiting time in the queue at a type i lane,
- W: total mean waiting time in the queue for all the arrivals at the toll plaza,
- s_i: the service standard for the mean waiting time for lane type i,

2.2 Lane Types

There are many kinds of lane types in different cities, states and countries. Lanes at the toll plaza are classified in (*1*) as follows: 1) dedicated manual toll lanes where transactions are handled by a toll collector, 2) dedicated automatic coin (or token) machine lanes (hereafter referred to as ACM lanes, 3) dedicated electronic toll collection (ETC) lanes with automatic vehicle identification technology, and 4) any mix of the above mentioned dedicated lane types, such as mixed ACM and ETC lanes, mixed ACM and manual lanes, and mixed ETC and manual lanes. These classification can be divided further into barrier and no-barrier toll lanes. Manual lanes are further classified in (*2*) depending on whether they provide service to semi-trucks or not.

Manual lanes need toll collectors who can issue change or receipt. Tolls are paid manually in ACM lanes, but ACM collects the toll instead of a toll collector. Drivers have to stop to pay the toll at the plaza in both manual and ACM lanes. Recently, there are some other type of toll lanes where credit card can be used with or without receipt. The average speed through the conventional lane types is the least because of the stop and pay system and, hence, queues extend up to long distances during peak hours.

These days ETC lanes are getting popularity more and more to many toll users who do not want to experience long waits at the toll plaza. ETC lanes utilize a radio frequency transponders and external sensors to collect toll. The external sensor decides whether the transponder attached to the vehicle are valid or not, then the class of the vehicle is identified and the toll is electronically debited from the driver's account. Transponders are like electronic tags attached to the vehicles which assess the toll by identifying the class of the vehicle. This ETC system has different names, for example, $E-Z\ PASS, HIGH\ PASS$, etc., in different cities, states and countries.

2.3 Arrival Process

For real queueing systems, the probability distribution of interarrival times can take on almost any form. But, to formulate a queueing theory model as a representation of the real system, we need to make some assumptions on the probability distribution which should be sufficiently realistic while, at the same time, being mathematically tractable.

The mathematics of queueing theory are most manageable if arrivals exhibit Poisson processes with exponential interarrival times. Arrivals may be counted from multiple toll booths for each lane type. From this observation, we can compute mean interarrival times. Statistical test recommends that the arrivals at the toll plaza follow the standard assumption of a Poisson process with exponential interarrival times. This is well supported in the literature (*3-8*).

Let λ denote the mean total arrival rate for service at the toll plaza, and λ_i mean arrival rate for lane type i, $i = 1, 2, \cdots, K$, where $\sum_{i=1}^{K} \lambda_i = \lambda$. Analysis of the arrival data also suggests that arrival percentage per lane type is uniformly distributed over all lanes (*3*). Consequently, the mean arrival rate per lane is determined by dividing the mean arrival rate per lane type by the number of lanes utilized for that type of collection, and we have $\frac{\lambda_i}{n_i}$.

2.4 Service Process

Service time distributions may also be obtained empirically in the same manner as described earlier. The mean service time for each payment type is computed through repeated observation at the plaza during peak load when no slack exists. Service time does not include waiting time of the vehicle in the queue.

If service times are exponentially distributed, the mathematics of the queueing theory are most tractable. However, they do not, in reality, follow the nice exponential distributions. An exponential service time distribution is used in (*3*) where the variance of the service time is greater than the real one, and upper bounds for mean waiting time are computed. Instead of using an exponential distribution, we assume in this study that service times have general distributions by measuring the mean service rate, μ_i, and the standard deviation, σ_i, for lane type i. It is reported in (*9*) from their collection of real data that the mean service rates per hour for ETC, ACM, and manual lanes were 1708, 503, and 376, respectively. Although these numbers are toll plaza specific, it is generally understood that mean service rate for the ETC lane is the highest and for the manual lane is the lowest. Generally again, the variance for the ETC lane is the smallest and for the manual lane is the largest.

2.5 Queue Length and Waiting time

With the assumption of the general service time distribution and the Poisson arrival process where the average arrival rate per lane type is uniformly distributed across each lane, we can formulate an $M/G/1$ queueing process model. If the mean arrival rates, mean service times and standard deviations are provided, mean queue length and mean waiting times in the queue can be obtained. Let L_i denote the mean queue length, i.e., mean number of vehicles waiting in the queue for a type i lane. Then, by using the *Pollaczek–Khintchine formula* for the $M/G/1$ queueing system *(10)*, we have

$$L_i = \frac{\lambda_i^2(\mu_i^2\sigma_i^2 + 1)}{2n_i^2\mu_i(\mu_i - \frac{\lambda_i}{n_i})} \tag{2.1}$$

Using the above formula, the queue length in each of the toll lanes is calculated.

Since service rate is different for each lane type, the mean queue length may not be a good measure of criteria for performance comparison. Mean waiting time in the queue, not the mean queue length, impacts on the perception of service quality at the toll plaza to the motorists. For example, while the queue in the ACM lane is longer than the one for the manual service, the mean waiting time for the ACM lane may be shorter due to a higher service rate with a smaller variance of the service.

Mean waiting time of the vehicle in the queue for lane type i can be obtained by the *Little's Law (10)*

$$W_i = \frac{L_i}{\frac{\lambda_i}{n_i}} = \frac{\lambda_i(\mu_i^2\sigma_i^2 + 1)}{2n_i\mu_i(\mu_i - \frac{\lambda_i}{n_i})} \tag{2.2}$$

and the total waiting time of all drivers at the toll plaza is

$$W = \sum_{i=1}^{K}\lambda_i W_i = \sum_{i=1}^{K}\frac{\lambda_i^2(\mu_i^2\sigma_i^2 + 1)}{2n_i\mu_i(\mu_i - \frac{\lambda_i}{n_i})}. \tag{2.3}$$

3 NONLINEAR INTEGER PROGRAMMING MODELS

From discussion with the toll plaza officials we find that the goals of operating a toll plaza may not be the same between peak and off-peak times. During the peak hours, for example, the objective may be minimizing the total (cost of) wait time with full capacity operation due to high traffic volume. That is, a decrease in user cost related to wait time as opposed to operating costs is preferred during the morning and evening rush hours. On the other hand, during the off-peak hours the traffic volume may not be that high and service standard may be well satisfied even with partial capacity operation. In other words, as long as the service standard is met, full capacity operations may not be necessary and the objective is simply minimizing the operating costs.

3.1 The User Cost Minimization Model

The user cost is directly related with the mean wait time in the queue. In order to find the optimal lane configuration $(n_1^*, n_2^*, \cdots, n_K^*)$ during the peak times at the toll plaza given several constraints, we develop a following nonlinear integer programming model minimizing the total waiting time.

$$\text{Minimize} \quad W = \sum_{i=1}^{K}\lambda_i W_i \tag{3.1}$$

$$\text{Subject to} \quad n_i \geq l_i \text{ for all } i \tag{3.2}$$

$$n_i \leq u_i \text{ for all } i \tag{3.3}$$

$$\frac{\lambda_i}{n_i} < \mu_i \text{ for all } i \tag{3.4}$$

$$\sum_{i=1}^{K} n_i = N \tag{3.5}$$

$$W_i = \frac{\lambda_i(\mu_i^2\sigma_i^2 + 1)}{2n_i\mu_i(\mu_i - \frac{\lambda_i}{n_i})} \tag{3.6}$$

$$n_i = \text{integer for all } i. \tag{3.7}$$

There are several constraints to achieve this goal at the toll plaza. Lanes to open for each collection type may have some lower and upper bounds, i.e., l_i and u_i. For example, at least one lane should be open for each lane type to provide service for all types of drivers, and the number of ACM and ETC lanes available is limited by the number of lanes equipped with ACM and ETC lane machines. Next, the arrival rate to each lane should be less than the service rate for the lane. Otherwise, queues will grow indefinitely to explode and the system will be unstable. Hence, in order to have a stable steady-state queueing system, we need the inequality (3.4). In addition, sum of open lanes for all lane types should be the total available lanes, i.e., full capacity operation during the peak times. Obviously, the number of lanes to open should be integer. Our objective is to find the optimal lane configuration $(n_1^*, n_2^*, \cdots, n_K^*)$, i.e., the number of lanes to open for different types in order to minimize the total wait times for all drivers. The optimal solution may be obtained by using any of commercial optimization packages.

3.2 The Operating Cost Minimization Model

The operating cost is directly related with the number of lanes open at the toll plaza. During the off-peak hours it may be desired to keep the minimum number of lanes to open for each lane type as long as the service standard is met in order to save operating cost. We develop a following nonlinear integer programming model to achieve this goal.

$$\text{Minimize} \quad \sum_{i=1}^{K} n_i \qquad (3.8)$$

$$\text{Subject to} \quad n_i \geq l_i \text{ for all } i \qquad (3.9)$$

$$n_i \leq u_i \text{ for all } i \qquad (3.10)$$

$$\frac{\lambda_i}{n_i} < \mu_i \text{ for all } i \qquad (3.11)$$

$$\sum_{i=1}^{K} n_i \leq N \qquad (3.12)$$

$$W_i \leq s_i \text{ for all } i \qquad (3.13)$$

$$W_i = \frac{\lambda_i(\mu_i^2\sigma_i^2 + 1)}{2n_i\mu_i(\mu_i - \frac{\lambda_i}{n_i})} \qquad (3.14)$$

$$n_i = \text{integer for all } i. \qquad (3.15)$$

Some of the constraints are modified and newly added from the previous model. First, inequality (3.12) implies that not all the lanes need to be open. In addition, inequality (3.13) states that average waiting time for each lane should be less than or equal to the predetermined service standard. Finally, the objective function (3.8) minimizes the number of lanes to open to keep the lowest operating costs.

4 CONCLUSIONS

This paper presented an $M/G/1$ queueing model, which is more realistic than a commonly-used $M/M/1$ model with a Poisson arrival process and a general service time distribution. We incorporated the $M/G/1$ queueing process with the two nonlinear integer programming models - the user cost minimization model during the peak times and the operating cost minimization model during the off-peak hours.

Although we obtain the optimal lane configuration based on the historical average traffic volume data, real-time traffic volume or real-time proportion of toll users may be considerably different from the historical average values, contingent on traffic accidents, weather conditions, sports events, and so on. When we have these unexpected changes of the total traffic volume or proportion of toll users, how to dynamically re-configure the toll plaza to maintain not long wait time is another future research direction as introduced in (*11*) with decision support system. In addition, finding the optimal work force scheduling to minimize operating cost in the manual lane based on real-time traffic observation may be an interesting but challenging research topic.

REFERENCES

1. J. Klodzinski and H. M. Al-Deek, Proposed Level of Service Methodology for a Toll Plaza, *Transportation Research Record* 1802, pp.86-96 (2002).
2. M. L. Zarrillo, A. E. Radwan, and H. M. Al-Deek, Modeling Traffic Operations at Electronic Toll Collection and Traffic Management Systems, *Computers and Industrial Engineering* 33, pp.857-860 (1997).
3. J. Boronico and P. H. Siegel, Capacity planning for toll roadways incorporating consumer wait time costs, *Transportation Research Part A* 32, pp.297-310 (1998).
4. A. Hasofer, On the single-server queue with non-homogeneous Poisson input and general service time, *Journal of Applied Probability* 1, pp.369-384 (1964).
5. B. Schwartz, Queueing models with lane selection: A new class of problems, *Operations Research* 22, pp.331-339 (1974).
6. W. Grassman, Transient and steady-state results for two parallel queues, *International Journal of Management Science* 8, pp.105-112 (1980).
7. L. Green, A queueing system with general-use and limited-use servers, *Operations Research* 33, pp.168-182 (1985).
8. G. Blackwell, Traffic monitoring system helps Ontario motorists, *Computing Canada* 14, pp.28+ (1988).
9. J. Klodzinski and H. M. Al-Deek, Evaluation of toll plaza performance after addition of express toll lanes at mainline toll plaza, *Transportation Research Record* 1867, pp.107-115 (2004).
10. F. Hillier and G. J. Lieberman, Introduction to operations research, McGraw-Hill, New York (2005).
11. M. L. Zarrillo, Using decision support system and graphical user interface to choose appropriate lane configurations at toll facilities, *Transportation Research Record* 1935, pp.8-16 (2005).