國立政治大學 應用數學系

碩士 學位論文

A Tiered Security Screening System at Airport

應用於安全檢查之等候模型

碩士班學生：黃鵬錕　撰

指導教授：陸　行　博士

中華民國 105 年 5 月 24 日

Abstract

This thesis proposes a tiered inspection system for airport security, wherein passengers are divided into three classes based on historical security records. A two-dimensional Markov process and a Markov modulated Poisson process (MMPP) queue were used in the formulation of the security inspection system. Simulated annealing was then used to obtain near-optimum solution for the model. The efficacy of the proposed model was evaluated using the arrival data of passengers at Taoyuan International Airport and other two international airports. A comparison with two conventional queueing models with regard to the average waiting time demonstrated the effectiveness of the proposed security inspection system in enhancing service efficiency and boosting the level of security.

中文摘要

本論文中，我們提出基於機場安全檢查的分層排隊理論模型，模型中的旅客基於歷史的安全數據被分成三組。我們運用二維馬可夫過程（two-dimensional Markov process）以及馬可夫調控卜瓦松過程（Markov modulated Poisson process）構建模型的排隊系統並加以分析。我們收集了台灣桃園國際機場和其它兩個機場的旅客數據以驗證我們提出的模型，並運用模擬退火法（simulated annealing）求得近似最佳解（near-optimum solution）。最後我們通過模型的旅客平均等候時間和另外兩種等候模型進行比較，之後得出我們的模型確實可以在不增加成本，甚至提升安全性的同時能夠有效地減少平均等候時間。

Contents

Abstract i

中文摘要 ii

Contents iii

List of Figures v

List of Tables vii

1 Introduction 1

2 Model Formulation and Analysis 4

2.1 Model formulation 4

2.1.1 A single-queue aggregated system without differentiation of passengers 5

2.1.2 A three-independent-lane without sharing system 5

2.1.3 A three-lane with sharing system 6

2.2 Model analysis 10

2.2.1 H-lane and M-lane 10

2.2.2 L-lane 21

2.2.3 Stability conditions 26

3 Simulation Study **29**

3.1 Simulation setup 29

3.2 Simulation assumptions 31

3.2.1 Case A: A single-queue aggregated system 31

3.2.2 Case B: A three-independent-lane without sharing system 32

3.2.3 Case C: A three-lane with sharing system 33

4 Numerical Results **34**

5 Conclusions **47**

A Approximation for GI/G/c Queue **48**

B Simulated Annealing Method **52**

Bibliography **55**

List of Figures

2.1 Tiered security inspection lanes 6

2.2 The state transition diagram for two-dimensional Markov process 11

2.3 Markov modulated Poisson process model 22

2.4 The state transition diagram for $MMPP(2)/M/1$ model 22

4.1 Process of finding the optimal solution with $D = 0.02$ at Taoyuan airport . . . 36

4.2 Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.02$ at Taoyuan airport 36

4.3 The average waiting time E_{W_C} of different D at Taoyuan airport 38

4.4 The thresholds h and m of different D at Taoyuan airport 38

4.5 The proportions p_M and p_L of different D at Taoyuan airport 39

4.6 The buffer size M of different D at Narita airport 41

4.7 Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.2$ at Narita airport 41

4.8 The average waiting time E_{W_C} of different D at Narita airport 42

4.9 The thresholds h and m of different D at Narita airport 42

4.10 The proportions p_M and p_L of different D at Narita airport 43

4.11 Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.2$ at Sydney airport 45

4.12 The average waiting time E_{W_C} of different D at Sydney airport 45

4.13 The thresholds h and m of different D at Sydney airport 46

4.14 The proportions p_M and p_L of different D at Sydney airport 46

List of Tables

3.1 Arrival schedules at Taoyuan airport 30

3.2 Parameters used in simulation at Taoyuan airport 31

4.1 Numerical results of Case A 34

4.2 Numerical results of Case B 34

4.3 The range of parameters used in simulated annealing at Taoyuan airport 35

4.4 Optimal solutions by simulated annealing of Case C with fixed D at Taoyuan airport 35

4.5 Optimal solutions at Taoyuan airport by simulated annealing of Case C with different D at Taoyuan airport 37

4.6 Arrival schedules at Narita airport 39

4.7 Parameters used in simulation at Narita airport 40

4.8 Optimal solutions by simulated annealing with different D at Narita airport . . 40

4.9 Arrival schedules at Sydney airport 43

4.10 Parameters used in simulation at Sydney airport 44

4.11 Optimal solutions by simulated annealing with different D at Sydney airport . . 44

Chapter 1

Introduction

Since its launch in Salt Lack City by the Transportation Security Administration (TSA) in February 2008, the "Black Diamond" self-select program has been expanded to 51 airports. [2] The self-selection process is meant to enable travelers familiar with TSA procedures to pass through checkpoints more quickly and efficiently, while giving families and others with special needs more time and assistance. Self-select lanes use familiar icons based on those used at ski resorts to guide people along trails or lanes in accordance with their skill level. Green designates a queue for families or beginners, blue is for casual travelers at an intermediate level, and the black diamond is reserved for expert travelers who are familiar with TSA rules and arrive at the checkpoint fully prepared. Self-selection also helps to reduce stress and anxiety levels among passengers, and infuse a sense of calm into the checkpoint environment. Reduced stress is a win-win situation for the traveling public as well as officers involved in maintaining transportation security. From the perspective of controlling waiting lines, this program can be seen as the conversion of a security screening system into a tiered service system.

Since September 11, 2001, considerable attention has been directed toward counter-terrorism. The TSA introduced the Computer Assisted Passenger Prescreening System II (CAPPS II) system, which is designed by the Office of National Risk Assessment (ONRA), a subsidiary office of the TSA. [1] On August 2009, TSA announced plans to replace CAPPS II with the Secure Flight program, another airline passenger prescreening program intended to partition passengers into separate classes according to the level of risk they pose. [3] CAPPS II and Security Flight are both risk-based passenger prescreening programs aimed at enhancing security by profiling low- and high-risk passengers prior to their arrival at the airport by matching their names with lists of trusted travelers and watchlists of potentially dangerous individuals. There is growing support for tiered risk-based security system at airports. [4] Again, the passengers are divided into three classes: trusted, regular, and risky. Different screening techniques are then applied to passengers of each class, based on what is known about them. This approach is based on the notion of applying at security checkpoints passenger-related data that the government and the airlines are already collecting. Trusted passengers are individuals who have undergone a background check in order to gain access to an expedited security lane. Risky passengers, as identified by government intelligence systems, are subjected to more intensive scrutiny, using body scanners and interviews with officers trained in behavioral analysis. Regular passengers in the middle group, who are neither vetted nor risky, would receive an intermediate level of screening; however, ideally the process would be made quicker and more efficient than current procedures by prescreening suspicious passengers. The "self-select" three-tier system has already been implemented in many airports; however, the "risk-based" three-tier system is still under consideration.

Whitt [25] discussed when and how to partition arriving passengers into service groups to be served separately. He provided a methodology by which to quantify the trade-off between economics and scale associated with larger systems and the benefits of having passengers requiring shorter service times separated from other passengers requiring longer service times. Poole and Passantino [19] demonstrated that dividing the processing of passengers into multiple levels can be more effective (from a security standpoint) than treating all passengers in the same way. They proposed a risk-based system capable of classifying passengers into two or more risk groups according to the level of risk they pose. When applying risk analysis, data mining, and applied probability to the analysis of prescreening systems, Barnett [5] concluded that Secure Flight program could be transformed from a security centerpiece to one of many components of aviation security systems. There has been tremendous interest in developing airport security screening systems using risk-based profiling methods due to the high false-alarm rates and undue pressure on security officers (see Ryu and Rhee [21]). Jacobson et al. [13] reported that a risk-based approach could greatly enhance security. Cavusoglu et al. [6] claimed that the deployment of a two-screening device architecture by TSA could blunt criticism that profiling is discriminatory while making procedures more convenient for normal passengers and reducing the economic burden of security systems. Nie et al. [17] investigated the means by which to assign passengers to queueing lanes in an airport screening system based on risk, using a steady-state nonlinear binary integer model. Song and Zhuang [23] provided a number of policy insights that are highly useful for security screening practices.

The aim of this study investigated the benefits of implementing the proposed three-tiered security screening system in terms of service quality and security screening. The proposed system was developed using a two-dimensional Markov process in conjunction with a Markov modulated Poisson process. Near-optimum solutions are then obtained using the matrix geometric method with simulated annealing. Actual data is used to evaluate the efficacy of the proposed model.

The remainder of the thesis is structured as follows: Chapter 2 outlines the formulation of a three-lane with sharing system for airport security. We then compare three configurations: (a) a single-queue aggregated system without differentiation of passengers; (b) a three-independent-lane without sharing system, and (c) a three-lane with sharing system. In Chapter 3, we analyze the configurations using simulations and numerical examples. Chapter 4 presents specific computing results and graphics. In Chapter 5, we draw conclusions and propose directions for future research into airport inspection and waiting time models. Details of the methods used in this study can be found in the Appendix.

Chapter 2

Model Formulation and Analysis

2.1 Model formulation

Passengers arriving at an airport security inspection can be classified into multiple classes. This classification can be made based on either risk level assessments or passenger service requirements. Each class of passengers has its own service time distribution and service goal. For a given service capacity such as total number of servers, we examine different service configuration to minimize the waiting time while maintaining the security level. Namely, the goal is to maximize the security inspection effectiveness which can be translated to minimize the probability of "false clear." From the passenger's viewpoint, the goal is to maximize the passenger service quality which can be achieved by minimizing the passenger waiting cost or the probability of "false alarm." Assume that there are three classes of passengers according to the risk level which are labeled as H-, M-, and L- classes. The arrival process of passengers of class-i ($i = H, M, L$) is a Poisson process with rate λ_i. The three arriving processes are independent from each other, and $\lambda_M \geq \lambda_L \geq \lambda_H$. Let $\Lambda = \lambda_H + \lambda_M + \lambda_L$ be the total arrival rate. The service time for class-i is independent and identically distributed (*i.i.d*). Denoted by $B_i(t)$ the service time's cumulative distribution function for class i is with service rate μ_i where $\mu_L \geq \mu_M \geq \mu_H$. In the following sections, we first study a single-queue aggregated system that does not discriminate the passengers in section 2.1.1, and study a three-independent-lane without sharing system in section 2.1.2. Finally, a three-lane with sharing system is developed in section 2.1.3.

2.1.1 A single-queue aggregated system without differentiation of passengers

The basic configuration is to combine arrivals of the three classes into a Poisson process with total arrival rate of Λ and serve these passengers with 3 inspection levels of homogeneous servers. Thus, the average waiting time can be estimated as $M/G/3$ in which the service time's c.d.f, denoted by $B(t)$, is a mixture of c.d.f's. i.e.

$$B\left(t\right)=\frac{\lambda_H}{\Lambda}B_H(t)+\frac{\lambda_M}{\Lambda}B_M(t)+\frac{\lambda_L}{\Lambda}B_L(t).$$

We will adopt the well-known approximations to evaluate the performance of such an $M/G/3$ system. This basic configuration is used as a bench mark for other two security inspection models. In this setting, there is no significantly different inspection procedure for security among all passengers. The specific procedure of approximation method is described in Appendix A.

2.1.2 A three-independent-lane without sharing system

This is a configuration with three dedicated server systems serving three classes. It implies that we may estimate the waiting time by three independent $M/M/1$ systems. The i-class system has a Poisson arrival process with rate λ_i. The service rate is set with respect to each inspection level. In contrast to the single-queue aggregated system, this setting is constructed separately without sharing the resources when passengers are examined at the security inspection.

2.1.3 A three-lane with sharing system

This is a configuration with the features of both the single-queue aggregated system and the three-independent-lane without sharing system. It can also be called a hybrid system. We need to develop the new procedures to compute the performance measures for this configuration. This is a more complex situation. Thus we first assume that the service times are all exponentially distributed and each queue is served by a single server. Due to the security inspection requirement, passengers in the lower class can share the service with the higher class but not the other way around. Such a sharing scheme will ensure that the required security level will not be hurt. Therefore, any improvement in passenger service measures indicates the overall improvement of the system performance. Figure 2.1 below shows the idea described above.

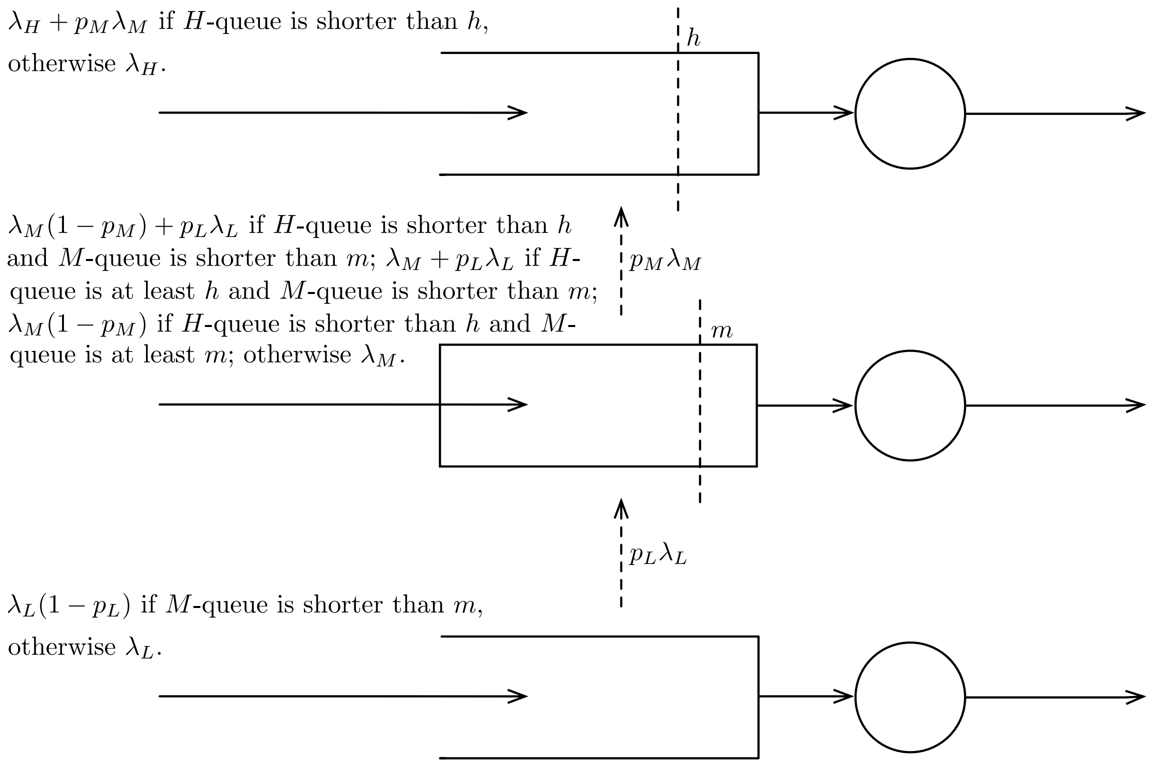


Figure 2.1: Tiered security inspection lanes

The p_M proportion of M-class passengers are sent to H-queue for security inspection, as long as the number of passenger in H-queue is below the threshold h. In the same way, the p_L proportion of L-class passengers are sent to M-queue for security inspection, as long as the number of passenger in M-queue is below its threshold value m. Obviously, this is a more general configuration as $p_M = p_L = 0$, reducing the three-lane with sharing system is to a three-independent-lane without sharing system. In contrast to the previous two configurations in section 2.1.1 and section 2.1.2, this configuration must be examined in detail. We use the queueing model to compute the expected waiting time and the simulation method to determine the optimal configuration of parameters. The computational approach provides a benchmark for numerically evaluating the performance of the system and the simulation approach elucidates the performance effects of the system and configuration of parameters such as M, h, m, p_M, and p_L.

System stability analysis

In this section, we will analyze the stability of the three-lane with sharing system using the method in [26] to show that the proposed model is solvable. Firstly, we define the following events: $A = \{\text{inspection system gives an alarm}\}$; $T = \{\text{the passenger is threat}\}$; $FI = \{$the passenger is selected for further inspection (to H-lane from M-lane)$\}$; FI^c is the complement of FI. Then we set the proportion of further inspection is $P(FI) = p$, which is the proportion of passengers who are sent to H-lane from M-lane, where p can be two: (i) when the number of passengers in H-lane is smaller than h; (ii) when the number of passengers in M-lane is full.

Secondly, we need to define True Alarm (TA), which is the case that the system gives an alarm and a threat exists. Our task now is to prove that the tiered queueing model can indeed improve the security level via maximizing the $P(TA)$, the probability of true alarm, and existence of p as the average waiting time of passengers is reduced. To achieve this goal, we define the following probabilities: the TA rate includes the threat from the further inspection of passengers in M-lane $\theta_{FI}(p) = P(A|T \cap FI)$, and the threat from the passengers in M-land $\theta_{FI^c}(p) = P(A|T \cap FI^c)$. We also need the information that the threat rate of further inspection passengers $\alpha(p) = P(T|FI)$, and threat rate of passengers of inspection in M-lane $\beta(p) = P(T|FI^c)$. Note that these probabilities have been denoted as functions of p, since they all can be controlled by p. Because the inspection procedure in H-lane is far more strict than in

M-lane, we assume $\theta_{FI}(p) > \theta_{FI^c}(p)$. By the law of total probability, we write the probability of true alarm $P(TA)$ as a function of p,

$$P(TA) = f(p) = \theta_{FI}(p)\alpha(p)p + \theta_{FI^c}(p)\beta(p)(1-p).$$

We also need to make the following specific assumptions in developing the model.

Assumption 2.1.1. *By profiling assumption, we know that the true threat from H-lane is greater than that of M-lane, i.e.,*

$$P(A \cap T|FI) > P(A \cap T|FI^c)$$

Assumption 2.1.2. *The sensitivity to change of further inspection is higher than that of its complement, which means*

$$\frac{\mathrm{d}P(A \cap T|FI)}{\mathrm{d}p} > \frac{\mathrm{d}P(A \cap T|FI^c)}{\mathrm{d}p}. \tag{2.1.1}$$

In other words, (2.1.1) *is equal to*

$$|\,(P(A|T \cap FI)P(T|FI))'\,p| > |\,(P(A|T \cap FI^c)P(T|FI^c))'\,(1-p)|,$$

and

$$|\,(\theta_{FI}(p)\alpha(p))'\,p| > |\,(\theta_{FI^c}(p)\beta(p))'\,(1-p)|.$$

Therefore, we get the following proposition of $P(TA)$.

Proposition 2.1.3. $P(TA)$ *is an increasing function of p, that is,*

$$\frac{\partial P(TA)}{\partial p} > 0.$$

Proof:

$$
\begin{aligned}
\frac{\partial P(TA)}{\partial p} =& p\left(\theta'_{FI}(p)\alpha(p)+\theta_{FI}(p)\alpha'(p)\right)+(1-p)\left(\theta'_{FI^c}(p)\beta(p)+\theta_{FI^c}(p)\beta'(p)\right) \\
&+\theta_{FI}(p)\alpha(p)-\theta_{FI^c}(p)\beta(p) \\
=& p\left(\theta_{FI}(p)\alpha(p)\right)'+(1-p)\left(\theta_{FI^c}(p)\beta(p)\right)'+\frac{P(A\cap T\cap FI)}{P(T\cap FI)}\frac{P(T\cap FI)}{P(FI)} \\
&-\frac{P(A\cap T\cap FI^c)}{P(T\cap FI^c)}\frac{P(T\cap FI^c)}{P(FI^c)} \\
=& p\left(\theta_{FI}(p)\alpha(p)\right)'+(1-p)\left(\theta_{FI^c}(p)\beta(p)\right)'+P(A\cap T|FI)-P(A\cap T|FI^c)
\end{aligned}
$$

According to Assumption 2.1.1 and 2.1.2, we know $\frac{\partial P(TA)}{\partial p} > 0$ and complete the proof.

So far, we have shown that $P(TA)$ is an increasing function of p, i.e., with an increase in the number of passengers sent to H-lane from M-lane, the true alarm rate and the safety performance of the proposed model will enhance stability. Herein, we modify the value of p to adjust the requirement of true alarm TA in the model. In addition to enhancing security, it was our intention that the model would decrease the average waiting time. Assume that $E_W(p)$ is the average waiting time for the model, which is obviously a function of p. The purpose of the model can be summarized using the following mathematical programming.

$$
\begin{gathered}
\text{Min } E_W(p) \\
\text{s.t. } D \leq p < 1,
\end{gathered}
$$

where D is a predetermined threshold for p, which is presented as the expected minimum requirement of the true alarm. The use of simulation to obtain the optimal model configuration is outlined in a later chapter.

2.2 Model analysis

The H- and M-class queue can be modeled as a two dimensional Markov process. The L-class queue can be treated as $MMPP(2)/M/1$ queue. We first solve for the stationary joint distribution for the H- and M-queues. Then we approximate the $MMPP(2)/M/1$ L-queue with the superimposition of the two $M/M/1$ queues.

2.2.1 *H*-lane and *M*-lane

Consider a two-dimensional Markov process $\{(X_H(t), X_M(t)), t \geq 0\}$ where $X_i(t)$ represents the number of passengers in queue i at time t. We assume that the stability condition (to be specified later) is satisfied and the steady state is reached. To numerically solve for the stationary distribution of the system, we assume that the buffer size of the passengers with middle risk is set to M.

Next, the state space of the Markov process $\{(X_H(t), X_M(t)), t \geq 0\}$ is

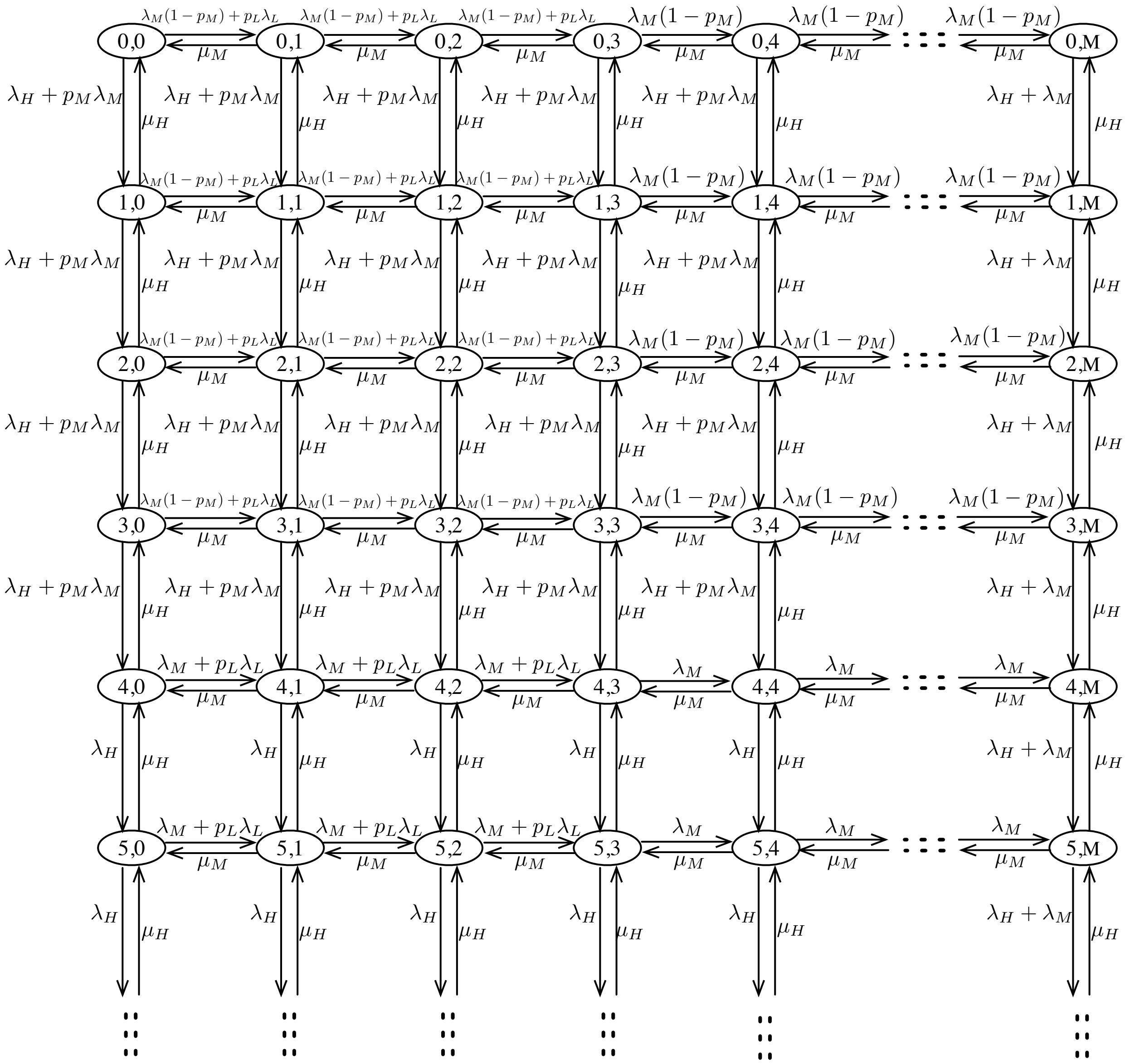
$$\Omega = \{(i,j) | \, i = 0,1,2,\ldots, \; j = 0,1,\ldots, M\}.$$

Let h and m be the thresholds at the H-lane and M-lane, respectively. The state transition diagram for a case with $h = 4$ and $m = 3$ is demonstrated in Figure 2.2 below.

Denoted by $\overline{n}$ is the set of state vectors with common first component n (the number of passengers in H-lane), where $n = 0, 1, 2, \ldots$

$$\overline{n} \; = \; \{(n,0), (n,1), (n,2), \ldots, (n,M)\}.$$

Let (X_H, X_M) be the limit of $(X_H(t), X_M(t))$ as $t \to \infty$. Define the stationary probabilities



$h = 4$, $m = 3$

Figure 2.2: The state transition diagram for two-dimensional Markov process

as follows:

$$
\begin{aligned}
\pi(i,j) &= \lim_{t\to\infty} P\{X_H(t)=i, X_M(t)=j\} = P(X_H=i, X_M=j) \\
\boldsymbol{\pi}_i &= (\pi(i,0), \pi(i,1), \ldots, \pi(i,M)),\ i \geq 0.
\end{aligned}
$$

Because of the high complexity of this queueing system, matrix analytic approach is employed to establish the steady-state equations in matrix form.

The infinitesimal generator $\boldsymbol{Q}$ is given by

$$
\boldsymbol{Q} = \begin{array}{c|ccccccc}
 & \overline{0} & \overline{1} & \cdots & \overline{h-1} & \overline{h} & \overline{h+1} & \cdots \quad \cdots \\
\hline
\overline{0} & \boldsymbol{A}_{00} & \boldsymbol{A}_{01} & & & & & \\
\overline{1} & \boldsymbol{A}_2 & \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & & & & \\
\vdots & & \ddots & \ddots & \ddots & & & \\
\overline{h-1} & & & \boldsymbol{A}_2 & \boldsymbol{A}_{(h-1)(h-1)} & \boldsymbol{A}_{(h-1)h} & & \\
\overline{h} & & & & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \\
\overline{h+1} & & & & & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 \\
\vdots & & & & & & \ddots & \ddots \quad \ddots
\end{array}
$$

Notation

M : the buffer size of M-lane for passengers with middle risk

h : the threshold of queue length associated with passengers with high risk

m : the threshold of queue length associated with passengers with middle risk

p_M : the percentage of passengers in M-lane who are sent to H-lane

p_L : the percentage of passengers in L-lane who are sent to M-lane

λ_H : the average arrival rate of passengers with high risk

μ_H : the average service rate per server for passengers with high risk

λ_M : the average arrival rate of passengers with middle risk

μ_M : the average service rate per server for passengers with middle risk

λ_L : the average arrival rate of passengers with low risk

μ_L : the average service rate per server for passengers with low risk

The state equations are given by $\boldsymbol{\Pi Q} = \boldsymbol{0}$ in which $\boldsymbol{\Pi}$ denotes the steady-state probability vector and $\boldsymbol{0}$ is the zero row vector. The sub-matrices in $\boldsymbol{Q}$ are defined in the following.

For $0 \le i \le h-1$, let $\theta = \lambda_M(1-p_m) + p_L\lambda_L$, then we have

$$\boldsymbol{A}_{ii} = \begin{array}{c|ccccccc} & (i,0) & (i,1) & \cdots & (i,m-1) & (i,m) & \cdots & (i,M) \\ \hline (i,0) & q_0^i & \theta & & & & & \\ (i,1) & \mu_M & q_1^i & \theta & & & & \\ \vdots & & \ddots & \ddots & \ddots & & & \\ (i,m-1) & & & \mu_M & q_{m-1}^i & \theta & & \\ (i,m) & & & & \mu_M & q_m^i & \lambda_M(1-p_m) & \\ \vdots & & & & & \ddots & \ddots & \ddots \\ (i,M) & & & & & & \mu_M & q_M^i \end{array} .$$

Let $\tau = \lambda_H + p_M\lambda_M$, then we have

$$\boldsymbol{A}_{01} = \begin{array}{c|cccc} & (1,0) & \cdots & (1,M-1) & (1,M) \\ \hline (0,0) & \tau & & & \\ \vdots & & \ddots & & \\ (0,M-1) & & & \tau & \\ (0,M) & & & & \lambda_H + \lambda_M \end{array} .$$

Notice that $\boldsymbol{A}_{01} = \boldsymbol{A}_{12} = \boldsymbol{A}_{23} = \cdots = \boldsymbol{A}_{(h-1)h}$.

Considering the repetitive portion $\boldsymbol{A}_2$, $\boldsymbol{A}_1$ and $\boldsymbol{A}_0$, we have $\boldsymbol{A}_2 = \mu_H\boldsymbol{I}$.

For $i \geq h$,

$$
\boldsymbol{A}_1 = \begin{array}{c|ccccccc}
 & (i,0) & (i,1) & \cdots & (i,m-1) & (i,m) & \cdots & (i,M) \\
\hline
(i,0) & q_0^i & \lambda_M + p_L\lambda_L & & & & & \\
(i,1) & \mu_M & q_1^i & \lambda_M + p_L\lambda_L & & & & \\
\vdots & & \ddots & \ddots & \ddots & & & \\
(i,m-1) & & & \mu_M & q_{m-1}^i & \lambda_M + p_L\lambda_L & & \\
(i,m) & & & & \mu_M & q_m^i & \lambda_M & \\
\vdots & & & & & \ddots & \ddots & \ddots \\
(i,M) & & & & & & \mu_M & q_M^i
\end{array}
$$

$$
\boldsymbol{A}_0 = \begin{array}{c|ccc}
 & (i+1,0) & \cdots & (i+1,M-1) & (i+1,M) \\
\hline
(i,0) & \lambda_H & & & \\
\vdots & & \ddots & & \\
(i,M-1) & & & \lambda_H & \\
(i,M) & & & & \lambda_H + \lambda_M
\end{array}
$$

where q_j^i is given as follows:

q_j^i	$i = 0$	$i \geq 1$
$j = 0$	$-\lambda_H - \lambda_M - p_L\lambda_L$	$-\lambda_H - \lambda_M - p_L\lambda_L - \mu_H$
$1 \leq j \leq m-1$	$-\lambda_H - \lambda_M - p_L\lambda_L - \mu_M$	$-\lambda_H - \lambda_M - p_L\lambda_L - \mu_H - \mu_M$
$m \leq j \leq M$	$-\lambda_H - \lambda_M - \mu_M$	$-\lambda_H - \lambda_M - \mu_H - \mu_M$

Before proceeding further, let us define the submatrix of $\boldsymbol{Q}$ in the initial portion as follows

$$\boldsymbol{B}_{00} = \begin{bmatrix} \boldsymbol{A}_{00} & \boldsymbol{A}_{01} & & & \\ \boldsymbol{A}_2 & \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & & \\ & \boldsymbol{A}_2 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} & \\ & & \ddots & \ddots & \ddots \\ & & & \boldsymbol{A}_2 & \boldsymbol{A}_{(h-1)(h-1)} \end{bmatrix}_{h(M+1)\times h(M+1)} \qquad \boldsymbol{B}_{01} = \begin{bmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{A}_{(h-1)h} \end{bmatrix}_{h(M+1)\times(M+1)}$$

$$\boldsymbol{B}_{10} = \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{A}_2 \end{bmatrix}_{(M+1)\times h(M+1)}.$$

Therefore, we can rewrite the infinitesimal generator $\boldsymbol{Q}$

$$\boldsymbol{Q} = \left[\begin{array}{cc|cccc} \boldsymbol{B}_{00} & \boldsymbol{B}_{01} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \boldsymbol{B}_{10} & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \hline \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right]$$

Below we analyze this model in more detail. We now turn to using matrix geometric method to analyze this model. After writing down the sub-matrices explicitly, we can obtain the steady-state probability by a computation recursively. Noting that for the repetitive portion, we have

$$\boldsymbol{\pi}_i\boldsymbol{A}_0 + \boldsymbol{\pi}_{i+1}\boldsymbol{A}_1 + \boldsymbol{\pi}_{i+2}\boldsymbol{A}_2 \;=\; \boldsymbol{0}, \quad i = h, h+1, \cdots. \tag{2.2.1}$$

Therefore, $\boldsymbol{\pi}_i$ is a function only of the transition rates between stationary probabilities with $(i-1)$ queued passengers and stationary probabilities with i queued passengers in H-lane. Then there exists a matrix $\boldsymbol{R}$ such that

$$\begin{aligned}\boldsymbol{\pi}_{i+1} &= \boldsymbol{\pi}_i\boldsymbol{R}, \quad i = h, h+1, \cdots \\ \text{or} \quad \boldsymbol{\pi}_i &= \boldsymbol{\pi}_h\boldsymbol{R}^{i-h}, \quad i = h, h+1, \cdots .\end{aligned} \tag{2.2.2}$$

Substituting (2.2.2) into (2.2.1), we get

$$\boldsymbol{\pi}_h\boldsymbol{R}^{i-h}\boldsymbol{A}_0 + \boldsymbol{\pi}_h\boldsymbol{R}^{i-h+1}\boldsymbol{A}_1 + \boldsymbol{\pi}_h\boldsymbol{R}^{i-h+2}\boldsymbol{A}_2 = \boldsymbol{0}, \quad i = h, h+1, \cdots .$$

Since it is true for $\pi_h\boldsymbol{R}^{i-h} \neq 0$, substituting $i = h$, we have

$$\boldsymbol{A}_0 + \boldsymbol{R}\boldsymbol{A}_1 + \boldsymbol{R}^2\boldsymbol{A}_2 = \boldsymbol{0}. \tag{2.2.3}$$

It is common practice to use the iterative procedure that is derived from the (2.2.3), namely

$$\boldsymbol{R} = -\{\boldsymbol{A}_0 + \boldsymbol{R}^2\boldsymbol{A}_2\}\boldsymbol{A}_1^{-1}.$$

Therefore the recursive solution is given by

$$\begin{aligned}\boldsymbol{R}(0) &= 0, \\ \boldsymbol{R}(k+1) &= -\{\boldsymbol{A}_0 + \boldsymbol{R}^2(k)\boldsymbol{A}_2\}\boldsymbol{A}_1^{-1},\end{aligned}$$

where $\boldsymbol{R}(k)$ is the value of $\boldsymbol{R}$ in the kth iteration. The iteration is repeated until the two successive iterations differs by less than a predefined parameter δ, that is

$$\|\boldsymbol{R}(k+1) - \boldsymbol{R}(k)\|_2 < \delta.$$

For the initial portion, we solve $(\boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_h)$ by using $\boldsymbol{R}$, as we already know,

$$[\boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_{h-1}]\boldsymbol{B}_{00} + \boldsymbol{\pi}_h \boldsymbol{B}_{10} = \boldsymbol{0},$$

$$[\boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_{h-1}]\boldsymbol{B}_{01} + \boldsymbol{\pi}_h \boldsymbol{A}_1 + \boldsymbol{\pi}_{h+1} \boldsymbol{A}_2 = \boldsymbol{0},$$

or

$$\begin{bmatrix} \boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_{h-1} & \boldsymbol{\pi}_h \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{00} & \boldsymbol{B}_{01} \\ \boldsymbol{B}_{10} & \boldsymbol{A}_1 + \boldsymbol{R}\boldsymbol{A}_2 \end{bmatrix} = \boldsymbol{0}. \tag{2.2.4}$$

Expanding the expression

$$\begin{bmatrix} \boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_{h-1} & \boldsymbol{\pi}_h \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{00} \\ \boldsymbol{B}_{10} \end{bmatrix} = \boldsymbol{0},$$

which is

$$\begin{bmatrix} \boldsymbol{\pi}_0 \boldsymbol{\pi}_1 \cdots \boldsymbol{\pi}_{h-1} & \boldsymbol{\pi}_h \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{00} & \boldsymbol{A}_{01} & & & & \\ \boldsymbol{A}_2 & \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & & & \\ & \boldsymbol{A}_2 & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} & & \\ & & \ddots & \ddots & & \ddots \\ & & & \boldsymbol{A}_2 & \boldsymbol{A}_{(h-1)(h-1)} & \\ & & & & \boldsymbol{A}_2 & \end{bmatrix} = \boldsymbol{0}.$$

Then we have the following equations

$$\begin{aligned}
\boldsymbol{\pi}_0 \boldsymbol{A}_{00} + \boldsymbol{\pi}_1 \boldsymbol{A}_2 &= 0 \\
\boldsymbol{\pi}_0 \boldsymbol{A}_{01} + \boldsymbol{\pi}_1 \boldsymbol{A}_{11} + \boldsymbol{\pi}_2 \boldsymbol{A}_2 &= 0 \\
\boldsymbol{\pi}_1 \boldsymbol{A}_{12} + \boldsymbol{\pi}_2 \boldsymbol{A}_{22} + \boldsymbol{\pi}_3 \boldsymbol{A}_2 &= 0 \\
&\vdots \\
\boldsymbol{\pi}_{h-3} \boldsymbol{A}_{(h-3)(h-2)} + \boldsymbol{\pi}_{h-2} \boldsymbol{A}_{(h-2)(h-2)} + \boldsymbol{\pi}_{h-1} \boldsymbol{A}_2 &= 0 \\
\boldsymbol{\pi}_{h-1} \boldsymbol{A}_{(h-1)(h-1)} + \boldsymbol{\pi}_h \boldsymbol{A}_2 &= 0.
\end{aligned}$$

We can get

$$\begin{aligned}
\boldsymbol{\pi}_1 &= -\boldsymbol{\pi}_0 \boldsymbol{A}_{00} \boldsymbol{A}_2^{-1} \\
\boldsymbol{\pi}_2 &= -\left(\boldsymbol{\pi}_0 \boldsymbol{A}_{01} + \boldsymbol{\pi}_1 \boldsymbol{A}_{11}\right) \boldsymbol{A}_2^{-1} \\
\boldsymbol{\pi}_3 &= -\left(\boldsymbol{\pi}_1 \boldsymbol{A}_{12} + \boldsymbol{\pi}_2 \boldsymbol{A}_{22}\right) \boldsymbol{A}_2^{-1} \\
&\vdots \\
\boldsymbol{\pi}_{h-1} &= -\left(\boldsymbol{\pi}_{h-3} \boldsymbol{A}_{(h-3)(h-2)} + \boldsymbol{\pi}_{h-2} \boldsymbol{A}_{(h-2)(h-2)}\right) \boldsymbol{A}_2^{-1} \\
\boldsymbol{\pi}_h &= -\left(\boldsymbol{\pi}_{h-2} \boldsymbol{A}_{(h-2)(h-1)} + \boldsymbol{\pi}_{h-1} \boldsymbol{A}_{(h-1)(h-1)}\right) \boldsymbol{A}_2^{-1}.
\end{aligned}$$

Also, we rewrite the expressions above in the following form:

$$\begin{aligned}
\boldsymbol{\pi}_1 &= \boldsymbol{\pi}_0 \left(-\boldsymbol{A}_{00} \boldsymbol{A}_2^{-1}\right) \\
\boldsymbol{\pi}_2 &= -\left(\boldsymbol{\pi}_0 \boldsymbol{A}_{01} + \boldsymbol{\pi}_1 \boldsymbol{A}_{11}\right) \boldsymbol{A}_2^{-1} \\
&= \boldsymbol{\pi}_0 \left(-\left(\boldsymbol{A}_{01} + \left(-\boldsymbol{A}_{00} \boldsymbol{A}_2^{-1}\right) \boldsymbol{A}_{11}\right) \boldsymbol{A}_2^{-1}\right) \\
\boldsymbol{\pi}_3 &= -\left(\boldsymbol{\pi}_1 \boldsymbol{A}_{12} + \boldsymbol{\pi}_2 \boldsymbol{A}_{22}\right) \boldsymbol{A}_2^{-1} \\
&= \boldsymbol{\pi}_0 \left(-\left(-\boldsymbol{A}_{00} \boldsymbol{A}_2^{-1} \boldsymbol{A}_{12} + \left(-\left(\boldsymbol{A}_{01} + \left(-\boldsymbol{A}_{00} \boldsymbol{A}_2^{-1}\right) \boldsymbol{A}_{11}\right) \boldsymbol{A}_2^{-1}\right) \boldsymbol{A}_{22}\right) \boldsymbol{A}_2^{-1}\right) \\
&\vdots
\end{aligned}$$

Therefore, $(\pi_0 \pi_1 \cdots \pi_h)$ can be expressed as

$$\begin{bmatrix} \pi_0 \pi_1 \cdots \pi_h \end{bmatrix} = \pi_0 X_1,$$

where

$$X_1 = \begin{bmatrix} I & -A_{00}A_2^{-1} & -\left(A_{01} + \left(-A_{00}A_2^{-1}\right) A_{11}\right) A_2^{-1} & \cdots \end{bmatrix}.$$

We also know that

$$1 \ = \ [\pi_0 \pi_1 \cdots \pi_{h-1}]e + \pi_h \sum_{i=1}^{\infty} R^{i-1} e \ = \ [\pi_0 \pi_1 \cdots \pi_{h-1}]e + \pi_h (I - R)^{-1} e,$$

and

$$\sum_{i=1}^{\infty} R^{i-1} \ = \ (I - R)^{-1}.$$

Eventually, from (2.2.4), we get the system of linear equations in the matrix form

$$\pi_0 \begin{bmatrix} e & B_{01}^* \\ (I - R)^{-1} e & (A_1 + RA_2)^* \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix},$$

where B_{01}^* and $(A_1 + RA_2)^*$ are B_{01} and $A_1 + RA_2$ with first column being eliminated, and e is the column vector of 1 with suitable size.

Now, it is apparent that we can get the expected waiting time and the expected number of passengers in the H- and M-lanes. Firstly, the arrival rates at different state conditions of M-lane are given below.

	$0 \le j \le m-1$	$m \le j \le M$
$0 \le i \le h-1$	$\lambda_M(1 - p_M) + p_L \lambda_L$	$\lambda_M(1 - p_M)$
$i \ge h$	$\lambda_M + p_L \lambda_L$	λ_M

Because of limitation of capacity of inspection for passengers with middle risk, the probability of insufficient inspection is estimated by

$$\gamma = \sum_{i=0}^{\infty} \pi(i, M).$$

So the effective arrival rate of passengers with middle risk is

$$\lambda_M (1 - \gamma).$$

We have the average queue length of M-lane

$$S_M = \sum_{j=1}^{M} j \sum_{i=0}^{\infty} \pi(i, j).$$

Also, the average effective arrival rate of passengers in M-lane is given by

$$\begin{aligned}
\lambda_{M_{ave}} =& \{\lambda_M (1-\gamma)(1-p_M) + p_L \lambda_L\} \sum_{i=0}^{h-1} \sum_{j=0}^{m-1} \pi(i,j) + \{\lambda_M (1-\gamma)(1-p_M)\} \sum_{i=0}^{h-1} \sum_{j=m}^{M} \pi(i,j) \\
&+ \{\lambda_M (1-\gamma) + p_L \lambda_L\} \sum_{i=h}^{\infty} \sum_{j=0}^{m-1} \pi(i,j) + \lambda_M (1-\gamma) \sum_{i=h}^{\infty} \sum_{j=m}^{M} \pi(i,j) \\
=& \lambda_M (1-\gamma) - p_M \lambda_M (1-\gamma) \sum_{i=0}^{h-1} \sum_{j=0}^{M} \pi(i,j) + p_L \lambda_L \sum_{i=0}^{\infty} \sum_{j=0}^{m-1} \pi(i,j).
\end{aligned}$$

Thus we can compute the average waiting time of passengers in M-lane

$$W_M = \frac{S_M}{\lambda_{M_{ave}}}.$$

Secondly, the arrival rates at different state conditions of H-lane are

	$0 \le j \le M-1$	$j = M$
$0 \le i \le h-1$	$\lambda_H + p_M \lambda_M$	$\lambda_H + \lambda_M$
$i \ge h$	λ_H	$\lambda_H + \lambda_M$

Then we have the average number of passengers in H-lane

$$S_H = \sum_{i=0}^{\infty} i \sum_{j=0}^{M} \pi(i,j).$$

Also, we can compute the average effective arrival rate of passengers in H-lane as

$$\begin{aligned}\lambda_{H_{ave}} &= \{\lambda_H + p_M \lambda_M\} \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i,j) + \lambda_H \sum_{i=h}^{\infty} \sum_{j=0}^{M-1} \pi(i,j) + (\lambda_H + \lambda_M) \sum_{i=0}^{\infty} \pi(i,M) \\ &= \lambda_H + p_M \lambda_M \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i,j) + \lambda_M \sum_{i=0}^{\infty} \pi(i,M).\end{aligned}$$

Therefore, the average waiting time of passengers in H-lane is given by

$$W_H = \frac{S_H}{\lambda_{H_{ave}}}.$$

2.2.2 *L*-lane

For L-lane, we consider an $MMPP(2)/M/1$ model. This model is a process that behaves like a Poisson process in phase 1 of which the duration is described by a random variable Y_1 with parameter ω_1 for a time that is exponentially distributed when the number of passengers in M-lane increase from $(m-1)$ to m with a mean $\frac{1}{\alpha_{12}}$. Then it switches to a Poisson process in phase 2 having Y_2 with parameter ω_2 for a time period that is exponentially distributed when the number of passengers in M-lane decrease from m to $(m-1)$with a mean $\frac{1}{\alpha_{21}}$. It then switches back to the phase of Y_1 back and force, again and again infinitely often. The Figure 2.3 shown below illustrates the model.

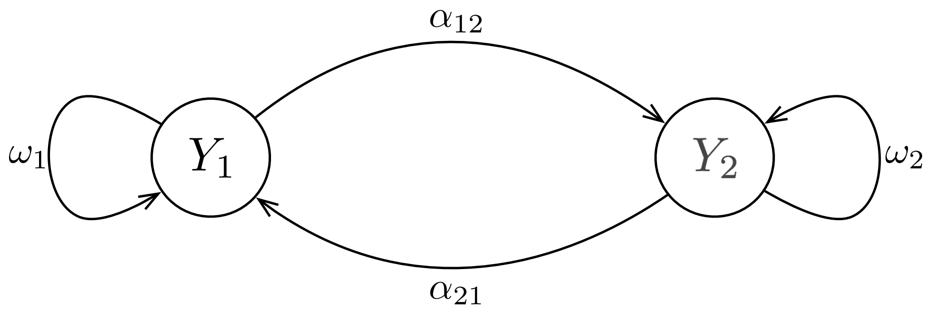


Figure 2.3: Markov modulated Poisson process model

The process of the $MMPP(2)/M/1$ queue is shown in Figure 2.4

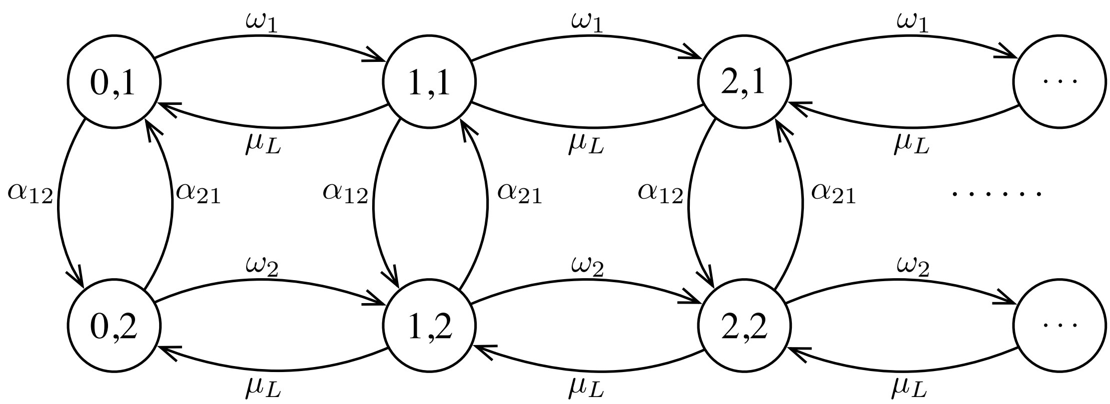


Figure 2.4: The state transition diagram for $MMPP(2)/M/1$ model

To motivate the discussion on the $MMPP(2)/M/1$ queue, we need the following notations. Let $\boldsymbol{\phi_i} = [\phi(i,1), \phi(i,2)]$ denote the vector of the steady-state probabilities that the process is in state with i passengers in L-lane, where $\phi(i,j)$ is the steady-state probability of being in state (i,j), and p_1 and p_2 denote the probabilities that the process is in phase 1 and 2, respectively.

Note that

$$p_j = \sum_{i=0}^{\infty} \phi(i,j) \quad j = 1, 2.$$

From the $MMPP(2)/M/1$ model, we know

$$\begin{aligned}\alpha_{12}p_1 &= \alpha_{21}p_2\\ 1 &= p_1 + p_2.\end{aligned}$$

Thus the average arrival rate of L-lane is given by

$$\lambda_{L_{ave}} = \omega_1 p_1 + \omega_2 p_2 = \omega_1 \frac{\alpha_{21}}{\alpha_{12}+\alpha_{21}} + \omega_2 \frac{\alpha_{12}}{\alpha_{12}+\alpha_{21}}.$$

From the discussion of H-lane and M-lane, we can compute the probabilities p_1 and p_2 as the following equations:

$$\begin{aligned}p_1 &= \sum_{i=0}^{\infty}\sum_{j=0}^{m-1}\pi(i,j)\\ p_2 &= \sum_{i=0}^{\infty}\sum_{j=m}^{M}\pi(i,j) = 1 - p_1.\end{aligned}$$

Changing between the phases only happens when the number of passengers in M-lane change between m and $(m-1)$. In other words, the effect arrival rate of L-lane depends on the number of passengers in M-lane, when it decreases from m to $(m-1)$, the MMPP(2) model switches to phase 1 from phase 2. Similarly, when the number of passengers in M-lane increases from $(m-1)$ to m, the model switches back to phase 2. Because we already know the service rate of M-lane, we can compute the probability from phase 2 to phase 1 in $MMPP(2)$ model, which is

$$\begin{aligned}\alpha_{21} &= P\{X_M = m \big| X_L \text{ is in phase } 2\}\mu_M\\ &= \frac{P\{X_M = m\}\mu_M}{P\{X_L \text{ is in phase } 2\}}\\ &= \frac{\sum_{i=0}^{\infty}\pi(i,m)\mu_M}{p_2}.\end{aligned}$$

We also know that $\alpha_{12} p_1 = \alpha_{21} p_2$, then we have

$$\begin{aligned}\alpha_{12} &= \frac{p_2 \alpha_{21}}{p_1} \\ &= \frac{p_2}{p_1} \frac{\sum_{i=0}^{\infty} \pi(i,m)\mu_M}{p_2} \\ &= \frac{\sum_{i=0}^{\infty} \pi(i,m)\mu_M}{p_1}.\end{aligned}$$

Therefore, we define the parameters of $MMPP(2)/M/1$ model of L-lane,

ω_1	ω_2	α_{12}	α_{21}
$\lambda_L(1-p_L)$	λ_L	$\frac{\sum_{i=0}^{\infty} \pi(i,m)\mu_M}{p_1}$	$\frac{\sum_{i=0}^{\infty} \pi(i,m)\mu_M}{p_2}$

We have the infinitesimal generator

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B} & \boldsymbol{L} & \boldsymbol{F} & \cdots \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

where

$$\boldsymbol{D}_0 = \begin{bmatrix} -(\omega_1+\alpha_{12}) & \alpha_{12} \\ \alpha_{21} & -(\omega_2+\alpha_{21}) \end{bmatrix},\ \boldsymbol{D}_1 = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} \mu_L & 0 \\ 0 & \mu_L \end{bmatrix},\ \boldsymbol{L} = \begin{bmatrix} -(\omega_1+\alpha_{12}+\mu_L) & \alpha_{12} \\ \alpha_{21} & -(\omega_2+\alpha_{21}+\mu_L) \end{bmatrix},\ \boldsymbol{F} = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}.$$

Let $\boldsymbol{\phi} = [\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots]$, because $\boldsymbol{\phi}\boldsymbol{U} = \mathbf{0}$, it produces

$$\begin{cases} \boldsymbol{\phi}_0\boldsymbol{D}_0 + \boldsymbol{\phi}_1\boldsymbol{B} & = \mathbf{0} \\ \boldsymbol{\phi}_0\boldsymbol{D}_1 + \boldsymbol{\phi}_1\boldsymbol{L} + \boldsymbol{\phi}_2\boldsymbol{B} & = \mathbf{0} \\ \boldsymbol{\phi}_{k-1}\boldsymbol{F} + \boldsymbol{\phi}_k\boldsymbol{L} + \boldsymbol{\phi}_{k+1}\boldsymbol{B} & = \mathbf{0} \ \ (k \geq 1). \end{cases}$$

Since $\boldsymbol{\phi}_j$ is a function only of the transition rates between states with $(k-1)$ queued passengers and states with k queued passengers, we have

$$\boldsymbol{\phi}_j = \boldsymbol{\phi}_{j-1}\boldsymbol{T} \ \ j = 2, 3, \cdots$$
$$\text{or } \ \boldsymbol{\phi}_j = \boldsymbol{\phi}_1\boldsymbol{T}^{j-1} \ \ j = 2, 3, \cdots.$$

It means that

$$\begin{cases} \boldsymbol{\phi}_0\boldsymbol{D}_0 + \boldsymbol{\phi}_1\boldsymbol{B} & = 0 \\ \boldsymbol{\phi}_0\boldsymbol{D}_1 + \boldsymbol{\phi}_1(\boldsymbol{L} + \boldsymbol{T}\boldsymbol{B}) & = 0 \end{cases}$$

Or

$$\begin{bmatrix} \boldsymbol{\phi}_0 & \boldsymbol{\phi}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_0 & \boldsymbol{D}_1 \\ \boldsymbol{B} & \boldsymbol{L} + \boldsymbol{T}\boldsymbol{B} \end{bmatrix} = \mathbf{0}.$$

We also need

$$\sum_{k=0}^{\infty} \boldsymbol{\phi}_k\boldsymbol{e} = \boldsymbol{\phi}_0\boldsymbol{e} + \boldsymbol{\phi}_1 \sum_{k=0}^{\infty} \boldsymbol{T}^k\boldsymbol{e} = \{\boldsymbol{\phi}_0 + \boldsymbol{\phi}_1[\boldsymbol{I} - \boldsymbol{R}]^{-1}\}\boldsymbol{e} = 1.$$

The expected number of passengers and the expected waiting time in the L-lane are given by

$$S_L = \sum_{j=1}^{2}\sum_{i=0}^{\infty} i\phi(i,j) = \sum_{j=1}^{2}\sum_{i=0}^{\infty} i\boldsymbol{\phi}_0(\boldsymbol{R}^i)_j$$
$$W_L = \frac{S_L}{\lambda_{L_{ave}}}$$

where $(\boldsymbol{R}^i)_j$ is the j^{th} column of the matrix $\boldsymbol{R}^i$.

2.2.3 Stability conditions

Before approaching simulation study, we first pose the problem of finding what is the stability conditions in the model. We use the following two lemmas to give these conditions. Then we will rewrite the mathematical programming of the proposed model to find the optimal solution.

Lemma 2.2.1. *To reach a stable state, we must make sure that the following inequalities are satisfied.*

$$\lambda_{H_{ave}} < \mu_H$$
$$\lambda_{L_{ave}} < \mu_L.$$

In other words, the inequalities can be simplified as

$$0 \le p_M < \frac{\mu_H - \lambda_H - \lambda_M \sum_{i=0}^{\infty} \pi(i, M)}{\lambda_M \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i, j)}$$
$$\frac{\lambda_L - \mu_L}{\lambda_L p_1} < p_L \le 1$$

where $p_1 = \sum_{i=0}^{m-1} \sum_{j=0}^{H} \pi(i, j)$ as defined before.

Proof:

From definition of the average arrival rate in H-lane and L-lane, it follows that

$$\begin{aligned}
\lambda_{H_{ave}} &= \{\lambda_H + p_M\lambda_M\} \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i, j) + \lambda_H \sum_{i=h}^{\infty} \sum_{j=0}^{M-1} \pi(i, j) + (\lambda_H + \lambda_M) \sum_{i=0}^{\infty} \pi(i, M) \\
&= \lambda_H + p_M\lambda_M \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i, j) + \lambda_M \sum_{i=0}^{\infty} \pi(i, M).
\end{aligned}$$

Since $\lambda_{H_{ave}} < \mu_H$, simplifying the expression, we get

$$\lambda_H + p_M\lambda_M \sum_{i=0}^{h-1}\sum_{j=0}^{M-1} \pi(i,j) + \lambda_M \sum_{i=0}^{\infty} \pi(i,M) < \mu_H.$$

Therefore, we have the following inequality

$$p_M < \frac{\mu_H - \lambda_H - \lambda_M \sum_{i=0}^{\infty} \pi(i,M)}{\lambda_M \sum_{i=0}^{h-1} \sum_{j=0}^{M-1} \pi(i,j)}.$$

By the same method, as $\lambda_{L_{ave}} < \mu_L$, we know

$$\lambda_{L_{ave}} = \lambda_L(1-p_L)p_1 + \lambda_L p_2,$$

and

$$p_1 + p_2 = 0.$$

Simplify the equation, we get

$$p_L > \frac{\lambda_L - \mu_L}{\lambda_L p_1}.$$

p_M and p_L are proportions, so this completes the proof of Lemma 2.2.1.

Next we will discuss the range of p by Lemma 2.2.2. First, we have

$$p = \sum_{i=0}^{h-1}\sum_{j=0}^{M-1} \pi(i,j) + \sum_{i=0}^{\infty} \pi(i,M). \tag{2.2.5}$$

Lemma 2.2.2. *To reach a stable state, the proportion p sent from M-lane to H-lane need to satisfy the inequality below*

$$0 < p < \frac{\mu_H - \lambda_H - (1-p_M)\lambda_M \sum_{i=0}^{\infty} \pi(i,M)}{p_M\lambda_M}.$$

The proof of this result is quite similar to that given for Lemma 2.2.1 and so is omitted. After we have Lemma 2.2.2, let D be the pre-given minimum requirement to adjust true alarm rate of the model. As a consequence, we rewrite the mathematical programming in section 2.1.3 as follows:

$$\begin{aligned}
&\text{Min } E_W(p) \\
&\text{s.t. } p = \sum_{i=0}^{h-1}\sum_{j=0}^{M-1}\pi(i,j) + \sum_{i=0}^{\infty}\pi(i,M) \\
&\qquad p \geq D \\
&\qquad 0 \leq p < \frac{\mu_H - \lambda_H - (1-p_M)\lambda_M \sum_{i=0}^{\infty}\pi(i,M)}{p_M\lambda_M} \\
&\qquad 0 \leq p_M < \frac{\mu_H - \lambda_H - \lambda_M \sum_{i=0}^{\infty}\pi(i,M)}{\lambda_M \sum_{i=0}^{h-1}\sum_{j=0}^{M-1}\pi(i,j)} \\
&\qquad \frac{\lambda_L - \mu_L}{\lambda_L p_1} < p_L \leq 1
\end{aligned} \tag{2.2.6}$$

Chapter 3

Simulation Study

Simulations were conducted using the actual data pertaining to Taiwan Taoyuan International Airport and other two international airports in order to determine the extent to which the tiered security system enhances airport security and reduces the expected waiting time.

The simulation results demonstrate that the proposed tiered security system is able to achieve both of these goals in (2.2.6).

3.1 Simulation setup

In the following, we consider three cases: (A) a single-queue aggregated system without differentiation of different type of passengers; (B) a three-independent-lane without sharing system; (C) a three-lane with sharing system. The purpose of simulation is to evaluate the efficacy of the model by comparing simulation results with data collected from three international airports. Finally, the simulation results are analyzed to find optimal queueing policy of the model for Case C.

Assume that when passengers join the security system, historical records of the passengers are used to categorized them H-, M-, L-classes in the formulation of the model.

Data included the scheduled times of departure on the website of Taiwan Taoyuan International Airport(http://www.taoyuan-airport.com/english/flight_depart/) for February 10th, 2015. The following fixed key parameters were used in the study: (1) arrival rate of passengers; (2) service rate of passengers.

In the Taoyuan International Airport website, we collect the scheduled fight between 10:30 and 14:30 in Terminal 1. According to the aircraft type of the flight, we can estimate the total number of passengers arriving the airport. Table 3.1 gives information about the arrival process.

Table 3.1: Arrival schedules at Taoyuan airport

Time Period	Number of Flights	Number of Passengers
10:30 - 11:30	9	1675
11:30 - 12:30	4	1057
12:30 - 13:30	8	2100
13:30 - 14:30	9	2255

The average arrival rate of passengers for the period 10:30-14:30 was 1771.75 per hour. We assume that 60% of them are assigned to M-class and 10% of them are assigned to H-class. On contrary, 30% of arrival passengers are allocated to L-class. We divide it into three classes according to this proportion.

It was also assumed that security lanes in different security level have the different service rate. We assume that the average service rates of each corresponding lane in H-lane, M-lane and L-lane are 185, 220 and 270 per hour respectively. Thus, the total average service rates of H-lane, M-lane and L-lane are 185, 1100 and 540 per hour respectively. Specific data is listed in Table 3.2.

In addition to the data collected in Taoyuan International Airport, we also have collected the scheduled times of departure on the website of Narita International Airport(http://www.narita-airport.or.jp/ais/flight/today/e_inter_dep.html) for July 4th, 2015 and on the website of Sydney Airport(http://www.sydneyairport.com.au/flights/flight-arrivals-and-departures/international-departures.aspx) for July 13th, 2015. Also, we use the data to validate the proposed model as well. The detailed data and numerical results will be presented in the next Chapter.

Table 3.2: Parameters used in simulation at Taoyuan airport

Classes	Proportion	Arrival Rate	Total Service Rate	Service rate per lane $\times$ Number of lanes
H-class	10%	177.175	185	185×1
M-class	60%	1063.05	1100	220×5
L-class	30%	531.525	540	270×2

3.2 Simulation assumptions

Investigation of the queueing system, requires a number of assumptions based on findings in the existing literature. The simulations do not account for flight delays, mechanical problems, balking, reneging, or the effects of dealing with families or groups.

3.2.1 Case A: A single-queue aggregated system

A single-queue aggregated system is an $M/G/3$ queueing system. Passengers arrive at the airport according to a Poisson process with mean arrival rate Λ and service time's cdf $B(t)$ is a mixture of cdf's. According to Choi et al. [7], the $M/G/3$ queueing model can be approximated by $GI/G/c/c+r$, where $c = 3$ and r is the length of the queue, which can be set sufficient large. The specific method of the approximation is outlined in Appendix A.

In the study, arrivals occur at rate Λ according to a Poisson process and service time has a hyper-exponential distribution. Then we obtain $c_A^2 = 1$ and

$$c_S^2 = \frac{Var[S]}{(E[S])^2},$$

where

$$
\begin{aligned}
E[S] &= \sum_{i=1}^{3} \frac{p_i}{\mu_i} = \frac{\lambda_H}{\Lambda}\frac{1}{\mu_H} + \frac{\lambda_M}{\Lambda}\frac{1}{\mu_M} + \frac{\lambda_L}{\Lambda}\frac{1}{\mu_L} \\
E[S^2] &= \sum_{i=1}^{3} \frac{2}{\mu_i^2} p_i = \frac{\lambda_H}{\Lambda}\frac{2}{\mu_H^2} + \frac{\lambda_M}{\Lambda}\frac{2}{\mu_M^2} + \frac{\lambda_L}{\Lambda}\frac{2}{\mu_L^2} \\
Var[S] &= E[S^2] - E[S]^2 \\
&= \left[\sum_{i=1}^{3} \frac{p_i}{\mu_i}\right]^2 + \sum_{i=1}^{3}\sum_{j=1}^{3} p_i p_j \left(\frac{1}{\mu_i} - \frac{1}{\mu_j}\right)^2 \\
&= \left(\frac{\lambda_H}{\Lambda}\frac{1}{\mu_H}\right)^2 + \left(\frac{\lambda_M}{\Lambda}\frac{1}{\mu_M}\right)^2 + \left(\frac{\lambda_L}{\Lambda}\frac{1}{\mu_L}\right)^2 + 2\frac{\lambda_H}{\Lambda}\frac{\lambda_M}{\Lambda}\left(\frac{1}{\mu_H} - \frac{1}{\mu_M}\right)^2 \\
&\quad + 2\frac{\lambda_H}{\Lambda}\frac{\lambda_L}{\Lambda}\left(\frac{1}{\mu_H} - \frac{1}{\mu_L}\right)^2 + 2\frac{\lambda_M}{\Lambda}\frac{\lambda_L}{\Lambda}\left(\frac{1}{\mu_M} - \frac{1}{\mu_L}\right)^2.
\end{aligned}
$$

Therefore, the estimated value of the average waiting time in Case A is given by the following

$$
E_{W_A} = \frac{1}{\Lambda}\sum_{n=1}^{\infty} i\tilde{P}_n.
$$

3.2.2 Case B: A three-independent-lane without sharing system

As described in Chapter 2, we designate a three-independent-lane without sharing system is a three independent $M/M/1$ system. This simulation includes a novel measure for the estimation the average waiting time of Case B, which is given as follows:

$$
E_{W_B} = \frac{S_H + S_M + S_L}{\Lambda},
$$

where S_i is the estimated average queue length of the $M/M/1$ queue for H, M, L classes, respectively.

E_{W_B} is set up to enable a comparison of the overall differences between Case B and Case C. This type of comparison provides a more direct indication of the performance of the proposed model as the three lanes are considered as a whole. Using this kind of objective function also makes it possible to determine the average waiting time for each lane in the numerical results. Keeping E_{W_B} as an average upper bound of Case C, the same method is used to assess the model performance in Case C.

3.2.3 Case C: A three-lane with sharing system

An optimization model provides an alternative means by which to obtain an optimal solution to the proposed tiered model. In this thesis, we employed simulated annealing to find the optimal solution for Case C. This begins with the establishment of an objective function of for the simulated annealing model, as follows:

$$E_{W_C} = \frac{S_H + S_M + S_L}{\Lambda}.$$

Simulated annealing method is based on the pre-given domain of these controllable parameters to determine the range to find the optimal feasible solution. Begin with initial point x_0, simulated annealing algorithm looks for the best neighbor point of x_0 and make it be x_k. After found every possible solution x_{k+1}, the algorithm will make sure x_{k+1} satisfy the conditions in mathematical programming, only coincident solution will be outputted. Otherwise it will use some specific approaches to deal with, which include the physics concepts of annealing, more material of simulated annealing method may be found in Appendix B.

Chapter 4

Numerical Results

In this chapter, we outline the numerical results of the three cases. First of all, there are extensive discussions on these cases by using the date collected in Taoyuan International Airport. We will get some useful conclusions of the proposed model. In the final part of this Chapter, we will use the data collected in the other two international airports to strengthen our conclusions.

To begin with Case A, Table 4.1 presents the parameters computed using the proposed estimation method.

Table 4.1: Numerical results of Case A

a	b	a_R	b_R	c_S^2	E_{W_A}
$\frac{1}{1771.75}$	0.0016	$\frac{1}{1771.75}$	0.0022	1.6336	1.434

Thus, we obtain the numerical results of Case B, as shown in Table 4.2.

Table 4.2: Numerical results of Case B

S_H	S_M	S_L	W_H	W_M	W_L	E_{W_B}
22.64	28.77	62.72	7.668	1.626	7.080	3.86

It should be noted that Case C achieves an average waiting time shorter than that of Case B, as anticipated. In the following, we present the optimal numerical results obtained via simulated annealing with a given fixed minimum requirement D. Table 4.3 below presents the range of controllable parameters used in simulated annealing.

Table 4.3: The range of parameters used in simulated annealing at Taoyuan airport

M	h	m	p_M	p_L
$[70, 100]$	$[1, 20]$	$[1, 20]$	$[0, 1]$	$[0, 1]$

Then, the optimal solutions can be found in Table 4.4, we assume $D = 0.02$ in this part, and the process of finding the optimal solutions is presented in Figure 4.1, the upper part of this figure shows that the changing of feasible solutions varies with iterations, the lower part demonstrates the changing of optimal solution E_{W_C}. In the search for an optimal solution, we also record the changes in proportion p of all passengers transferred from M-lane to H-lane. Figure 4.2 presents a comparison between p and the current feasible solution E_{W_C}.

Table 4.4: Optimal solutions by simulated annealing of Case C with fixed D at Taoyuan airport

D	M	h	m	p_M	p_L	E_{W_C}	W_H	W_M	W_L
0.02	100	2	4	0.28	0.99	2.359	10.419	1.403	1.474

As indicated above, the relationship between E_{W_C} and p is unpredictable rather than linear. Nonetheless, the simulation has an optimal solution, which can be found via simulated annealing. Compared to Case B, Case C shortens the average waiting time in L-lane and M-lane but extends the waiting time in H-lane, resulting in an overall decrease. A lack of passengers in higher level lanes will tend to attract passengers from lower level lanes. However, it is necessary to ensure that p is not less than D at any point during this period. Thus, in the following, we discuss the changes to the configuration of parameters and the average waiting time with different values for D.

Table 4.5 lists the optimal parameter configurations with various values for D. An increase in D was shown to increase E_{W_C}; however, the effect is not significant. Figure 4.3 can reflect

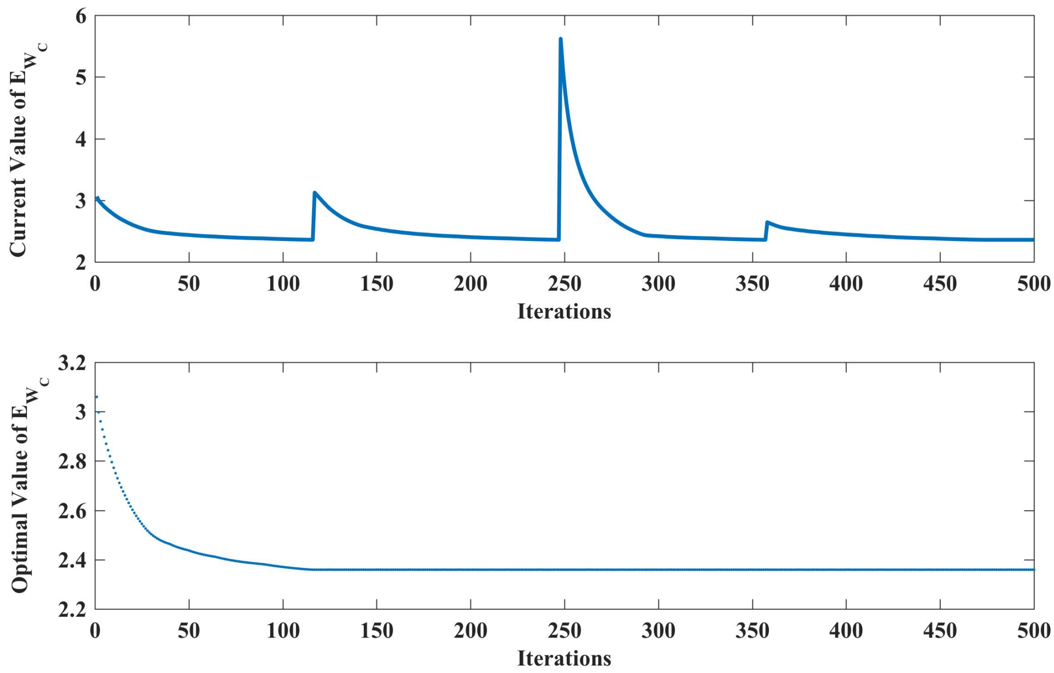


Figure 4.1: Process of finding the optimal solution with $D = 0.02$ at Taoyuan airport

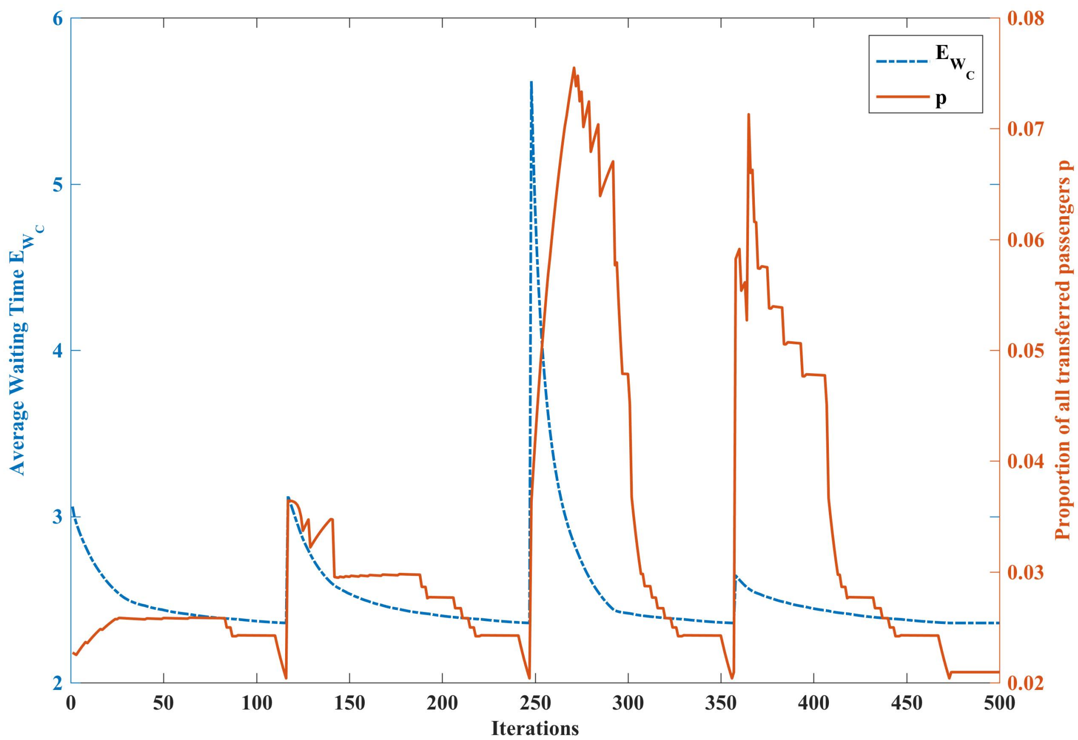


Figure 4.2: Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.02$ at Taoyuan airport

this trend more directly. The thresholds h and m can be adjusted in accordance with the value for D. As shown in Figure 4.4, an increase in D makes the requirements for the security inspection system more stringent, which leads to an increase in h, while the value of m remains unchanged.

Table 4.5: Optimal solutions at Taoyuan airport by simulated annealing of Case C with different D at Taoyuan airport

D	M	h	m	p_M	p_L	E_{W_C}
0	100	1	4	1	1	2.339
0.01	100	1	4	0.53	0.99	2.351
0.02	100	2	4	0.28	0.99	2.359
0.03	100	2	4	0.15	1	2.370
0.04	100	3	4	0.13	0.99	2.379
0.05	100	3	4	0.09	0.99	2.384
0.06	100	3	4	0.06	0.99	2.394
0.07	100	4	4	0.06	0.99	2.399
0.08	100	4	4	0.04	1	2.406
0.09	100	5	4	0.05	1	2.410
0.10	100	5	4	0.04	1	2.412

In the next part, we discuss the issue of proportions p_M and P_L, which send passengers to a higher level lane from lower lane. Figure 4.5 presents the results of this situation. The result is somewhat non-intuitive: an increase in D results in the transfer of a smaller proportion of arriving passengers from M-lane to H-lane, such that the value of p_M drops steadily, as shown in Figure 4.5. Nonetheless, we know that the threshold h gradually increases, which drives up the overall number of passengers being transferred. At the same time, the threshold of M-lane remains constant; therefore, the value of p_L remains equal to 1.

Following the statements above, we will change some of the model configurations to check that the conclusions are still holds. Table 4.6 shows the arrival schedules at Narita Airport which are collected from 10:30 to 14:30. So we have the average total arrival rate during the period at Narita Airport is 1770.5 per hour, which is very close the data we collected at Taoyuan airport. But we use different proportions to assign the number of passengers to the three lanes. In this time, we assume that 70% of arrival passengers, in which 65% of them are assigned to M-lane and 5% of them are assigned to H-lane. Thus 30% of arrival passengers are allocated to L-lane.

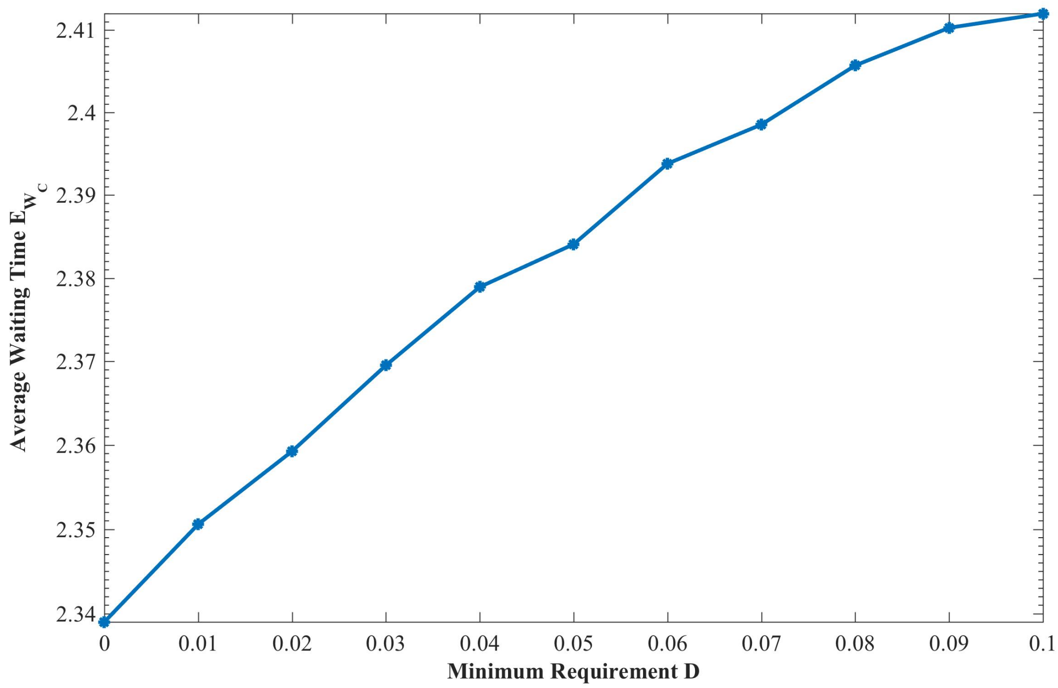


Figure 4.3: The average waiting time E_{W_C} of different D at Taoyuan airport

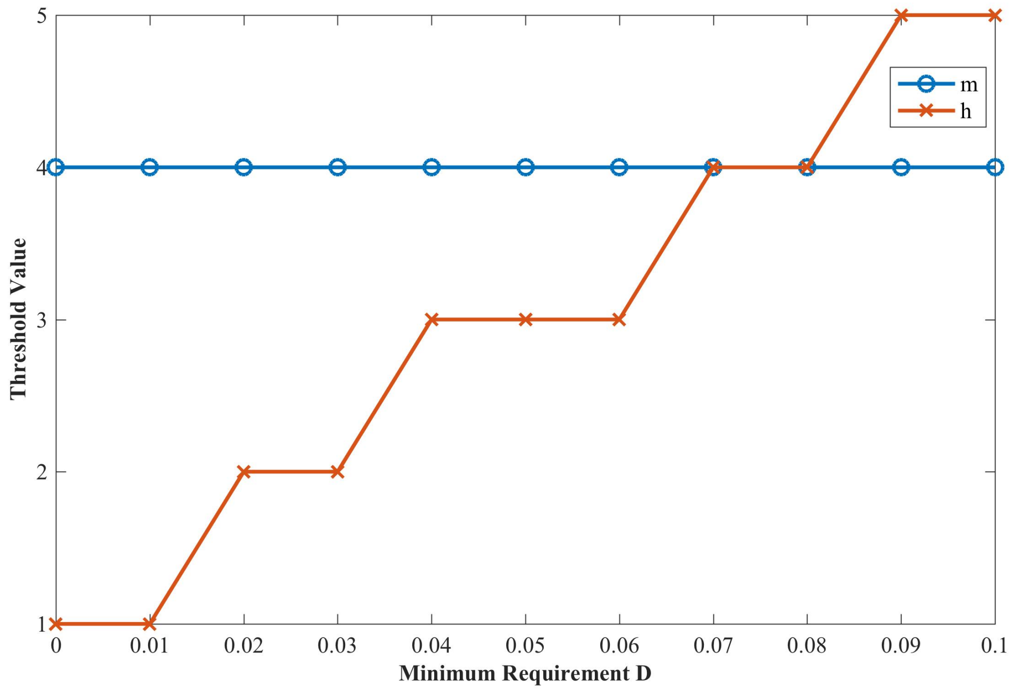


Figure 4.4: The thresholds h and m of different D at Taoyuan airport

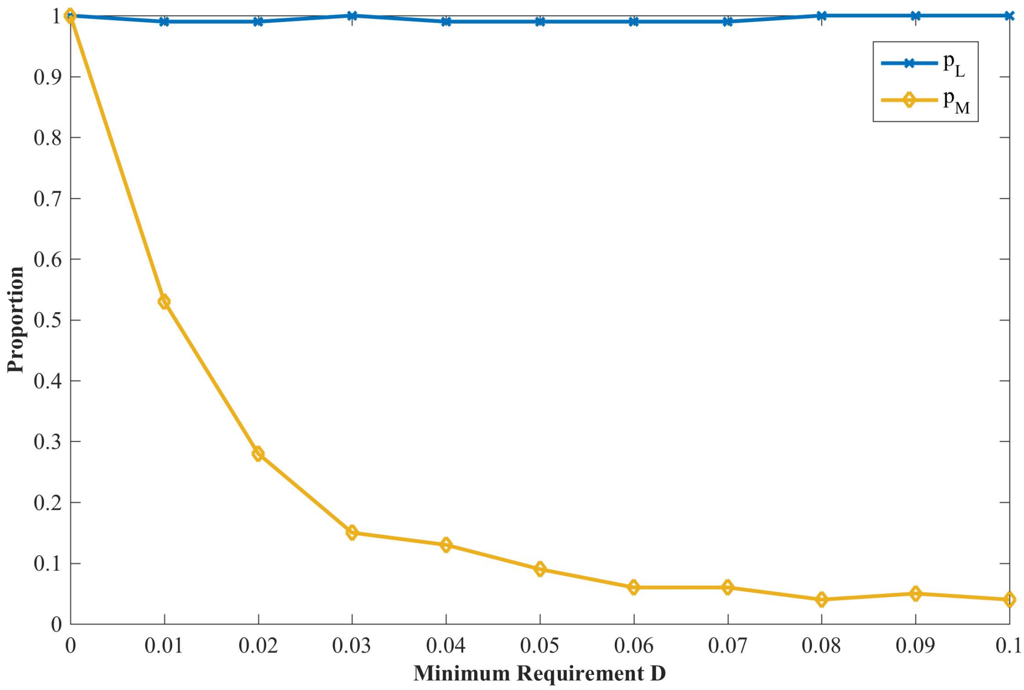


Figure 4.5: The proportions p_M and p_L of different D at Taoyuan airport

Since the number of arrival passenger is close to that at Taoyuan airport, we use the same service rates of three lanes. Table 4.7 demonstrates this assignment. The range of parameters we used in simulated annealing method is still the same as we did at Taoyuan airport.

Table 4.6: Arrival schedules at Narita airport

Time Period	Number of Flights	Number of Passengers
10:30 - 11:30	16	4075
11:30 - 12:30	5	986
12:30 - 13:30	3	630
13:30 - 14:30	6	1391

We directly obtain the optimal solutions via simulated annealing method with different D in the Table 4.8. In this configuration, we can still get the similar conclusions. Nevertheless, the only difference is that the optimal configurations of M at Taoyuan airport occur at the upper bound of its range, while at Narita airport, the optimal configurations of M become varied with the increase of the minimum requirement D but there is no strict rules. It can be intuitively

Table 4.7: Parameters used in simulation at Narita airport

Classes	Proportion	Arrival Rate	Total Service Rate
H-class	5%	88.525	185
M-class	65%	1150.825	1100
L-class	30%	531.15	540

observed from Figure 4.6. As the same methods we dealing with at Taoyuan airport, the comparison between proportion p and current average waiting time E_{W_C} at Narita airport is given in Figure 4.7, the average waiting time E_{W_C} of different D is given in Figure 4.8, and the thresholds h and m of different D are given in Figure 4.9. Also, Figure 4.10 demonstrates the proportions p_M and p_L of different D.

Table 4.8: Optimal solutions by simulated annealing with different D at Narita airport

D	M	h	m	p_M	p_L	E_{W_C}
0	70	2	4	0.99	0.99	1.516
0.05	70	2	4	0.99	0.99	1.516
0.10	75	3	4	0.82	0.99	1.522
0.15	71	3	4	0.52	0.99	1.532
0.20	75	4	4	0.39	1	1.549
0.25	70	4	4	0.30	1	1.560
0.30	70	5	4	0.25	0.99	1.577
0.35	70	5	4	0.21	1	1.590
0.40	79	6	4	0.19	0.99	1.606
0.45	90	6	4	0.17	1	1.625
0.50	82	7	4	0.15	0.99	1.641

Since the average arrival rates collected at Taoyuan airport and Narita airport are almost equal, we have collected data from different time periods at Sydney International Airport. Table 4.9 gives the arrival schedules.

From Table 4.9, the average arrival rate at Sydney airport during 18:00-22:00 is 1140.5 per hour, which is quite smaller than that at Taoyuan airport. We allocate the passengers according to the same proportions at Taoyuan airport. Therefore, 60% of them and 10% of them are assigned to M-lane and H-lane. On contrary, 30% of arrival passengers are allocated to L-lane. Similarly,

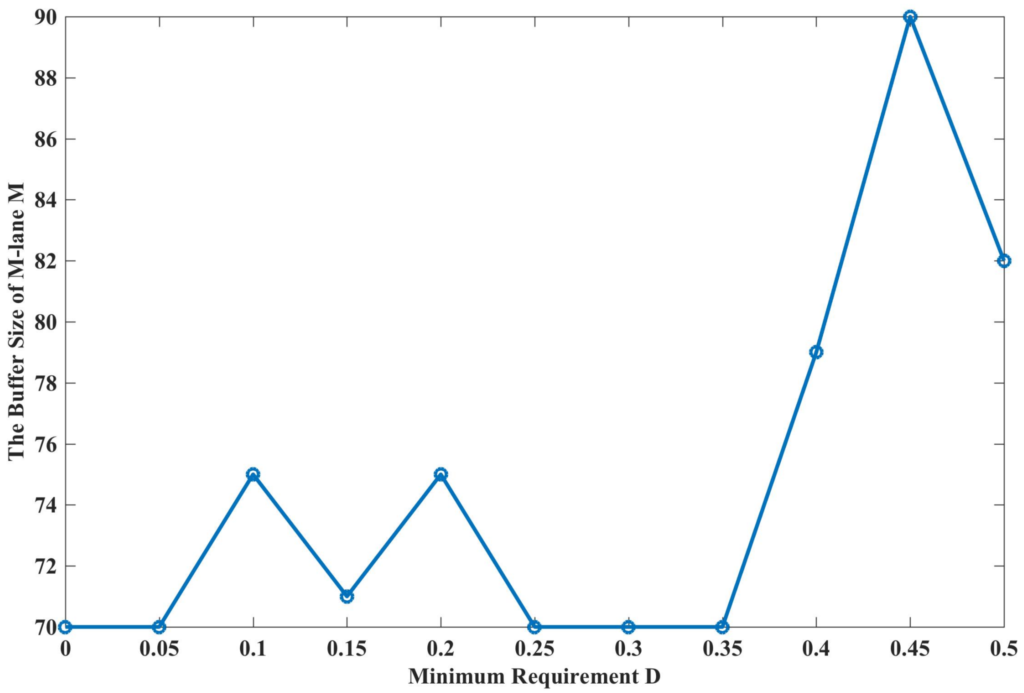


Figure 4.6: The buffer size M of different D at Narita airport

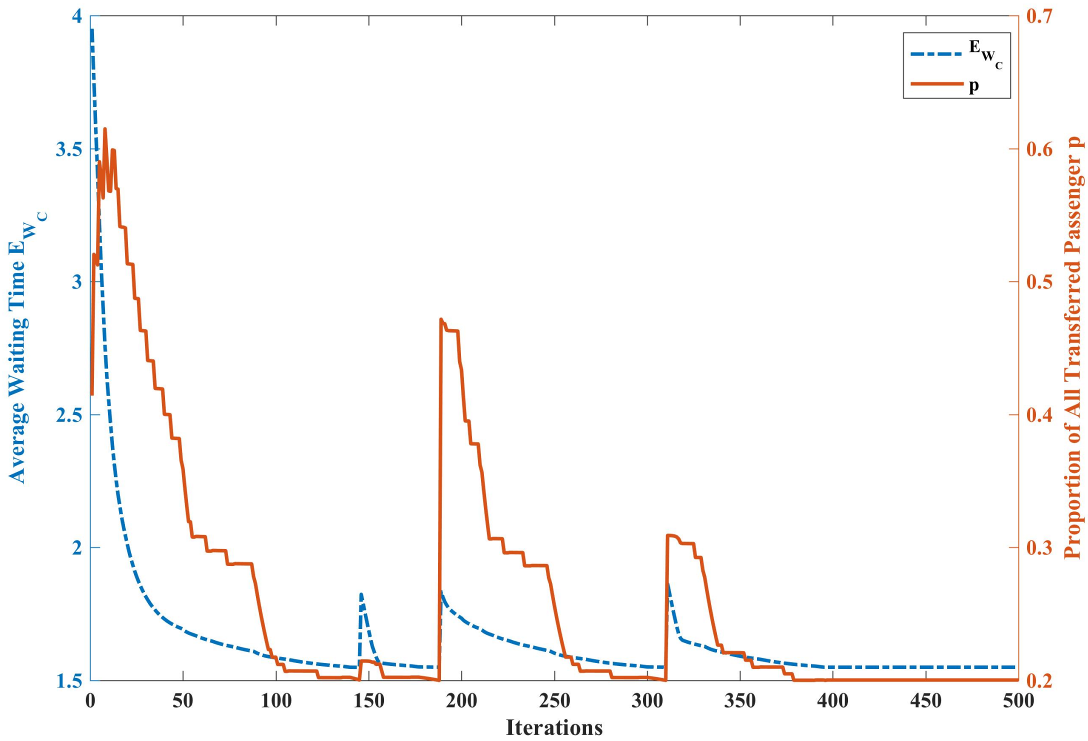


Figure 4.7: Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.2$ at Narita airport

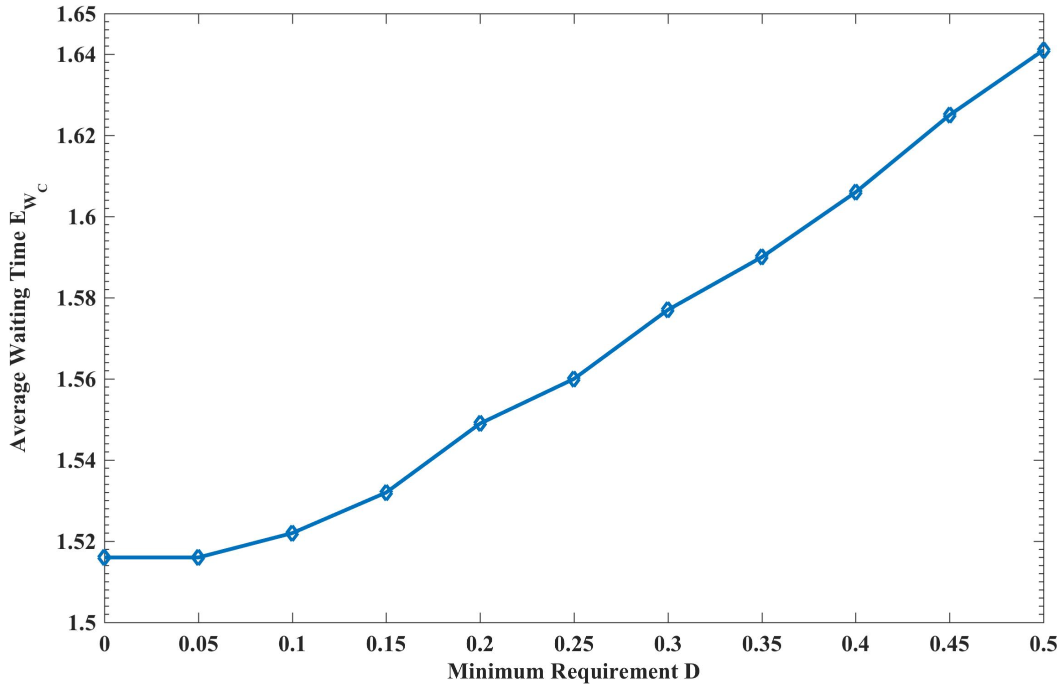


Figure 4.8: The average waiting time E_{W_C} of different D at Narita airport

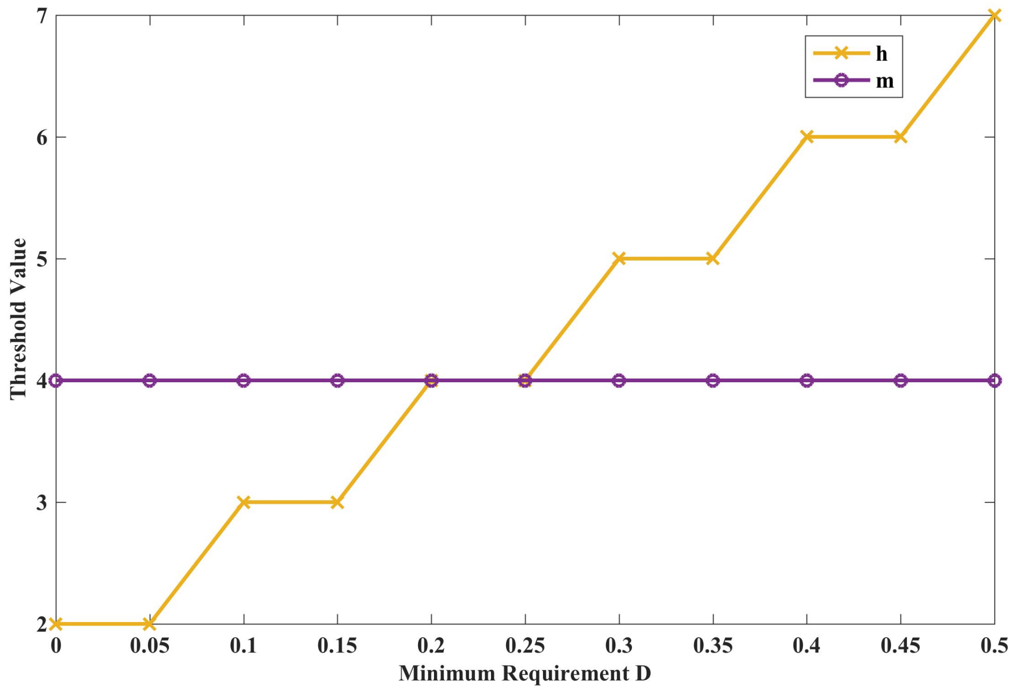


Figure 4.9: The thresholds h and m of different D at Narita airport

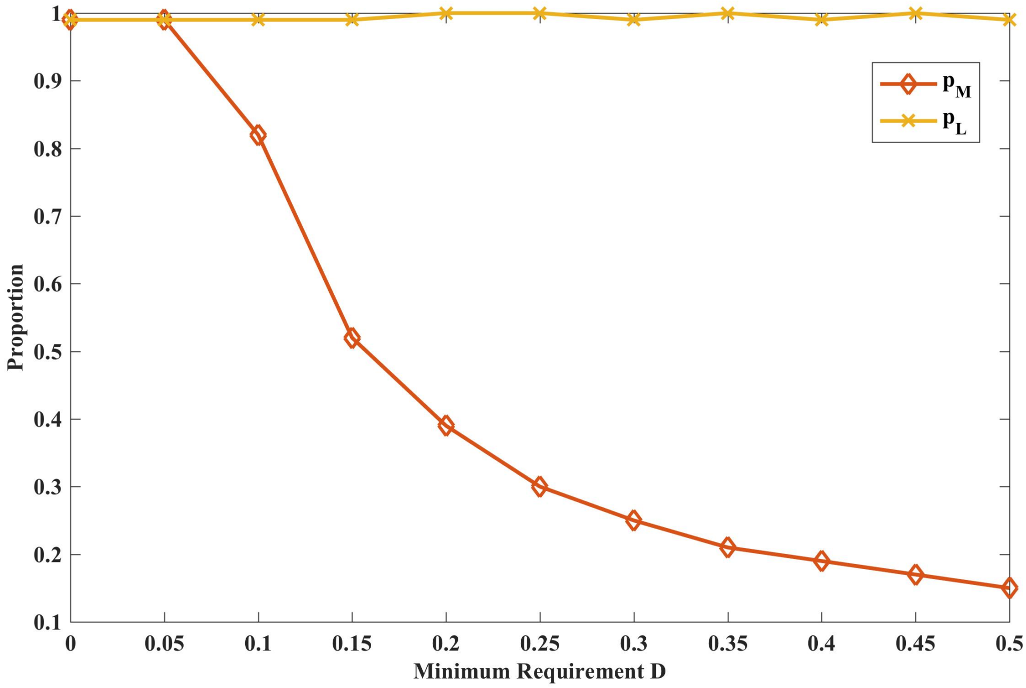


Figure 4.10: The proportions p_M and p_L of different D at Narita airport

Table 4.9: Arrival schedules at Sydney airport

Time Period	Number of Flights	Number of Passengers
18:00 - 19:00	8	1956
19:00 - 20:00	3	595
20:00 - 21:00	0	0
21:00 - 22:00	6	2311

we give the appropriate service rate in this case, Table 4.10 summarizes the information about parameters used in simulation at Sydney airport.

Table 4.10: Parameters used in simulation at Sydney airport

Classes	Proportion	Arrival Rate	Total Service Rate
H-class	10%	114.05	125
M-class	60%	684.30	720
L-class	30%	342.15	380

After optimization using the simulated annealing method, we get the optimal solutions at Sydney airport. Table 4.11 demonstrates the configuration of parameters with different D. Therefore we can find that, although the number of passenger is declined, the parameters with two sets of data have the same variation with increase of D. Figure 4.11 to Figure 4.14 are showing the changing of various parameters, which all have maintained the same trend with the previous case at Taoyuan airport.

Table 4.11: Optimal solutions by simulated annealing with different D at Sydney airport

D	M	h	m	p_M	p_L	E_{W_C}
0	100	1	2	0.99	0.99	1.752
0.05	100	2	2	0.26	1	1.782
0.10	100	3	2	0.12	0.99	1.809
0.15	100	3	2	0.05	0.99	1.824
0.20	100	4	2	0.04	0.99	1.836
0.25	100	5	2	0.03	1	1.849
0.30	100	6	2	0.02	0.99	1.864
0.35	100	7	2	0.02	1	1.873
0.40	100	7	2	0.01	0.99	1.879
0.45	100	9	2	0.01	0.99	1.890
0.50	100	10	2	0.01	1	1.896

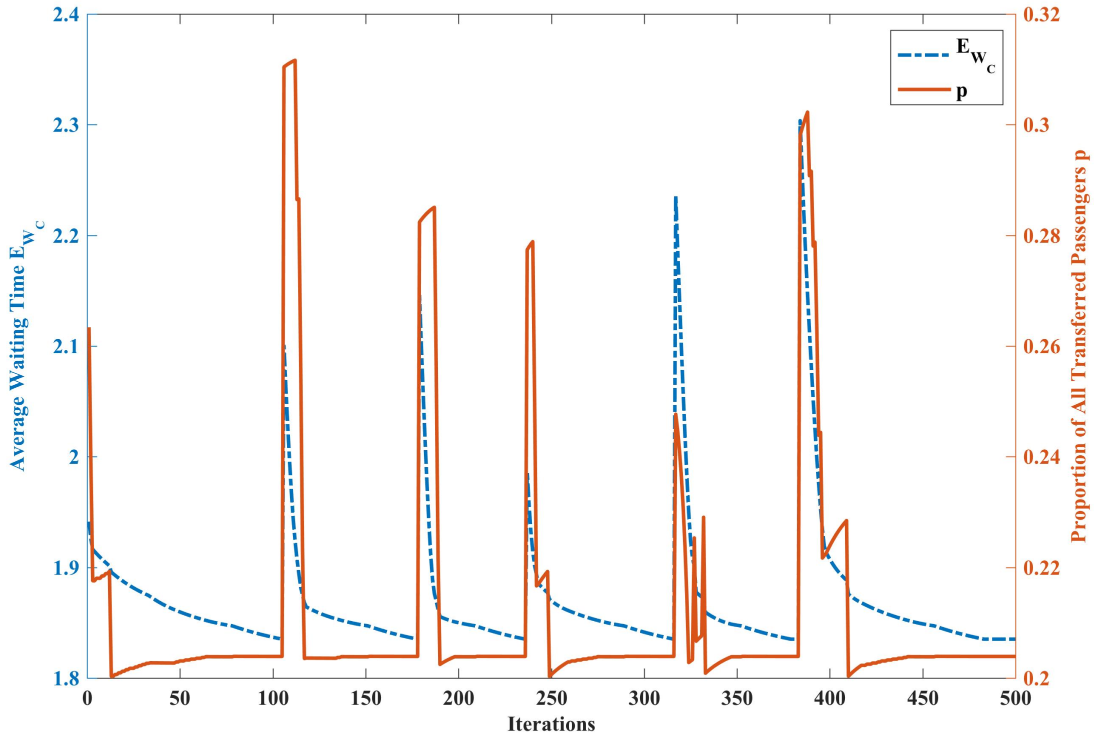


Figure 4.11: Comparison between proportion p and current average waiting time E_{W_C} with $D = 0.2$ at Sydney airport

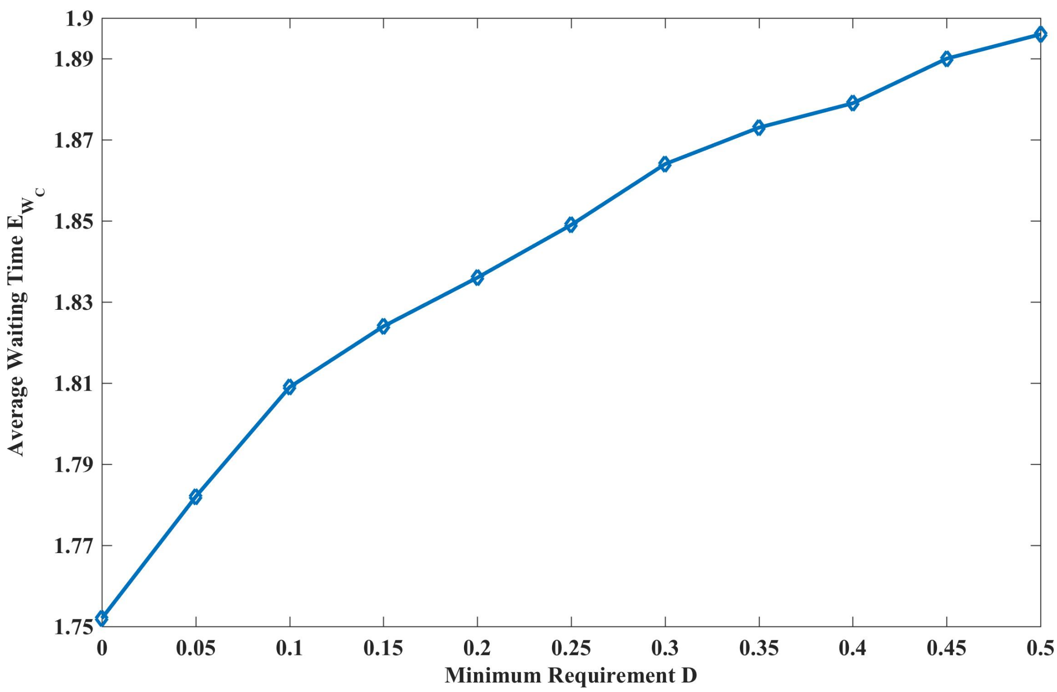


Figure 4.12: The average waiting time E_{W_C} of different D at Sydney airport

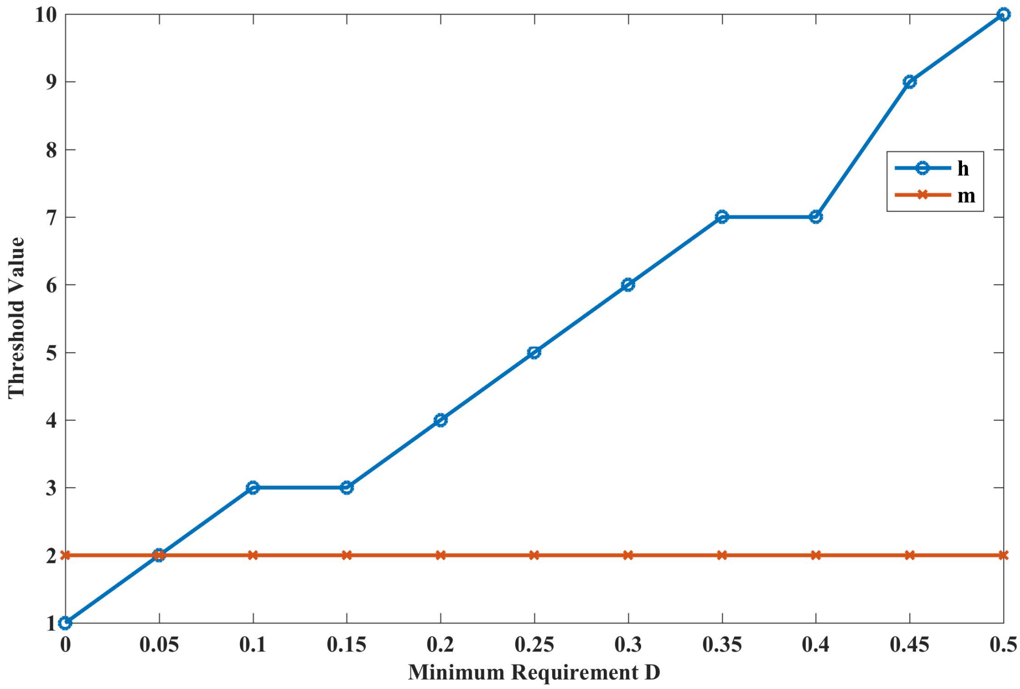


Figure 4.13: The thresholds h and m of different D at Sydney airport

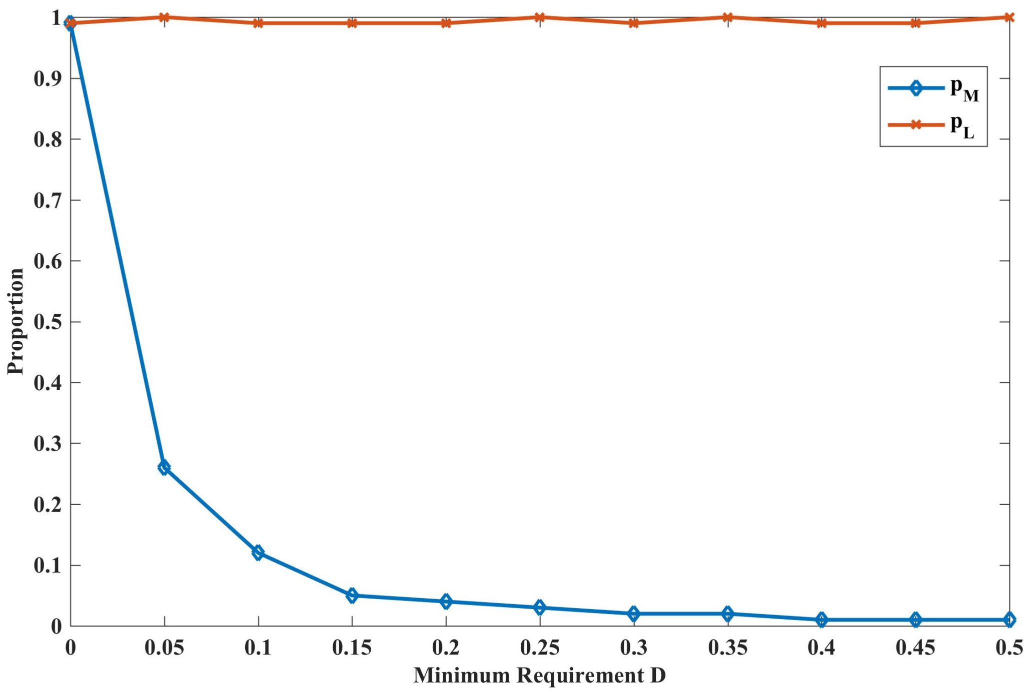


Figure 4.14: The proportions p_M and p_L of different D at Sydney airport

Chapter 5

Conclusions

This thesis reports a tiered security screening system for airports based on a two-dimensional Markov process and a Markov modulated Poisson process. The proposed model was evaluated using the matrix geometric method, wherein the optimal configuration of parameters is determined using simulated annealing. The proposed security screening system was shown to reduce the overall average waiting time, even as security was improved. We also constructed a comprehensive queueing strategy and a novel approach to calculating optimal model parameters. This makes it possible to adjust the configuration of the model according to the number of arriving passengers or the specific requirements of the system security.

The efficacy of the proposed methodology depends on the validity of profiling, in the form of a background pre-check by the TSA and the airline company. This makes it possible to send passengers to different queueing lanes according the risk they pose.

We present the following recommendations for future study. Despite reductions in the computation time of the proposed model, it still lags behind real-time calculations. Enabling calculations in real time will require adjustment of the queueing strategy according to the number of passengers arriving in real time. Secondly, in the determination of optimal solutions, it will be necessary to take into account not only the overall average waiting time but also the inspection cost and inconvenience cost of passengers.

Appendix A

Approximation for GI/G/c Queue

According to Choi et al. [7], they can approximate the $M/G/3$ queueing model by the $GI/G/c/c+r$ model. Let N denote the number of passengers in the system at an arbitrary time, and N^A (N^D) denotes the number of passengers that an arriving passenger finds (that a departing passenger leaves behind) in steady state. In particular, passengers who arrive to find $c+r$ passengers in the system are assumed to depart immediately from the system with those $c+r$ passengers left behind (note that we take into account those rejected passengers who find and leave behind $c+r$ passengers). Now let P_n, P_n^A, and P_n^D denote the probabilities that $N = n$, $N^A = n$, and $N^D = n$, respectively, for $0 \le n \le c+r$. These probabilities will be expressed in terms of the following quantities:

$$
\begin{aligned}
a &= E[A] = \frac{1}{\lambda}, \\
b &= E[S], \\
a_n^D &= E[A_n^D],\ 0 \le n \le c+r, \\
b_n^A &= E[S_n^A],\ 0 \le n \le c+r, \\
b_n^D &= E[S_n^D],\ 0 \le n \le c+r,
\end{aligned}
$$

where $A_n^D, 0 \le n \le c+r$, is the residual inter-arrival time at the departure instant of a customer who leaves behind n passengers in the system, and S_n^A (S_n^D), $1 \le n \le c+r$, is the residual service time of a randomly chosen busy server at the arrival instant (the departure instant) of a customer who finds (leaves behind) n passengers in the system. From these definitions, we have $a^D_{c+r} = a$ and $b^D_{c+r} = b^A_{c+r}$ (recall that passengers who leave behind $c+r$ passengers in the system are those who are rejected from the system immediately after their arrivals). Finally, we set $b^A_0 = b^D_0 = 0$ for convenience. We assume that all the above quantities are well defined and finite. Also, we have the steady-state queue-length distributions of the $GI/G/c/c+r$ queue as the following theorem.

Theorem. *The steady-state queue-length distributions of the $GI/G/c/c+r$ queue are given by*

$$P^A_n = P^D_n = P^A_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}},\ 0 \le n \le c+r,$$

and

$$P_n = P^A_n \gamma_n,\ 0 \le n \le c+r,$$

where

$$\frac{1}{\mu_i} = \begin{cases} b - i(a - a^D_{i-1}) + (i-1)(b^A_{i-1} - b^D_{i-1}), & 1 \le i \le c, \\ -c(a - a^D_{i-1}) + b^A_{i-1} + (c-1)(b^A_{i-1} - b^D_{i-1}), & c+1 \le i \le c+r, \end{cases}$$

$$\frac{1}{\lambda_i} = \begin{cases} (i+1)(a^D_{i+1} + b^A_{i+1} - b^D_{i+1}), & 0 \le i \le c-2, \\ ca^D_{i+1} + b^A_{i+1} - b + (c-1)(b^A_{i+1} - b^D_{i+1}), & c-1 \le i \le c+r-2, \\ ca, & i = c+r-1, \end{cases}$$

$$\gamma_i = \begin{cases} \lambda a^D_0, & i = 0, \\ \lambda \left[\frac{\mu_i (a - a^D_{i-1})}{\lambda_{i-1}} + a^D_i \right], & 1 \le i \le c+r, \end{cases}$$

and by normalization, $\sum_{n=0}^{c+r} P^A_n = 1$,

$$P^A_0 = \left[1 + \sum_{n=1}^{c+r} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} \right]^{-1}.$$

Since equations in the theorem above involve quantities a^D, b_n^A, and b_n^D, which are not easy to compute in general, except for some special cases such as Poisson arrivals or exponential service times. The approximations of steady-state queue-length distributions are obtained by replacing (unknown) arrival and departure-average quantities a_n^D, b_n^A, and b_n^D of the above theorem by their corresponding (well-known) time-average counterparts. That is

$$
\begin{aligned}
a_n^D \approx & a_R = \frac{E[A]^2}{2E[A]} = \frac{(1+c_A^2)a}{2}, \; 0 \le n \le c+r-1, \; and \\
b_n^A(b_n^D) \approx & b_R = \frac{E[S]^2}{2E[S]} = \frac{(1+c_S^2)b}{2}, \quad 1 \le n \le c+r-1,
\end{aligned}
$$

where $c_X^2 = Var[X]/(E[X])^2$ is the *squared coefficient of variation* (SCV) of a r.v. X with distribution function F. Note that $E\left[X^2\right]/2E\left[X\right]$ is the mean of the *equilibrium distribution* of F.

Finally, we obtain a two-moment approximation for the steady-sate queue-length distribution as follows:

$$
\begin{aligned}
&\tilde{P}_n^A = \tilde{P}_n^D = \tilde{P}_0^A \prod_{i=0}^{n-1} \frac{\tilde{\lambda}_i}{\tilde{\mu}_{i+1}}, \; 0 \le n \le c+r, \text{ and} \\
&\tilde{P}_n = \tilde{P}_n^A \tilde{\gamma}_n, \; 0 \le n \le c+r,
\end{aligned}
$$

where

$$
\frac{1}{\tilde{\mu}_i} = \begin{cases} b - i(a - a_R), & 1 \le i \le c, \\ -c(a-a_R) + b_R, & c+1 \le i \le c+r, \end{cases}
$$

$$
\frac{1}{\tilde{\lambda}_i} = \begin{cases} (i+1)a_R, & 0 \le i \le c-2, \\ ca_R + b_R - b, & c-1 \le i \le c+r-2, \text{ and} \\ ca, & i = c+r-1, \end{cases}
$$

$$\tilde{\gamma}_i = \begin{cases} \lambda a_R, & i = 0, \\ \lambda \left[\tilde{\mu}_i \left(a - a_R\right) / \tilde{\lambda}_{i-1} + a_R\right], & 1 \leq i \leq c + r - 1, \text{ and} \\ \lambda \left[\tilde{\mu}_i \left(a - a_R\right) / \tilde{\lambda}_{i-1} + a\right], & i = c + r. \end{cases}$$

Appendix B

Simulated Annealing Method

In this Appendix, we summarize the simulated annealing method which is reported by Kirkpatrick et al. [14] and Laarhoven and Aarts [24]. Simulated annealing is an effective method for finding a good optimal solution to an optimization problem. Usually, finding an optimal solution for some optimization can be a difficult task. This is not only because when the problem becomes sufficiently large we need to search enormous number of feasible solutions to find the optimal, but also because even using modern computing methods there are still too many feasible solutions to consider. In our model, we need to search five parameters, which are M, p_M, p_L, h, and m. The vector x_k denotes the five parameters at state k in the searching precess, and $f(x)$ is the objective function of our optimization model. It has over 90 million feasible solutions to be considered. Therefore, we can extremely decrease computing time via simulated annealing.

Firstly, let us figure out how simulated annealing works. The main idea of simulated annealing algorithm is inspired from the procedure of annealing in metallurgy. Annealing involves heating and cooling a metal to change its internal structure and physical properties. When the structure becomes fixed, metal consequently can retain its new obtained properties. So, in simulated annealing, we will set up a parameter called temperature to simulate the heating or cooling process. As the algorithm runs, the temperature will be allowed to decrease slowly. While the temperature parameter at high value, the simulated annealing algorithm will be allowed to accept feasible solutions that are worse than the current solution more frequency. This useful property gives the algorithm the ability to jump out of the local optimums. Therefore, simulated annealing algorithm is remarkably effective for finding a close to optimum solution when dealing with searching numerous local optimums.

Secondly, before shows the algorithm, we should introduce the acceptance probability function. Since simulated annealing will occasionally accept worse solutions, it depends on Boltzmann's function to decide which solutions to accept. That is

$$PR(f(x_k)) = \min\{1, e^{\left(\frac{-\Delta f}{T_k}\right)}\},$$

where $PR(f(x_k))$ is the probability that the algorithm will accept the current state x_k, Δf is the difference of the objective function value between current state x_k and the previous state x_{k-1}, and T_k is the control temperature in state k. After having the $PR(f(x_k))$, we compute a random variable V between 0 and 1 to compare with it. If $V > PR(f(x_k))$, then we abandon the state x_k, otherwise accept it. Eventually, in our algorithm, we check if the neighbor solution is better than our current solution. If it is, we accept it unconditionally. If however, the neighbor solution is worse than the current solution, we use Boltzmann's function to decide whether accept it or not. From the Boltzmann's function we know that the algorithm is more likely accept solutions at high temperatures. Actually, the temperature parameter T is not constant, we set $T_{k+1} = 0.95T_k$.

Finally, we give the pseudo-code of the simulated annealing in Algorithm 1.

Algorithm 1 Simulated Annealing Algorithm

Require: $f(x)$: objective function; x_0: initial solution; T_0: initial temperature; K: step size;
Ensure: optimal best value x_b

Initial $k = 0$, current best state $x_b = x_0$, current best value $f(x_b) = f(x_0)$, current temperature $T_C = T_0$;
while $k < K$ **do**
    Compute the neighbor of x_k, which is x_{k+1};
    if $f(x_{k+1}) \leq f(x_k)$ **then**
        $x_b = x_{k+1}$, $f(x_b) = f(x_{k+1})$;
    else
        Compute the $PR(f(x_k))$ and the random variable V;
        if $V \leq PR(f(x_k))$ **then**
            $x_b = x_{k+1}$, $f(x_b) = f(x_{k+1})$;
        else
            $x_b = x_k$, $f(x_b) = f(x_k)$;
        end if
    end if
    $T_{k+1} = 0.95T_k$ and $k = k + 1$;
end while

Bibliography

[1] Computer-Assisted Passenger Prescreening System, 2004. https://en.wikipedia.org/wiki/Computer-Assisted_Passenger_Prescreening_System.

[2] TSA announces expansion of Black Diamond self-select lanes to Norfolk international airport, 2008. http://www.marketwired.com/press-release/tsa-announces-expansion-black-diamond-self-select-lanes-norfolk-international-airport-927005.htm.

[3] Secure Flight Program, 2009. https://en.wikipedia.org/wiki/Secure_Flight.

[4] Support grows for tiered risk system at airports, 2011. http://www.nytimes.com/2011/02/08/business/08security.html?_r=0.

[5] Arnold Barnett. CAPPS II: The foundation of aviation security? *Risk Analysis*, 24(4): 909–916, 2004.

[6] Huseyin Cavusoglu, Byungwan Koh, and Srinivasan Raghunathan. An analysis of the impact of passenger profiling for transportation security. *Operations Research*, 58(5): 1287–1302, 2010.

[7] Dae W. Choi, Nam K. Kim, and Kyung C. Chae. A two-moment approximation for the $GI/G/c$ queue with finite capacity. *INFORMS Journal on Computing*, 17(1):75–81, 2005.

[8] Robert de Lange, Ilya Samoilovich, and Bo van der Rhee. Virtual queuing at airport security lanes. *European Journal of Operational Research*, 225(1):153–165, 2013.

[9] Wolfgang Fischer and Kathleen Meier-Hellstern. The Markov-modulated Poisson process ($MMPP$) cookbook. *Performance evaluation*, 18(2):149–171, 1993.

[10] Donald Gross and Donald Gross. *Fundamentals of queueing theory*. Wiley series in probability and statistics. Wiley, Hoboken, N.J., 4th edition, 2008.

[11] Boudewijn R Haverkort, Aad PA van Moorsel, and Arvid Dijkstra. MGMtool: A performance modelling tool based on matrix geometric techniques. 1992.

[12] Oliver C. Ibe. *Markov processes for stochastic modeling*. Academic Press, Amsterdam, Boston, 2009.

[13] Sheldon H Jacobson, Tamana Karnani, John E Kobza, and Lynsey Ritchie. A cost-benefit analysis of alternative device configurations for aviation-checked baggage security screening. *Risk Analysis*, 26(2):297–310, 2006.

[14] Scott Kirkpatrick, C Daniel Gelatt, Mario P Vecchi, et al. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[15] Q.-L. Li and J. Cao. A Computational Framework for the Mixing Times in the QBD Processes with Infinitely-Many Levels. *ArXiv e-prints*, aug 2013.

[16] Randolph Nelson. *Probability, stochastic processes, and queueing theory : the mathematics of computer performance modeling*. Springer-Verlag, New York, 1995.

[17] Xiaofeng Nie, Gautam Parab, Rajan Batta, and Li Lin. Simulation-based selectee lane queueing design for passenger checkpoint screening. *European Journal of Operational Research*, 219(1):146–155, 2012.

[18] Md Mostafizur Rahman and Attahiru Sule Alfa. Computational procedures for a class of $GI/D/k$ systems in discrete time. *Journal of Probability and Statistics*, 2009:1–18, 2009.

[19] George Passantino Robert W. Poole Jr. A risk-based airport security policy. *Policy Study*, 308, 2003.

[20] Sheldon M. Ross. *Stochastic processes*. Wiley series in probability and statistics Probability and statistics. Wiley, New York, 2nd edition, 1996.

[21] Young U Ryu and Hyeun-Suk Rhee. Evaluation of intrusion detection systems under a resource constraint. *ACM Transactions on Information and System Security (TISSEC)*, 11(4):20, 2008.

[22] Hans-Peter Schwefel. *Performance Analysis of Intermediate Systems Serving Aggregated ON/OFF Traffic with Long-Range Dependent Properties*. Dissertation, Technische Universität München, München, 2000.

[23] Cen Song and Jun Zhuang. Two-stage security screening strategies in the face of strategic applicants, congestions and screening errors. *Annals of Operations Research*, pages 1–26, 2015.

[24] Peter J Van Laarhoven and Emile H Aarts. *Simulated annealing: theory and applications*, volume 37. Springer Science & Business Media, 1987.

[25] Ward Whitt. Queues with service times and interarrival times depending linearly and randomly upon waiting times. *Queueing Systems*, 6(1):335–351, 1990.

[26] Zhe George Zhang, Hsing Paul Luh, and Chia-Hung Wang. Modeling security-check queues. *Management Science*, 57(11):1979–1995, 2011.