Received 14 August 2021
Revised 5 October 2021
Accepted 5 November 2021

Optimising truck arrival management and number of service gates at container terminals

Chu Cong Minh
Department of Bridge and Highway Engineering, School of Civil Engineering,
Ho Chi Minh City University of Technology,
Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam, and
Nguyen Van Noi
Department of Information Technology, Faculty of Technology Engineering,
Tien Giang University, My Tho, Vietnam

Abstract

Purpose – Truck appointment systems have been applied in critical container ports in the United States due to their potential to improve handling operations. This paper aims to develop a truck appointment system to optimise the total cost experiencing at the entrance of container terminals by managing truck arrivals and the number of service gates satisfying a given level of service.

Design/methodology/approach – The approximation of $M_t/G/n_t$ queuing model is applied and integrated into a cost optimisation model to identify (1) the number of arrival trucks allowed at each time slot and (2) the number of service gates operating at each time slot that ensure the average waiting time is less than a designated time threshold. The optimisation model is solved by the Genetic Algorithm and tested with a case study. Its effectiveness is identified by comparing the model's outcomes with observed data and other recent studies.

Findings – The results indicate that the developed truck appointment system can provide more than threefold and twofold reductions of the total cost experiencing at the terminal entrance compared to the actual data and results from previous research, respectively.

Originality/value – The proposed approach provides applicably coordinated truck plans and operating service gates efficiently to decrease congestion, emission and expenses.

Keywords Truck appointment system, Truck arrival management, Container terminals, Service gates, Queuing model

Paper type Research paper

Introduction

The operational effectiveness of container terminals plays a vital role in the success of supply chains. However, the tremendous growth of international and domestic trade has produced many problems for container terminals. One of them is the fluctuation in drayage truck arrivals during working hours, which causes both over and underutilisation of the terminal's capacity. Large numbers of trucks arriving at the maritime port during peak periods cause congestion at the entrance. Heavy congestions lead to high waiting time and long queues and high air pollution. Although terminals can strengthen their capacity by investing in more service gates or expanding the waiting area, the investment usually is substantial and limited by land availability. To overcome this problem, some major container terminals apply truck appointment systems, like New Orleans, Los Angeles, New Bedford, Vancouver and Long Beach. These systems were studied in many research studies, such as Giuliano and O'Brien (2007), Chen *et al.* (2011, 2013a, b) and Zhang *et al.* (2013). With a truck appointment system, terminal operators divide daily working hours at the container terminal into time slots and set

Maritime Business Review
Vol. 8 No. 1, 2023
pp. 18-31
Emerald Publishing Limited
2397-3757
DOI 10.1108/MABR-08-2021-0060

The authors would like to thank Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for supporting time and facilities for this study.

entry quotas for each time slot. Drayage truckers can make reservations for their arrivals at a preferred time slot. Using this system, terminal operators limit the maximum number of truck arrivals at every time slot by flattering truck arrival rates in a working day, especially at peak periods. Besides, Guan and Liu (2009a, b), Fleming *et al.* (2013) and Minh and Huynh (2014, 2017) proposed another way to reduce the congestion at the entrance by rearranging the entry gate layout plus optimising the number of service gates. Those studies assess the effectiveness of different gate layouts, including pooled and non-pooled truck queuing strategies, identify the optimal design for marine container terminals and compute the number of service gates necessary to achieve a given level of service. Those methods help terminal operators make more beneficial utilisation of labour, equipment and facilities. Drayage truckers benefit from their shorter turn times if long queues at the entrance disappear.

This study proposes the method applied to a truck appointment system to minimise the total cost experiencing at the entrance of container terminals by organising truck arrivals and estimating the number of operating service gates that satisfies a given level of service. The diffusion approximation of $M_{(t)}/G/n_{(t)}$ queuing model is applied and integrated into a total cost optimisation model. It determines (1) the number of arrival trucks allowed at each slot and (2) the number of service gates operating at each time slot satisfying a designated level of service for a given truck arrival and truck service rate. The optimisation model is solved by the Genetic Algorithm (GA), tested with a case study, and compared with other recent studies in the literature.

The rest of this paper is structured as follows. The following section provides a summary of previous studies that are related to this research. The Methodology section describes the truck appointment system and details of the optimisation model. It continues with the Case Study section that represents the application of the proposed method to the actual terminal and the comparative results with other models in the literature. The following section, Sensitivity Analysis, examines how different parameters influence the model results and how better the proposed model is to previous ones. Finally, the Conclusion section gives concluding notes and suggestions for future studies.

Literature review

Considerable research in the literature was studied truck queuing problems at the entrance gates of marine port terminals. Some have concentrated on operational approaches to decrease truck waiting time by expanding operating hours or managing truck arrival time. Following this trend is the study of Huynh *et al.* (2016), which provides a comprehensive analysis of both the state of the art and the state of the practice of truck appointment systems. The authors discussed and compared the advantages and disadvantages of truck appointment studies. They also described vital elements and needs for successful and effective strategies and recommended further research. Similarly, Abdelmagid *et al.* (2021) presented a comprehensive review of truck appointment models classified into three categories, i.e. control and decision perspectives, modelling methodologies and collaboration between terminal stakeholders. Additionally, they discussed the role of IoT in smart terminals and suggested directions for further studies. They concluded that truck appointment systems are the most reliable solution to boost the performance of a terminal. Bentolila *et al.* (2016) evaluated the “Good Night Program”, applied at the Port of Haifa, Israel, to increase traffic movements at the terminal. The authors estimated the market utility and examined customer satisfaction regarding the extension of working hours and diverting drayage moves at the port from daytime to nighttime with an incentive program. Torkjazi *et al.* (2018) propose an approach to serve the maritime container terminal operators and drayage operators by distributing the truck arrivals throughout the day and providing appointment times so that truck drivers do not have to change their original schedule significantly. The model is formulated as a non-linear program and solved using Lingo software. Chen *et al.* (2011, 2013a, b) proposed a model using time-varying tolls to optimise truck arrivals at maritime ports. They introduced the fluid-based approximation functions to analyse time-dependent truck queuing processes with stochastic service time distributions. The authors applied the single-server queuing model (*M/G/1*), given that the truck inter-arrival times followed the exponential distribution, and the service times followed an independent and identical distribution. The non-linear programming model was developed to minimise the total truck turn-time and discomfort due to shifted arrival times. Numerical experiments at the port of Los Angeles are conducted to test the proposed optimisation model's computational efficiency and accuracy. Guan and Liu (2009a, b) proposed a multi-server queuing model to examine congestion costs at entrance gates and calculate truck waiting costs at New York port. The authors applied the *M/Ek/n* queuing model and developed an optimisation model to optimise the total price, including the cost of trucks waiting in front of the gates and the cost of operating service gates. The authors finally recommended optimal operational service gates to minimise the port authority's cost to handle the given truck arrival rate. Im *et al.* (2021) developed a cooperation model that simultaneously considered the perspective of forwarders and terminal operators. They compared numerical results between cases in which truck companies and port operators make their own decisions and cooperate. Finally, the authors provided a solution that profits both participants. The model, however, does not reflect the subordinate relationship between stakeholders. While most studies investigate truck appointment systems from the viewpoint of drayage firms and port operators, Jovanovic (2018) focuses on truck drivers' characteristics, including their number of arrivals, the number of daily working hours and travel distances. An optimisation model and the integer program were developed to indicate that a higher number of scheduled truck visits can be achieved at the ports that installed a truck appointment system.

Other studies focus on planning and designing approaches to optimise the gate layout. Fleming *et al.* (2013) proposed an agent-based simulation approach to evaluate two types of pooled and non-pooled queues at an existing terminal gate system. They used a car-following model to represent how trucks move and queue up at the container terminal entrance. Netlogo, the agent-based simulation software, was used to evaluate queuing strategies under various operational conditions. The authors concluded that a pooled queue produced considerably lower average truck queuing times than a non-pooled queue. Minh and Huynh (2014) developed a planning-level tool that can be applied to assess the effectiveness of different gate layouts and determine the optimal design for marine container terminals. The analytical tool is used to estimate the average truck queuing time for a given gate layout and optimised the number of required service gates and average truck queue length needed to obtain the given level of service for a particular truck arrival rate and truck service rate. This tool accounts for trucks arriving as a Poisson distribution, served by n gates, and gate service time is independent and identically distributed random variables (*M/G/n* queuing model). Minh and Huynh (2017) provided a methodology for terminal operators and port planners to design a gate layout to reduce truck queuing time and gate operation costs. It provides practitioners with a preciser approach than the state of the practice, which assumes the truck inter-arrival times follow the Poisson distribution. The gate service times follow the Erlang distribution ($M/E_k/n$ model). Wang *et al.* (2019) proposed the mathematical model based on the energy consumption of numerous operating devices to select the optimal terminal layout in terms of total carbon emission. The proposed model considers the interaction between different areas of the typical terminal while the size of the width and length of the terminal are unchanged. Following this trend, Chamchang and Niyomdecha (2021) evaluated several service policies at a terminal gate, including lane expansion, exit-gate sharing policies and reorganisation of work processes using a simulation approach. They concluded that redesigning a work process by separating customs work from gate operations was most significant in reducing truck queue lengths and waiting time.

The present study combines both operational and planning approaches by (1) providing a truck appointment system at the terminal to schedule truck arrivals, (2) using a pooled queue policy and (3) selecting the number of service gates, which minimises the total cost experiencing at the entrance of a container terminal and satisfies the given level of service. This study modifies Minh and Huynh's work (2014 and 2017), which provided the analytical tool to determine how many service gates are needed to achieve a desired level of service. Additionally, this study estimates the truck arrivals and the number of operating gates at every time slot which minimises the total cost experiencing at the entrance of container terminals and satisfies the given level of service. More specifically, the proposed methodology integrates two critical variables, the truck arrival rates and the number of operating service gates at different time slots which fulfil the given level of service, into the optimisation model.

Methodology

This section describes the methodology applied to the truck appointment system to optimise total cost experiencing at the entrance gates of container terminals by managing truck arrivals and the number of service gates.

Description of truck appointment model

With a truck appointment system implemented, operators at container terminals divide working hours into many appointment time slots, 1, . . ., T. T is the total number of time slots within an analysing period. They set the maximum appointment quotas for each time slot. All truckers are required to reserve the time slot to enter the terminal. If the preferred time slot allowance is not available, truckers need to select another time slot. That is, truck arrivals with preferred time slots $(\lambda_1, \ldots, \lambda_T)$ will be adjusted to truck arrivals with appointed time slots $(\lambda_1^*, \ldots, \lambda_T^*)$ in the way that total truck preferred-time arrivals equal total truck appointed-time arrivals. One of the optimisation model's objectives is to determine the maximum appointment quota of each time slot, meaning to estimate $\lambda_1^*, \ldots, \lambda_T^*$.

The objective of this methodology is to minimise the total cost (z), including the cost for shifting truckers' arrival times from their preferred arrival time, gate operating cost and truck operating cost by managing the arrival rate (λ_t^*) and number of service gate (n_t):

Objective function:

$$\min z = C_a \times \sum_{t=1}^{T} \left| \lambda_t - \lambda_t^* \right| + C_g \times \sum_{t=1}^{T} n_t + C_t \times \sum_{t=1}^{T} N_t \tag{1}$$

where

C_a: Cost of shifting truckers' arrival times from their preferred arrival time to another time ($/truck)

C_g: Gate operating cost ($/hour/gate)

C_t: Truck waiting cost per hour per truck ($/hour/truck)

T: Index of the total analysing time period

t: Index of time frame defined by the terminal operation characteristics

λ_t: Mean of preferred truck arrival rate at time frame t (trucks per hour)

λ_t^*: Mean of truck appointed arrival rate at time frame t (trucks per hour)

n_t: Number of service gates at time frame t

N_t: Average queue length at the gates at time frame t

All constraints will be described in the next section, combining with queuing length estimation. The diffusion approximation formula is utilised to estimate average queue length at the gates at time frame t, N_t.

Queuing length estimation

The type of a queuing model is determined by the statistical distributions of truck arrival intervals and gate service times. The gate layout at maritime terminals can be considered a multi-server queuing system. Each service gate is treated as a server by assuming that truck drivers will automatically select the service lane with the lowest queuing length. It assumes that arrival trucks follow an exponential distribution at each time slot with the mean λ_t^*. They are served by n service gates and gate service times follow a nonnegative independent and identical distribution with the mean μ. The multi-server queuing model (*M/G/n*), therefore, is used. Based on work by Kimura (1983) and Yao (1985) developed a diffusion approximation method to estimate the average waiting time for this system, as shown below:

$$QT_t \approx \pi_0 \times \theta_{n_t} \times \frac{\left(\frac{\rho_t}{1-\rho_t}\right)}{\lambda_t^* \left[1 - \exp(r_{n_t})\right]} \tag{2}$$

where,

$$\rho_t = \frac{\lambda_t^*}{n_t \times \mu} < 1, \text{ is traffic intensity or server utilisation} \tag{3}$$

$$\pi_0 = \left[\sum_{k=0}^{n_t-1} \theta_k + \frac{\theta_{n_t}}{1-\rho_t} + \frac{n_t \times \rho_t}{r_1} \times \left(\exp\left(\frac{r_1}{2}\right) - \exp\left(\frac{-r_1}{2}\right) - r_1\right)\right]^{-1}, \text{ is the steady} - \text{state of the queuing process} \tag{4}$$

With $k = 0, \ldots, n_t,$ the following coefficients are defined as follows,

$$\theta_k = \frac{(n_t \times \rho_t)^k}{k!} \tag{5}$$

$$r_k = \frac{2b_k}{a_k} \tag{6}$$

$$b_k = \lambda_t^* - k \times \mu \tag{7}$$

$$a_k = \lambda_t^* + k \times \mu \times c_s^2 \tag{8}$$

$$c_s = \mu \times \sigma \tag{9}$$

According to Little's law,

$$N_t = \lambda_t^* \times QT_t \tag{10}$$

Arrival constraints:

$$\sum_{t=1}^{T} \lambda_t = \sum_{t=1}^{T} \lambda_t^* \tag{11}$$

$$\lambda_t, \lambda_t^* \geq 0 \tag{12}$$

Service constraints:

$$0 < QT_t \leq QT^* \tag{13}$$

$$n_t = \text{Integer} \tag{14}$$

$$0 \leq n_t \leq n_{max} \tag{15}$$

$$\mu, \sigma > 0 \tag{16}$$

where QT_t: Average queuing time at time frame t (hour)

QT^*: Specified waiting time threshold (hour)

σ: Standard deviation of gate service time per truck

$n_{\max}$: Maximal number of service gates in the terminal, practically $n_{\max} < 15$.

Equation (11) represents that the total number of truck arrivals should equal the number of appointed trucks. Equation (12) ensures the number of truck arrivals and assigned for each period is nonnegative. Equation (13) interprets that the average waiting time of trucks (QT_t) must be less than the given time as a specified waiting time threshold (QT^*) to maintain the proper level of service and to avoid a fine applied at some port terminals. In other words, it means that the arrival rates and the number of service gates need to satisfy a given level of service.

The proposed model aims to minimise the total cost, z, in Equation (1) by selecting λ_t^* and n_t satisfying all constraints above.

Solution for the model

A method based on the Genetic Algorithm is designed to solve the model. The GA is applied to search for the optimal solution considering the constraints of adjustment quota of arrival rates and the number of service gates. The total cost is used as the fitness value of the algorithm and computed through Equation (1). The steps of the GA are as below:

Step 1: Initialisation

The GA is based on chromosomes representing a group number of maximum appointment quotas ($\lambda_1, \ldots, \lambda_T$) of the appointment period. The initial population is randomly generated, and each individual consists of a set of T random numbers. The population size is set to be 5,000. The criteria to stop the GA are (1) the maximum of iteration set to be 2,000, (2) the amount of time in seconds that the GA stops after running and is set to be 1,000 and (3) the fitness value of the best point which is unchanged after last iterations and set to be 10^{10}. This value is updated when the GA reaches a better point.

Step 2: Evaluation

The fitness value of each individual is computed based on Equation (1) and set to be a considerable number (10^{10}) if the individual does not subject to equation (2) - (16).

Step 3: Selection

The selection improves the fitness value by discarding unsuitable individuals with considerable fitness value and keeping the population's best individuals. The fitter individuals will be selected for the next generation. The GA parameters are set to be (*2/3* × *population size*) for the tournament size, meaning that two-thirds of individuals based on the smallest to the largest fitness value are kept for crossover.

Step 4: Crossover

Crossover creates new individuals by combining aspects of selected individuals. Integrating specified traits from two or more individuals will produce an even fitter offspring, including the most excellent features from its parents. The GA parameters are set to be 0.5 for uniform rate, meaning that the offspring's genome comes from a half genome of its father and a half genome of its mother.

Step 5: Mutation

A little randomness is added to the population by making minimal random changes to individuals' genomes. The GA parameters are set to be 0.1 for the mutation rate.

Step 6: Repeat

The next generation can again start from step 2 until it reaches a termination condition, in which the algorithm has found a solution that meets the criteria to stop the GA mentioned in Step 1.

Case study

The data in one working day used in this research were taken from Guan and Liu's study (2009). The maritime terminal starts at 6 a.m. and closes at 5 p.m. from Monday to Friday with no lunch break. Truck arrival rates, λ, are ranging from 1.5 to 2.43 trucks per minute and follow the Poisson distribution. The gate service time of drayage trucks follows the Erlang distribution with the mean (μ), and the standard deviation of gate service times (σ) at this terminal is 2.44 min per truck and 1.18 min per truck respectively.

1. Hourly gate operating cost C_g: $125.75 per hour.
2. Truck waiting cost C_t: $32.15 per truck hour.
3. Since no information on the cost of shifting truckers' arrival times from their preferred arrival time (C_a), it is assumed to be $0/truck.

The objective of this methodology is to minimise the total cost z (Equation 1) by selecting truck arrivals (λ_t^*) and service gates (n_t) to ensure the average truck queuing time (QT_t) is not more than a given time threshold (QT^*).

Table 1 shows the arrival rates, the number of operating gates and the total costs of the on-site survey and different models using 30 min of the waiting time threshold ($QT^* = 30$ min) that assuming the cost of shifting truckers' preferred arrival time to another time (C_a) is $0. In Table 1.

(1) expresses the actual data in one day observed at one of the maritime terminals in the port of New York referred to Guan and Liu's study (2009);

(2) represents the optimisation outputs of Guan and Liu's model (2009);

(3) shows the optimal results of Minh and Huynh's model (2014) regarding observed truck arrival rates at each hour;

Table 1. The comparison of observed performance, Guan and Liu's model, Minh and Huynh's model, optimisation with equal truck arrivals and the proposed model

Hour	Observed performance (i)			Guan and Liu's model (ii)			Minh and Huynh's model (iii)			Optimum with equal arrivals (iv)			The proposed model (v)		
	λ	n	z	λ	n	z	λ	n	z	λ^*	n	z	λ^*	n	z
06:00–07:00	1.50	4	697	105	5	731	1.50	4	678	1.69	5	686	0.26	1	148
07:00–08:00	1.58	4	1,015	105	5	731	1.58	4	1,021	1.69	5	686	0.32	1	182
08:00–09:00	2.30	6	1,055	105	5	731	2.30	6	992	1.69	5	686	0.23	1	141
09:00–10:00	1.57	6	773	105	5	731	1.57	4	889	1.69	5	686	0.23	1	141
10:00–11:00	1.75	6	788	105	5	731	1.75	5	706	1.69	5	686	2.10	6	828
11:00–12:00	1.58	6	772	105	5	731	1.58	4	1,021	1.69	5	686	3.68	10	1,373
12:00–13:00	2.42	6	2,014	105	5	731	2.42	6	1,833	1.69	5	686	2.13	6	840
13:00–14:00	2.43	6	3,282	105	5	731	2.45	7	951	1.69	5	686	3.69	10	1,378
14:00–15:00	1.35	5	648	105	5	731	1.35	4	562	1.69	5	686	2.48	7	960
15:00–16:00	1.32	4	558	105	5	731	1.32	4	552	1.69	5	686	3.26	9	1,229
16:00–16:30	0.83	4	257	43	4	380	0.83	3	397	1.69	5	686	0.25	1	145
Total gates and cost		57	*11,859*		54	*7,690*		51	*9,602*		55	*7,549*		53	*7,365*

(4) presents the optimal results taken from Minh and Huynh's model (2014) with the same truck arrival per hour by taking the total truck arrivals at the day divided by the number of operating hours per day;

(5) indicates the optimal results of the proposed model in this study.

The observed performance (1) shows that the terminal's current operating state and the total cost dealing with observed arrival rates is \$11,859. The total cost of Guan and Liu's model (2) is \$7,690. The expected total cost of Minh and Huynh's model regarding observed truck arrival rates (3) is \$9,602, and that regarding the same truck arrivals (same λ^*) (4) is \$7,549. The total cost of the proposed methodology (5) is \$7,365, the lowest among the five, implying that this is the best model for this case study. However, the total number of gates operating in the day, defined as the sum of gates serving at each hour in a day, in the proposed methodology, model v, is 53 gates and is the second-lowest model compared to model iii, 51 gates. It can be explained that with the assumptions as mentioned earlier, model iii optimises the number of service gates only and does not mean the total cost is optimised accordingly. It can be concluded that the present methodology can achieve the minimal cost experiencing at the terminal entrance. Still, it may not obtain the minimum number of service gates necessary to serve these truck arrivals and satisfy the given level of service.

Sensitivity analysis

The cost of shifting truckers' arrival times from their preferred arrival time (C_a) on the total cost (z) is analysed using the practical range of C_a, from 0 to \$30/truck in \$5 increments and shown in Figure 1. Other parameters are kept the same as observed values. In detail, truck arrivals, λ, are ranging from 1.5 to 2.43 trucks per minute and follow the exponential distribution. The gate service times of trucks follow the Erlang distribution with the mean (μ), and the standard deviation of gate service times (σ) at this terminal is 2.44 min per truck and 1.18 min per truck, respectively. The time threshold, QT^*, is assumed to be 30 min, meaning that the average truck queuing time does not exceed 30 min. Models (i) and (iii) provide the same outputs, which are

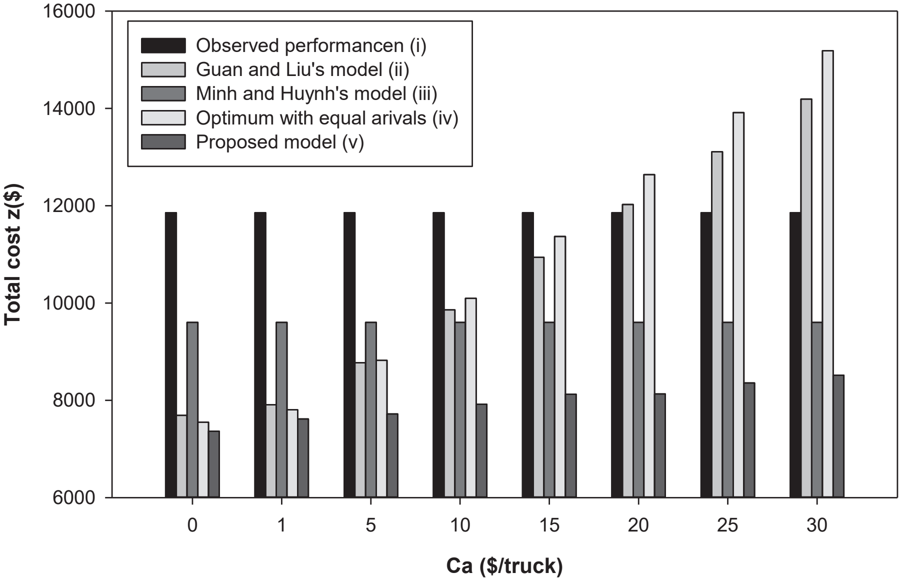

Figure 1. Effect of the cost of shifting truckers' arrival times from their preferred arrival time to another time (C_a) on the total cost in different models

\$11,859 and \$9,601, respectively, regardless of the variety of C_a. It can be explained that no truck is requested for coming to the terminal at another time in those models. The total cost of other models (ii), (iv) and (v) is directly proportional to the value of C_a, meaning that if the cost of shifting truckers' arrival times from their preferred arrival time to another time increases, the total cost of these models (z) increases as well. The results also indicate that disregarding C_a ranging from 0 to \$30 per truck, the total cost estimated from the proposed model is lower than other models considered in this study. The difference in total costs between the proposed model and others is slight when C_a is low and vice versa. In other words, the higher the C_a is, the better the result the proposed model produces when compared with other recent studies.

Similarly, the effect of truck arrival rate on the total cost is examined by assuming truck arrivals to be Poisson distributed with the mean (λ) varied from 0.5 times to 1.5 times of the observed arrival rate at each hour. In this analysis, the truck service time follows the Erlang distribution with the mean (μ), and the standard deviation of gate service times (σ) at this terminal is 2.44 min per truck and 1.18 min per truck, respectively. The time threshold, QT^*, is assumed to be 30 min, and the cost for shifting truck arrival times from their preferred arrival time to another time (C_a) is \$0/truck. Figure 2 represents that increasing the arrival rate with a specified time threshold may increase the total cost accordingly, but it is not always true. For example, in model iv, when the truck arrival rate increases from 1.2 times (1.2x) to 1.3 times (1.3x) of the actual values, the number of operating gates at each hour need to increase from 5 to 6 gates to satisfy the given time threshold ($QT^* = 30$ min). The average waiting time of the truck arrival rate of 1.2x and 1.3x is 9.52 s and 1.39 s, respectively. The total cost (z) of the case of a 1.2x actual arrival rate is, therefore, higher than that of 1.3 times of observed arrival rate, \$13,612, and \$9,359. It has a higher average waiting time than another (1.3x), 9.52 s, and 1.39 s, although it needs fewer service gates (5 gates). A similar explanation for other cases includes 0.5x and 0.6x, 1x, and 1.1x in model iii, 0.7x and 0.8x, 1.4x, and 1.5x in model iv. Figure 2 also indicates that the proposed model's output is the lowest. It means that with changing truck arrival rates at each hour and number of service gates, the truck appointment system proposed in this study performs better than others irrespective of the truck arrival rate's different values.

The gate service times at maritime terminals may fluctuate considerably according to terminals' checking procedures and facilities. The effect of different gate service rates on the total cost is therefore investigated. For this analysis, they are assumed to follow the

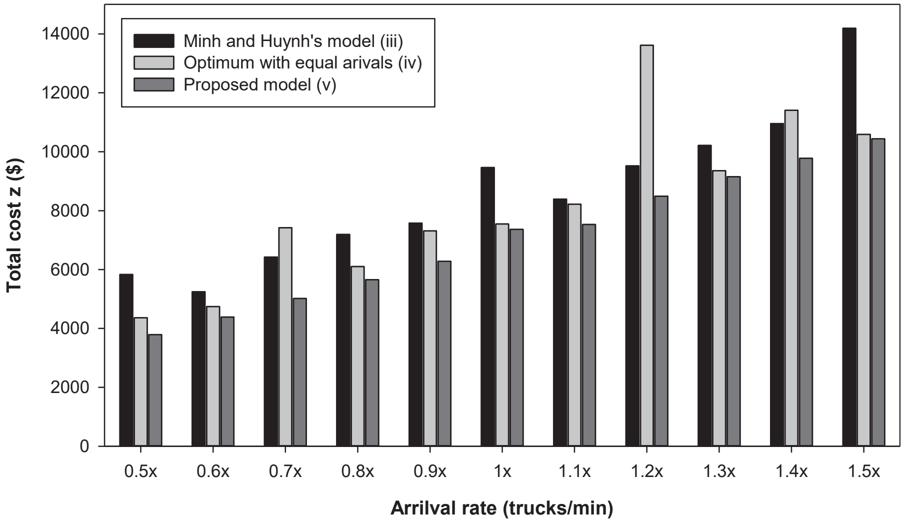

Figure 2.
Effect of truck arrival rate (trucks/min) on the total cost in different models

exponential distribution. The mean value, μ, ranges from 1 to 4 min per truck in 0.25-min increments and the standard deviation, σ, is 1.18 min/truck. Values of truck arrival times, λ, are kept as observed. The time threshold, QT^*, is assumed to be 30 min and the cost for shifting truckers' arrival times from their preferred arrival time to another time (C_a) is again \$0/truck. Figure 3 represents the effect of gate service rate, μ, on different models' total cost.

This figure describes increasing gate service rate may increase the total cost, but it is not always correct. In model iv, if the gate service time increases from 3.5 to 3.75 trucks/min, the number of gates at each hour needs to increase from 6 to 7 gates to satisfy the given time threshold ($QT^* = 30$ min). The average waiting time reduces from 26.07 s to 2.35 s, accordingly. The total cost (z) of the case of gate service rate at 3.5 trucks/min is, therefore, higher than that of gate service time at 3.75 trucks/min, \$23,914, and \$11,091, respectively. It can be explained because it has a higher average waiting time than another, 26.07 s and 2.35 s, even though it needs fewer service gates (6 gates). A similar clarification is also applied for other cases, including the possibility of gate service rate of 1.25 min/truck and 1.5 min/truck, 2.44 min/truck and 2.75 min/truck in model iii, 1.00 min/truck and 1.25 min/truck, 2.25 min/truck and 2.44 min/truck in model iv. Figure 3 also represents that the proposed model's total cost is the lowest among the models considered in this study, meaning that the truck appointment system proposed in this study performs better than others, regardless of the variety of gate service time.

Those models' total cost is computed by considering the effect of the time threshold, QT^*, which is assumed to vary between 1 and 50 min in 5-min increments. In this sensitivity analysis, truck arrival times, λ, are again kept as observed from the case study and follow the Poisson distribution. The gate service time of trucks follows the Erlang distribution with the mean (μ) of 2.44 min per truck and the standard deviation of gate service times (σ) of 1.18 min per truck. The cost of shifting truckers' arrival times from their preferred arrival time to another time (C_a) is \$0/truck. The effect of the time threshold, QT^*, on the total cost of different models is analysed, and the results are in Figure 4 as follows.

This figure describes that if the time threshold increases, the number of gates may reduce, and the total cost does not always decrease. In model iii, for example, if the time threshold increases from 10 to 15 min, the total number of gates operated in a day can be reduced.

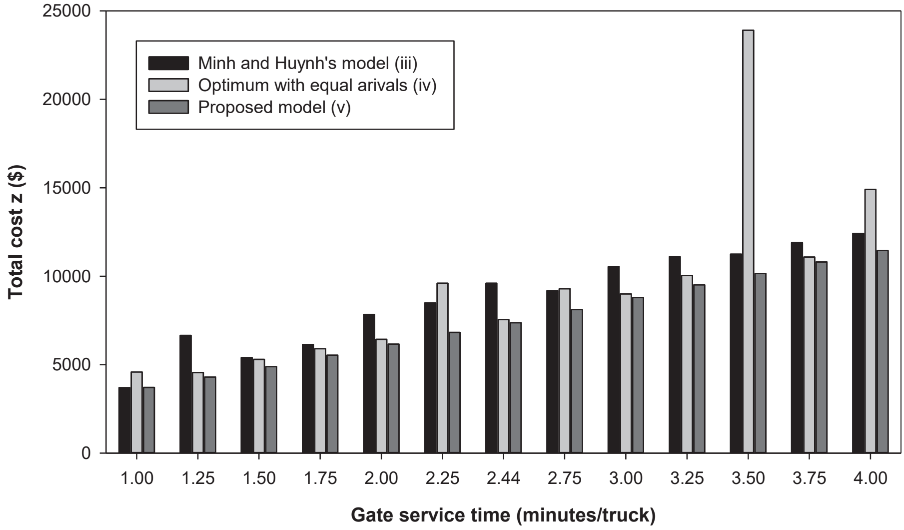

Figure 3.
Effect of gate service time (minutes/truck) on the total cost in different models

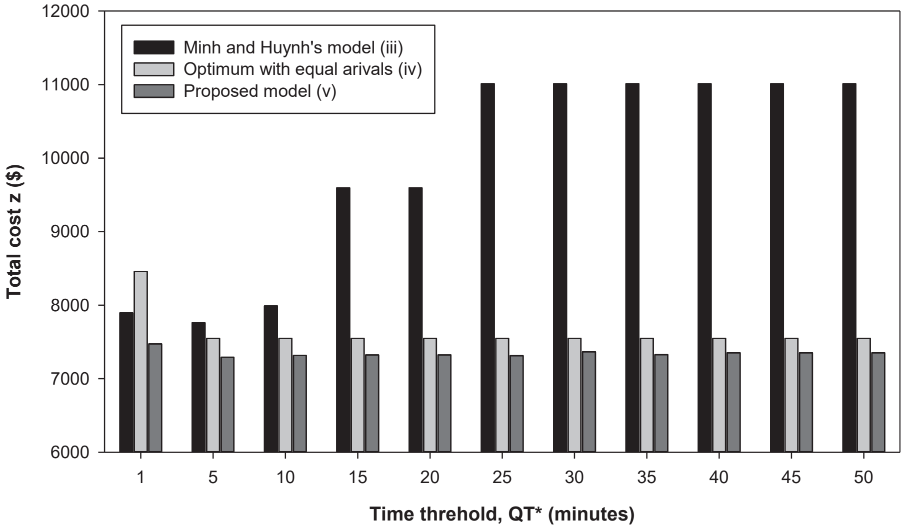

Figure 4.
Effect of the time threshold (minutes) on the total cost in different models

The average waiting time increases accordingly; therefore, the total cost of the 15-min time threshold is higher than that of the 10-min time threshold, $9,596 and $7,989, respectively. In other cases, the total cost can be the same regardless of the variety of the time threshold. This figure also indicates that the proposed model provides the lowest total cost compared to other studies in the literature, disregarding the effect of the time threshold.

More specifically, the effect of the time threshold on the total number of gates with other parameters assuming above displays in Figure 5.

The figure shows that when the time threshold is small, i.e. 1, 5, 10 min, the total number of gates necessary to serve these truck arrivals of the proposed model is the lowest among the

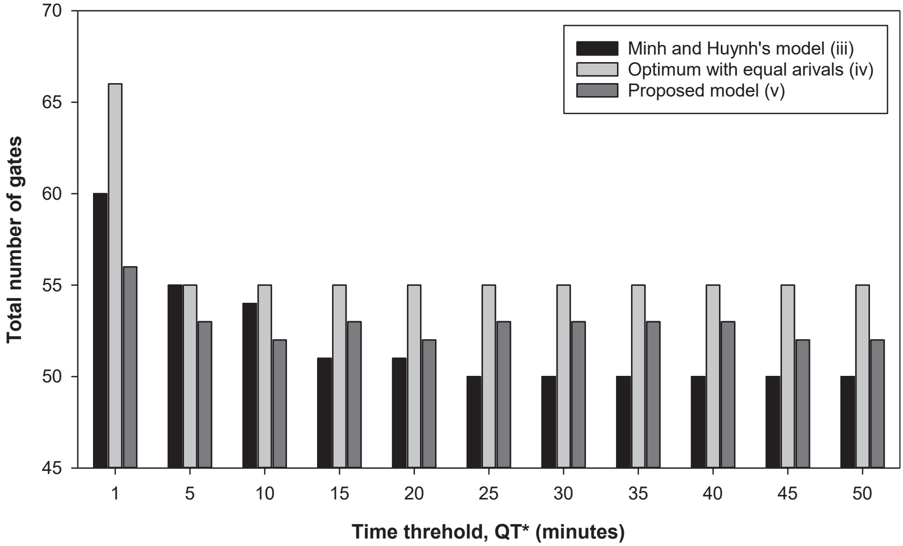

Figure 5.
Effect of the time threshold (minutes) on the total number of gates

three. However, when QT^* is 15 min and higher, the result indicates that this model is not the best compared to model iii. Therefore, the present methodology obtains the minimal cost experiencing at the terminal entrance, not the minimal number of service gates.

Conclusion

This paper applied the truck appointment system concept to minimise the total cost experiencing at the entrance of container terminals by managing truck arrivals at each time interval and estimating the number of operating service gates satisfying a given level of service. The diffusion approximation of $M_t/G/n_t$ queuing model is applied and integrated into a total cost optimisation model. It determines (1) the number of arrival trucks allowed at each time slot and (2) the number of service gates operating at each time slot, satisfying a designated level of service for a given truck arrival rate and truck service rate. The proposed optimisation model is tested with a case study and compared with other models from the literature.

Numerical experiments indicated that the proposed model yields a lower total cost compared to other studies. Specifically, when the cost of shifting truckers' arrival times from their preferred arrival time to another time (C_a) increases, the proposed model's performance is significantly better than that of others. In general, the results indicate that the proposed methodology can offer less than triple and double the cost experiencing at the entrance of container terminals compared to the actual data and existing studies, respectively. However, the present model may not be the best in terms of the total number of service gates necessary to serve these truck arrivals and satisfy the given level of service.

There are several recommendations for further studies though the model is robust and flexible. The application of the diffusion approximation model used in this study may not be precise due to closed-form queuing equations, such as $M/M/n$ and $M/E_k/n$. The model assumed that the truck arrivals follow a Poisson distribution. This assumption may not be correct since the arrival rate at a busy marine terminal is possibly uniformly distributed during peak periods. The developed model does not consider a complicated gate layout that classifies different truck transactions by types, such as export and import trucks, empty and full trucks, full documents handling versus lack of documents managing trucks, etc. Further studies could also be extended and applied to the model to cope with truck arrivals for weeks or even months, not within a day in this study. Furthermore, additional models could modify the proposed model by integrating the management of vessel arrival times.

Data availability statement

Some data, models or code that support the findings of this study are available from the corresponding author upon reasonable request:

1. Code of Truck Appointment Model
2. Data analysis

References

Abdelmagid, A.M., Gheith, M.S. and Eltawil, A.B. (2021), "A comprehensive review of the truck appointment scheduling models and directions for future research", *Transport Reviews*, Vol. 41 No. 4, pp. 1-24.

Bentolila, D.J., Ziedenveber, R.K., Hayuth, Y. and Notteboom, T. (2016), "Off-peak truck deliveries at container terminals: the 'good night' program in Israel", *Maritime Business Review*, Vol. 1 No. 1, pp. 2-20.

Chamchang, P. and Niyomdecha, H. (2021), "Impact of service policies on terminal gate efficiency: a simulation approach", *Cogent Business and Management*, Vol. 8 No. 1, pp. 1-18.

Chen, X., Zhou, X. and List, G. (2011), "Using time-varying tolls to optimise truck arrivals at ports", *Transportation Research Part E*, Vol. 47 No. 6, pp. 965-982.

Chen, G., Kannan, G., Yang, Z., Choi, T. and Jiang, L. (2013a), "Terminal appointment system design by non-stationary M(t)/Ek/c(t) queueing model and genetic algorithm", *International Journal of Production Economics*, Vol. 146 No. 2, pp. 694-703.

Chen, G., Govindan, K. and Yang, Z. (2013b), "Managing truck arrivals with time windows to alleviate gate congestion at container terminals", *International Journal of Production Economics*, Vol. 141 No. 1, pp. 179-188.

Fleming, M., Huynh, N. and Xie, Y. (2013), "An agent-based simulation tool for evaluating pooled queue performance at marine container terminals", *Transportation Research Record: Journal of the Transportation Research Board*, No. 2330, pp. 103-112.

Giuliano, G. and O'Brien, T. (2007), "Reducing port-related truck emissions: the terminal gate appointment system at the ports of Los Angeles and Long Beach", *Transportation Research Part D: Transport and Environment*, Vol. 12 No. 7, pp. 460-473.

Guan, C.Q. and Liu, R. (2009a), "Modeling gate congestion of marine container terminals, truck waiting cost, and optimisation", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2100, pp. 58-67.

Guan, C.Q. and Liu, R. (2009b), "Container terminal gate appointment system optimisation", *Maritime Economics and Logistics*, No. 11, pp. 378-398.

Huynh, N., Smith, D. and Harder, F. (2016), "Truck appointment systems: where we are and where to go from here", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2548, pp. 1-9.

Im, H., Yu, J. and Lee, C. (2021), "Truck appointment system for cooperation between the transport companies and the terminal operator at container terminals", *Applied Sciences*, Vol. 11 No. 1, pp. 1-16.

Jovanovic, R. (2018), "Optimizing truck visits to container terminals with consideration of multiple drays of individual drivers", *Journal of Optimization*, Vol. 2018, pp. 1-8.

Kimura, T. (1983), "Diffusion approximation for an M/G/m queue", *Operation Research*, Vol. 31 No. 2, pp. 304-321.

Minh, C.C. and Huynh, N. (2014), "Planning-level tool for assessing and optimizing gate layout for marine container terminals", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2409, pp. 31-39.

Minh, C.C. and Huynh, N. (2017), "Optimal design of container terminal gate layout", *International Journal of Shipping and Transport Logistics*, Vol. 9 No. 5, pp. 640-650.

Torkjazi, M., Huynh, N. and Shiri, S. (2018), "Truck appointment systems considering impact to drayage truck tours", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 116, pp. 208-228.

Wang, N., Chang, D., Shi, X., Yuan, J. and Gao, Y. (2019), "Analysis and design of typical automated container terminals layout considering carbon emissions", *Sustainability*, Vol. 11 No. 2957, pp. 1-40.

Yao, D.D. (1985), "Refining the diffusion approximation for the M/G/m queue", *Operation Research*, Vol. 33 No. 6, pp. 1266-1277.

Zhang, X., Zeng, Q. and Chen, W. (2013), "Optimisation model for truck appointment in container terminals", *Procedia - Social and Behavioral Sciences*, Vol. 96, pp. 1938-1947.

Corresponding author
Chu Cong Minh can be contacted at: ccminh@hcmut.edu.vn