

Marginal Cost Congestion Pricing under Approximate Equilibrium Conditions

by

Mattias Jansson

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 22, 1998 [June 1998]

© Mattias R. Jansson, MCMXCVIII. All rights reserved.

The author hereby grants MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document in whole or in part, and to grant others to do so.

Author.....

Department of Electrical Engineering and Computer Science
May 22, 1998

Certified by.....

Amedeo R. Odoni
Professor of Aeronautics and Astronautics and of
Civil and Environmental Engineering
Thesis Supervisor

Accepted by.....

Arthur C. Smith
Chairman, Department Committee on Graduate Theses

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUL 14 1998

LIBRARIES

ENG

Marginal Cost Congestion Pricing under Approximate Equilibrium Conditions

by

Mattias Jansson

Submitted to the Department of Electrical Engineering and Computer Science
on May 22, 1998, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

We formulate four stochastic steady-state bottleneck models to calculate economically optimal equilibrium congestion fees. The models use Little's Law to obtain closed form expressions for steady-state marginal cost congestion fees in M/M/1 and M/G/1 systems for n types of users with and without nonpreemptive priorities. We show that the models can be used to calculate equilibrium congestion fees and demand levels for any monotonically decreasing demand function. We also demonstrate that marginal cost congestion pricing can decrease delays significantly in a socially optimal way at a relative low decrease in utilization of the facility.

Thesis Supervisor: Dr. Amedeo R. Odoni

Title: Professor of Aeronautics and Astronautics and of Civil and Environmental
Engineering

Acknowledgements

I would like to express my gratitude to the following people for making this thesis possible:

Dr. Amedeo Odoni, for his continual support and advice throughout the project. He provided direction and showed extreme clear-sightedness in times of confusion.

My parents Rolf and Barbro, and my sister Karin, and all members of my family, for encouraging me to realize my dreams and for their unconditional love and concern.

My friends Anders Rensmo, Erik Dahren, Anders Carlsson, Louise Johnsson, Lisa Dalsbro, Tove Franzson, Petter Skantze, Henrik and Christine Martin, Ann-Christin Axelsson, Eva Nygren, Roger Söderkvist, and all other people that I have had the fortune to enjoy being with and learn from

Lastly, You know who You are. For all the incredible times, for all the great wisdom, and for showing me not only who I am but who I can be.

Contents

1 Introduction	7
2 Literature Review	8
3 General Approach	11
3.1 Model Description	11
3.2 Steady-State Considerations	13
3.3 Definitions	15
3.4 A Fundamental Observation	16
3.5 Cost of Queueing: Without Priorities	18
3.6 M/M/1 Queue: Without Priorities	19
3.7 M/G/1 Queue: Without Priorities	20
3.8 Cost of Queueing: Nonpreemptive Priorities	20
3.9 M/M/1 Queue: Nonpreemptive Priorities	22
3.10 M/G/1 Queue: Nonpreemptive Priorities	23
3.11 Elasticity of Demand	24
3.12 Finding Equilibrium Congestion Fees	25
3.13 Equilibrium without Priorities	26
3.14 Equilibrium with Nonpreemptive Priorities	28
4 Applications	31
4.1 Examples with One Type of User	31
4.2 Examples with Multiple Types of Users	35
4.2.1 Fee versus No Fee Comparison	37
4.2.2 M/M/1 versus M/G/1 Comparison	39
4.2.3 Nonpreemptive Priorities versus No Priorities Comparison	40
5 Conclusions and Discussion	44
A Examination of What Types of Users Benefit From Priorities	49

List of Figures

3-1	Supply Curve for Marginal Cost Congestion Pricing	25
4-1	M/G/1 Equilibrium Demand for 1 Type of User	32
4-2	Equilibrium Demand Level	33
4-3	Equilibrium Congestion Fee	33
4-4	Demand and Congestion Fee as a Function of Elasticity (2D)	34
4-5	Demand and Congestion Fee as a Function of Elasticity (3D)	35
4-6	Demand Functions for Multiple Types of Users Examples	36
4-7	Comparison of Congestion Fee vs. No Fee	37
4-8	Comparison of M/M/1 and M/G/1 Models	39
4-9	Nonpreemptive Priorities vs. No Priorities	40
4-10	Internal and External Costs with Different Priority Orders	43

List of Tables

2-1 Different measures aimed to decrease congestion at airports	9
3-1 The four different models examined	12
4-1 Parameter Settings for One User Type Example	31
4-2 Parameter Settings for Multiple Types of Users Examples	35
4-3 Comparison of No Congestion Fee vs. Optimal Congestion Fee	38
4-4 Parameters for Ordering Example	41
4-5 Equilibrium Costs with Different Priority Orders	42
A-1 Factorial Design Parameters	49

Chapter 1

Introduction

How many times have you found yourself being delayed waiting for takeoff when flying out of Logan Airport? Why is it that so many flights need to wait before they gain access to the runway? The reason is simple: when the airport is operating near capacity the runways become congested and during peak hours delays become very long.

The problem of congestion is naturally not limited to airports, but applies to any facility with a finite capacity. Thus, the models and results presented in this thesis can be applied to any congested facility, even if the examples shown here are specific to the framework of an airport.

Throughout the thesis, we examine peak-load pricing as a means to decrease delays at congested facilities. We develop four related queueing models that calculate equilibrium congestion tolls. In Section 2, we present the background and look at previous work on this problem. In Section 3, we introduce the queueing models, which calculate optimal congestion fees given the characteristics of the facility as well as the users. In Section 4, we examine some numerical applications intended to clarify the main characteristics of the models. In section 5, we draw conclusions and discuss directions for future work.

Chapter 2

Literature Review

As air transportation has evolved into a nearly ubiquitous mode of travel, airports all over the world have struggled with the increasingly challenging problem of runway congestion. The FAA (Federal Aviation Administration) estimates that in the U.S. alone, congestion costs airlines and passengers about five billion dollars every year [DANI 95]. In Europe, that number is closer to 6.5 billion dollars [REED 92]. Traditionally, capital investments such as the construction of new runways or entire new airports have been the means of expanding airport capacity to alleviate congestion problems [ODON 76]. However, making these capital investments has become increasingly difficult due to their high costs in conjunction with the prohibitive political climate for airport expansion, due to increased noise levels near airports, environmental and safety concerns associated with air traffic, etc. In some cases, there is simply no additional physical space available for further airport expansion.

Thus, airports have attempted to devise measures to accommodate increased demand. At most airports there is one or two large demand peaks during the day, and there is very little demand in the evening and at night. It would therefore be desirable to balance demand between the peak and the off-peak periods. The main three attempted types of measures in this respect are listed in Table 2-1 [ROSE 75].

Type of Measure	Description
Administrative	Imposing limits on the number of aircraft that may use the runways during given time periods.
Economic	Charging higher runway usage fees at peak hours.
Hybrid	Combinations of the two.

Table 2-1: Different measures aimed to decrease congestion at airports

Administrative methods have been implemented at several airports around the world, and many have proven to work reasonably well to decrease congestion. However, it is far from clear that administrative limitations provide desirable long-term effects. The system tends to be biased toward maintaining status quo, and does not give preference to the users that derive the most benefit from using the airport [ECKE 72 and PULL 72]. Pure economic measures try to link the private as well as social costs and benefits of utilizing the runway. An additional aircraft should not only consider its own cost when deciding whether to use the airport, but should also consider the impact on all other users, the external cost [ODON 76].

The seminal paper on marginal cost congestion pricing is “Congestion Theory and Transport Investment” by Nobel laureate William Vickrey [VICK 69]. Vickrey introduces modeling congestion as a queue behind a bottleneck as well as the concept of time-varying congestion tolls. After Vickrey, other studies have used simple deterministic processes to generate equilibrium fees using the bottleneck model. Good examples include Arnott, de Palma, and Lindsey [ARNO 93], and Henderson [HEND 85]. Koopman [KOOP 72] uses a stochastic queueing model to calculate congestion fees, but

assumes that arrival rates are fixed and thus does not determine the equilibrium fees.

Daniel [DANI 95] uses numerical methods to combine a stochastic queueing model and the bottleneck model to find equilibrium congestion fees at airports. This thesis utilizes Little's Law [LITT 61] to obtain closed form expressions for optimal steady-state congestion fees. Furthermore, we calculate equilibrium congestion fees and associated equilibrium demand levels for any monotonically decreasing demand function.

Chapter 3

General Approach

3.1 Model Description

In this section, four different single server queueing models will be developed. Each provides a framework for marginal cost pricing of congested facilities. All of the models assume that there are several distinct types of facility users, e.g. private aircraft, small commercial aircraft, and large commercial aircraft, in the case of an airport. The models can be divided into two categories; one that uses a pure first come, first served (FCFS) queue discipline, and one that gives certain types of users priority over other types. Before getting into the details of the various models, a general description of queueing systems is needed.

Most queueing systems can be described using the $A/B/m$ notation where

- A indicates the probability distribution of the user interarrival times
- B the distribution of the service times
- m is an integer indicating the number of servers (e.g. runways) available.

In this paper the $M/M/1$ and the $M/G/1$ queue will be discussed, where M stands for “memoryless” or “Markovian” meaning a Poisson distribution, and G means a “general” distribution (i.e. any distribution). Each queueing model will be examined in the context of nonpreemptive priorities (the user currently in service will not be forced back to the

queue if a user of higher priority arrives) as well as in the context of no priorities. See Table 3-1 below.

	M/M/1	M/G/1
Without priorities	Yes	Yes
Nonpreemptive priorities	Yes	Yes

Table 3-1: The four different models examined

The probability density function for the first-order interarrival times of a Poisson process is given by $f_T(t) = \lambda e^{-\lambda t}$, where λ is the user arrival rate, and the random variable T is the time until the next arrival. From this definition we can obtain the mean interarrival time $E[T]=1/\lambda$, and the variance $\sigma_T^2 = 1/\lambda^2$. Thus, a Poisson process can be fully described by λ , the arrival rate (users/unit time). The properties of the Poisson process have been thoroughly examined and it is arguably the best understood non-trivial arrival process. Due to its relative simplicity, the Poisson distribution is commonly used in urban operations research when modeling any type of arrivals (e.g. emergency calls to 911, arrivals of people to a customer service desk etc.), and it often turns out to be a reasonable model. For example, the Poisson model works quite well for modeling the arrivals (and departures) of airplanes at an airport during periods when average demand is approximately constant and reasonably high.

However, since the Poisson distribution is fully characterized by the arrival rate, λ , the variance and the mean cannot be set independently. This means that we lose one degree of freedom for the advantage of the relative analytical ease of use. If the process we wish to model does not fit this assumption, the need for a more flexible model arises.

For instance, in the airport setting, most planes in a specific category need approximately the same time (i.e. the variance is low), for landings or takeoffs, which motivates the use of the general distribution for the service times. In the M/G/1 queueing model, the distribution of the service time, S , is characterized by the average service time, $E[S]=1/\mu$ (service rate = μ) and the variance, σ_s^2 , which can both be set independently of each other.

3.2 Steady-State Considerations

Queueing theory is quite powerful at estimating the low moments and central moments of quantities such as expected waiting time and expected length of the queue under equilibrium or “steady-state” conditions. However, using steady-state expressions clearly leads to a problem if the arrival rate exceeds the available service rate. This would lead to an infinite build-up in the queue over time and the system is said to “blow up.” The implications of this dilemma are quite different depending on whether the system operates with or without priorities. In the context of first come, first served (FCFS), the simple answer is that no steady-state expressions can be found and there is no equilibrium solution to the problem. However, if there are classes of users that have priority over other classes, there may be cases where the higher priority customers receive full service and reach an equilibrium state, but customers in lower priority classes do not, as will be illustrated below.

First, we define the overall arrival rate and the available service rate as λ and μ , respectively. Furthermore we define ρ as the utilization ratio,

$$\rho = \frac{\lambda}{\mu} \tag{3.1}$$

If $\rho < 1$, the arrival rate is less than the available service rate and steady-state expressions for such quantities as average queue length and average waiting time can indeed be found.

Now assume that there are r distinct types of users at a facility and let ρ be defined by $\rho = \sum_{i=1}^r \rho_i$. In the context of no priorities, as described above, if $\rho < 1$ all types of users will receive service and otherwise the queue grows to infinity. In the context of nonpreemptive priorities, however, there could be the case that

$$\sum_{i=1}^m \rho_i < 1 \quad (3.2)$$

but

$$\sum_{i=1}^{m+1} \rho_i \geq 1 \quad (3.3)$$

where the first m classes receive full service and reach an equilibrium state, and the others do not.

There are two appropriate conclusions we can draw here. Without priorities the system may or may not reach steady-state depending on whether the overall arrival rate is smaller or greater than the available service capacity. In the context of nonpreemptive priorities the system will always reach steady-state for one or more classes except if the arrival rate of the highest priority class exceeds the available service capacity. For the remainder of this thesis, m will denote the number of high priority types of users, that satisfies the two conditions in (3.2) and (3.3) when steady-state is not reached for all classes. Note that $m=r$ when steady-state is reached for all classes. In the context of no priorities m is always equal to r (if ρ is not less than one, no steady-state expressions can

be obtained and the analysis breaks down). In other words: m indicates the number of types of users that reach an equilibrium state and is fully served at the facility. However, for the remainder of this thesis it will be assumed that only m types of users continue to use the facility, and that users of types that do not experience bounded waiting times drop out. This assumption guarantees that $\rho < 1$ and steady-state is reached for all remaining types of users.

3.3 Definitions

Assume that each type of users arrives in a Poisson manner with an arrival rate of λ_i . Thus, the overall arrival rate is given by

$$\lambda = \sum_{i=1}^r \lambda_i \quad (3.4)$$

Similarly, the average service times for the different types are defined as $1/\mu_i$, and the weighted average service time can be written as

$$\frac{1}{\mu} = \sum_{i=1}^m \left(\frac{\lambda_i}{\lambda} \times \frac{1}{\mu_i} \right) \quad (3.5)$$

where $\lambda_i/\lambda = \text{Prob}(\text{ a user is of type } i)$. Note that only the m first types are included, since those are the only customers receiving full service. Furthermore, the overall utilization ratio is thus given by

$$\rho = \sum_{i=1}^m \rho_i = \sum_{i=1}^m \frac{\lambda_i}{\mu_i} \quad (3.6)$$

and the service rate

$$\mu = \frac{\lambda}{\rho} \quad (3.7)$$

For the M/G/1 system, we define σ_i^2 to be the variance of the service time of users of type i . From basic probability theory, the second moments of the service times can be written as

$$E[S_i^2] = \frac{1}{\mu_i^2} + \sigma_i^2 \quad (3.8)$$

and,

$$E[S^2] = \sum_{i=1}^m \left(\frac{\lambda_i}{\lambda} E[S_i^2] \right) \quad (3.9)$$

is the overall second moment of the service times of users receiving full service.

Finally, c_i is defined to be the cost of delay/unit time per user of type i . In the context of no priorities, the average cost of delay/unit time per user can be written as

$$c = \sum_{i=1}^m \left(\frac{\lambda_i}{\lambda} c_i \right) \quad (3.10)$$

However, in the nonpreemptive priority scheme, only the specific type costs will be used.

3.4 A Fundamental Observation

We begin with a fundamental observation on the relation between queueing theory and marginal cost pricing. To keep things simple we focus on a single queue without priorities, but entirely analogous ideas apply to priority queues. Furthermore, these ideas are discussed in the context of congested airport runways, but again the concepts are easily transferred to other settings.

Pure economic measures try to link the private and the social costs and benefits of utilizing the runways of a congested airport. An aircraft should not only consider its own

cost when deciding whether to use the airport, but should also consider the impact on all other airport users [ODON 76]. This leads to the concept of social marginal cost pricing, which is what this thesis intends to explore in some detail.

The nature of the impact of an additional aircraft on other users is eloquently described by Carlin and Park:

During a period when an airport is continuously busy, each user imposes some delay on all following users until the end of the busy period. That is, an additional user shoves those following him one space back in the queue, and the effect persists until the queue dissipates [CARL 70].

The marginal cost of an additional user at a facility can conveniently be broken into two parts, which are referred to as “internal cost” and “external cost” respectively. The internal cost represents the waiting cost experienced by the additional user due to the already present congestion, and the external cost is the increased cost for all other users due to the increased congestion caused by the additional user.

Thus, if the runway landing fee were set to equal the external cost, every user would have to consider the impact of its use of the runway on all other users, in addition to the internal cost which the user already pays implicitly (gas, flight crew salary, opportunity cost of lost time, etc.). If the benefit to a user of using the runway exceeds the external cost, that user will be willing to pay the landing fee. Conversely, if the negative impact imposed on the other users is greater than the benefit the additional user derives from using the airport, that user will choose not to use the airport.

Hence, the basic idea of marginal cost pricing is as follows: in order to maximize social economic welfare, when deciding whether to use a facility, prospective users should be forced to consider not only their own internal cost but also the external costs, i.e., the increased cost imposed on other users of the facility. In plain terms, if someone wants to use a facility, he or she needs to consider the increased cost to those users that experience delay due to his or her occupying the facility for some period of time.

3.5 Cost of queueing: Without priorities

In order to appropriately calculate the congestion fees, it is necessary to find an expression for the total expected cost/unit time, C , associated with delay in the system. In the context of no priorities, C can be written as the average cost of delay/unit time per user multiplied by the expected number of users in the queue,

$$C = c\bar{L}_q \quad (3.11)$$

where \bar{L}_q is the steady-state expected length of the queue. Using Little's Law,

$$\bar{L}_q = \lambda\bar{W}_q \quad (3.12)$$

where \bar{W}_q is the expected waiting time in the queue, (3.11) can be re-written as

$$C = c\lambda\bar{W}_q \quad (3.13)$$

Thus, the marginal cost of an additional user of type i ,

$$MC(i) = \frac{dC}{d\lambda_i} = c_i\bar{W}_q + c\lambda \frac{d\bar{W}_q}{d\lambda_i} \quad (3.14)$$

The marginal cost is thus divided into two terms, which can be referred to as the internal cost, $(c_i \bar{W}_q)$, and the external cost, respectively. The internal cost is indeed the cost experienced by the user due to the already existing delay, and the external cost is the increased cost for all other users due to the additional user.

3.6 M/M/1 Queue: Without Priorities

For the M/M/1 queue, the steady-state expected waiting time in the queue is well-known [LARS 81 and GROS 85]

$$\bar{W}_q = \frac{\rho}{\mu - \lambda} \quad (3.15)$$

Differentiating \bar{W}_q ,

$$\frac{d\bar{W}_q}{d\lambda_i} = \frac{\left(\frac{2\mu - \lambda}{\mu_i}\right) - (1 - \rho)}{(\mu - \lambda)^2} \quad (3.16)$$

which, using (3.14), gives the final expression,

$$\frac{dC}{d\lambda_i} = c_i \frac{\rho}{\mu - \lambda} + c\lambda \frac{\left(\frac{2\mu - \lambda}{\mu_i}\right) - (1 - \rho)}{(\mu - \lambda)^2} \quad (3.17)$$

For the case where there is only one type of user, this expression simplifies considerably to

$$\frac{dC}{d\lambda} = c \frac{\rho}{\mu - \lambda} + \frac{c\lambda}{(\mu - \lambda)^2} \quad (3.18)$$

3.7 M/G/1 Queue: Without Priorities

The analysis for the M/G/1 queue follows closely the M/M/1 analysis, but now the different service time variances must be taken into account. Again, the expected waiting time in the queue is well-known [LARS 81 and GROS 85],

$$\bar{W}_q = \frac{\lambda E[S^2]}{2(1-\rho)} \quad (3.19)$$

where $E[S^2]$ is defined as in (3.9). Differentiating this expression with respect to λ_i

$$\frac{d\bar{W}_q}{d\lambda_i} = \frac{(1-\rho)E[S_i^2] + \frac{\lambda}{\mu_i}E[S^2]}{2(1-\rho)^2} \quad (3.20)$$

and substituting into (3.14)

$$\frac{dC}{d\lambda_i} = c_i \frac{\lambda E[S^2]}{2(1-\rho)} + c\lambda \frac{(1-\rho)E[S_i^2] + \frac{\lambda}{\mu_i}E[S^2]}{2(1-\rho)^2} \quad (3.21)$$

Again, when there is only one type of user, the expression simplifies notably:

$$\frac{dC}{d\lambda} = c \frac{\lambda E[S^2]}{2(1-\rho)} + \frac{c\lambda E[S^2]}{2(1-\rho)^2} \quad (3.22)$$

3.8 Cost of Queuing: Nonpreemptive Priorities

In a queue where certain types of users have priority over other types, the low-priority users are likely to experience longer delays than the high-priority users.

Equations (3.11), (3.13) and (3.14) are not valid in this context, and need to be replaced.

The overall cost can be calculated by treating each class separately, and adding up all of the costs.

$$C = \sum_{k=1}^m \lambda_k c_k \bar{W}_{qk} \quad (3.23)$$

where \bar{W}_{qk} is the expected time of waiting in the queue for a user of type k . Users in classes that never receive service are ignored, under the assumption that they do not affect the situation for the users that gain full access to the facility. Thus,

$$\frac{dC}{d\lambda_i} = c_i \bar{W}_{qi} + \sum_{k=1}^m \left(\lambda_k c_k \frac{d\bar{W}_{qk}}{d\lambda_i} \right) \quad (3.24)$$

and again, the two terms represent the internal cost and the external cost as defined above. The expected waiting time in queue for a user has been found [COBH 54] to be

$$\bar{W}_{qk} = \begin{cases} \frac{\bar{W}_0}{(1-a_{k-1})(1-a_k)} & , k \leq m \\ \infty & , k > m \end{cases} \quad (3.25)$$

$$a_k \equiv \sum_{i=1}^k \rho_i \quad (3.26)$$

$$a_0 \equiv 0 \quad (3.27)$$

where \bar{W}_0 represents the expected remaining time in service for the user occupying the server (e.g. using the runway) when a new user arrives, and a_k is used to simplify the notation for the terms that describe how different types of users experience different delays. In general, \bar{W}_0 consists of two parts: one represented by the m classes of users that experience bounded waiting times at the server, and another by those users in class $m+1$ that receive service only during the fraction of the time when there are no queues in the top m classes. Using random incidence and the fact that when we have Poisson arrivals

the probability of finding a user of a particular class occupying the server is simply equal to the fraction of the time the users of that class occupy the server. Those fractions in turn are equal to the utilization ratios ρ_i [LARS 81].

$$\bar{W}_0 = \sum_{i=1}^m \frac{\rho_i E[S_i^2]}{2E[S_i]} + \frac{(1-a_m)E[S_{m+1}^2]}{2E[S_{m+1}]} \quad (3.28)$$

The term $(1-a_m)$ corresponds to the fraction of the time that class $m+1$ occupies the server. For the case of negative exponential service times, (3.28) reduces to

$$\bar{W}_0 = \sum_{i=1}^m \frac{\rho_i}{\mu_i} + \frac{(1-a_m)}{\mu_{m+1}} \quad (3.29)$$

However, under the assumption that only the first m classes continue to use the facility, expressions (3.28) and (3.29) reduce to

$$\bar{W}_0 = \sum_{i=1}^m \frac{\rho_i E[S_i^2]}{2E[S_i]} \quad (3.30)$$

and,

$$\bar{W}_0 = \sum_{i=1}^m \frac{\rho_i}{\mu_i} \quad (3.31)$$

respectively.

3.9 M/M/1 Queue: Nonpreemptive Priorities

Thus, by combining (3.25) and (3.31) the expected waiting time in queue for the M/M/1 nonpreemptive priority queue for a user of type k can be written as

$$\bar{W}_{qk} = \begin{cases} \frac{\sum_{i=1}^m \rho_i}{\mu_i} & , k \leq m \\ \frac{(1-a_{k-1})(1-a_k)}{\infty} & , k > m \end{cases} \quad (3.32)$$

Differentiating (3.32) with respect to λ_i for the case where $k \leq m$,

$$\frac{d\bar{W}_{qk}}{d\lambda_i} = \frac{\frac{(1-a_{k-1})(1-a_k)}{\mu_i^2} - \bar{W}_0 \frac{d[(1-a_{k-1})(1-a_k)]}{d\lambda_i}}{[(1-a_{k-1})(1-a_k)]^2} \quad (3.33)$$

where,

$$\frac{d[(1-a_{k-1})(1-a_k)]}{d\lambda_i} = \begin{cases} \frac{a_k + a_{k-1} - 2}{\mu_i} & , i < k \\ \frac{a_{k-1} - 1}{\mu_i} & , i = k \\ 0 & , i > k \end{cases} \quad (3.34)$$

The expressions in (3.32) and (3.33) can then finally be substituted back into (3.24) to find the marginal cost for the M/M/1 queue with nonpreemptive priorities. For the case with only one class, this complex expression reduces to (3.18), which is expected since priorities should not matter in the absence of multiple types of users.

3.10 M/G/1 Queue: Nonpreemptive Priorities

Similarly to the M/M/1 system, the expected waiting time for the M/G/1 queue can be found by combining (3.25) and (3.30)

$$\bar{W}_{qk} = \begin{cases} \frac{\sum_{i=1}^m \frac{\rho_i E[S_i^2]}{2E[S_i]}}{(1-a_{k-1})(1-a_k)} & , k \leq m \\ \infty & , k > m \end{cases} \quad (3.35)$$

Again, differentiating with respect to λ_i for the case where $k \leq m$,

$$\frac{d\bar{W}_{qk}}{d\lambda_i} = \frac{\left(\frac{E[S_i^2]}{2\mu_i E[S_i]} \right) (1-a_{k-1})(1-a_k) - \bar{W}_0 \frac{d[(1-a_{k-1})(1-a_k)]}{d\lambda_i}}{[(1-a_{k-1})(1-a_k)]^2} \quad (3.36)$$

with all terms defined as above. Now, the expressions in (3.35) and (3.36) can then finally be substituted back into (3.24) to find the marginal cost for the M/G/1 queue with nonpreemptive priorities. Again, the resulting expression reduces to (3.18) for the case of only one class.

3.11 Elasticity of Demand

We have shown that for Poisson arrivals, if the cost of delay for each type of user and the first and second moments of the service time are known, it is possible to calculate the appropriate congestion fee. But, if the demand for using the facility is price sensitive, introducing a congestion fee will have a dampening effect on the arrival rates to the system. One approach to taking this effect into account would be to wait and see what the new conditions are, and then recalculate the appropriate congestion fee. However, that would again alter the conditions and the process would have to be repeated until convergence. Hopefully, equilibrium would eventually be reached where the congestion fee and demand would be consistent, but this process is very cumbersome and time-consuming. For instance, in the case of airports the process would take many years. Thus, it is likely that demand and service characteristics will change during this time, and the system may never get sufficiently close to equilibrium.

However, if the demand function for each type of user is known, it should be possible to immediately calculate a congestion fee such that equilibrium is obtained. The following sections intend to demonstrate this proposition. Examining the form of the expressions used to calculate the congestion fees, it should be clear that the appropriate congestion fee is a monotonically increasing function of demand. The closer to capacity the system is operating, the more congested it is and more users are affected by the

marginal user which translates to a higher external cost. Using similar reasoning, we can argue that the internal cost is a monotonically increasing function of demand as well, meaning that the total cost for a user is monotonically increasing which intuitively makes sense.

Assuming that price sensitive users take their total cost of using a facility into account when deciding whether to use a facility, we need to include the internal cost into the user demand functions. Furthermore, if we assume that the demand function is monotonically decreasing with the total cost of using the facility there must be a unique equilibrium. See Figure 3-1 below.

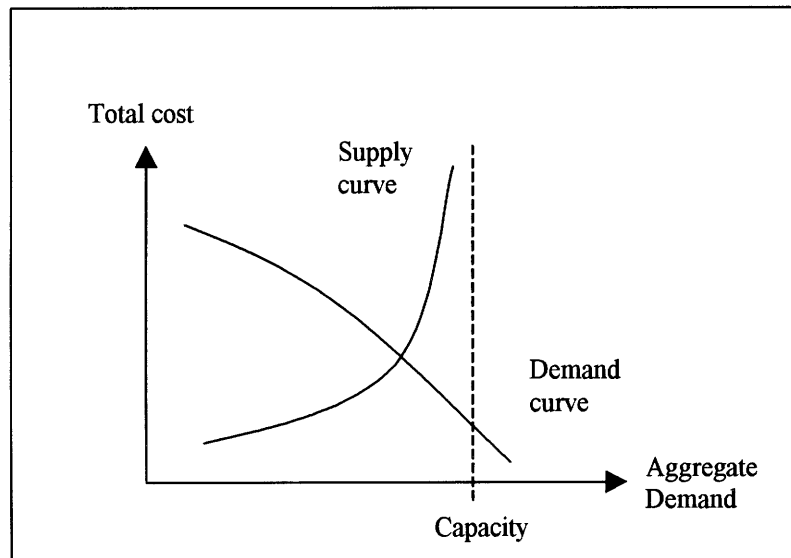


Figure 3-1: Supply Curve for Marginal Cost Congestion Pricing

3.12 Finding Equilibrium Congestion Fees

We can summarize the problem as follows: given a demand curve for each type of user, calculate the optimal equilibrium congestion fee for each type. This can be done by finding the intersection between the supply curve (the total cost of using the facility) and the aggregate demand curve. In order to find this aggregate equilibrium we must

simultaneously find m equilibrium points since the total cost for each type depends on the arrival rate of all types of users. First, we need to define some notation for the problem.

x_i - Total cost of using the facility for a user of type i .

$x = \{x_1, x_2, \dots\}$ - The set of all total costs.

$\lambda_i(x_i)$ - Arrival rate of type i users as a function of the total cost for type i ; it can be any monotonically decreasing function.

$\lambda(x)$ - Total arrival rate.

$\bar{W}_q(x)$ - The expected waiting time.

$\mu(x)$ - Total available service rate.

$E[S^2(x)]$ - Second moment of the service time as a function of all total costs.

It may not be immediately obvious why the overall service rate and the second moment of the overall service time may be dependent on the total costs of using the facility. However, remember that these quantities are dependent on the arrival rates, which in turn are cost dependent. Some users may be more price sensitive than others and change the proportion of arrivals due to different user classes, which in turn alters the overall service characteristics.

3.13 Equilibrium without Priorities

Recall the expression (3.14) for finding the total marginal cost for a system without priorities. Combining (3.10) and (3.14), using our slightly extended notation, the total cost of using the facility for type i users can be written as

$$x_i = c_i \bar{W}_q(x) + \left(\sum_{j=1}^m c_j \lambda_j(x_j) \right) \frac{d\bar{W}_q(x)}{d\lambda_i(x_i)}, \quad \forall i \quad (3.37)$$

Thus we have a system of $m=r$ equations that need to be solved simultaneously to find the total cost for each type of user. Generally, it is not possible to obtain a closed form expression due to the non-linearity of the expected waiting time and the general nature of the demand functions. However, it is not difficult to find numerical solutions to (3.37) using a simple iterative search. Any guess in the right-hand side of (3.37) will yield a result that indicates if the guess is below or above the equilibrium point.

For the M/M/1 system we can re-write (3.15) as

$$\bar{W}_q(x) = \frac{\lambda(x)}{\mu(x)(\mu(x) - \lambda(x))} \quad (3.38)$$

and (3.16) as

$$\frac{d\bar{W}_q(x)}{d\lambda_i(x_i)} = \frac{\left(\frac{2\mu(x) - \lambda(x)}{\mu_i}\right) - (1 - \rho(x))}{(\mu(x) - \lambda(x))^2} \quad (3.39)$$

which leads to the following set of equations

$$x_i = c_i \frac{\rho(x)}{\mu(x) - \lambda(x)} + \left(\sum_{j=1}^m c_j \lambda_j(x_j)\right) \frac{\left(\frac{2\mu(x) - \lambda(x)}{\mu_i}\right) - (1 - \rho(x))}{(\mu(x) - \lambda(x))^2}, \quad \forall i \quad (3.40)$$

If there is only one type of user, (3.38) reduces to

$$x = c \frac{\rho(x)}{\mu - \lambda(x)} + \frac{c\lambda(x)}{(\mu - \lambda(x))^2} \quad (3.41)$$

Note that in this case the service rate is independent of the total cost.

Similarly, for the M/G/1 queue we can re-write (3.19) and (3.20)

$$\bar{W}_q(x) = \frac{\lambda(x)E[S^2(x)]}{2(1 - \rho(x))} \quad (3.42)$$

$$\frac{d\bar{W}_q(x)}{d\lambda_i(x_i)} = \frac{(1-\rho(x))E[S_i^2] + \frac{\lambda(x)}{\mu_i}E[S^2(x)]}{2(1-\rho(x))^2} \quad (3.43)$$

which together yield the following set of equations

$$x_i = c_i \frac{\lambda(x)E[S^2(x)]}{2(1-\rho(x))} + \left(\sum_{j=1}^m c_j \lambda_j(x_j) \right) \frac{(1-\rho(x))E[S_i^2] + \frac{\lambda(x)}{\mu_i}E[S^2(x)]}{2(1-\rho(x))^2}, \quad \forall i \quad (3.44)$$

For the case of only one type, (3.44) reduces to

$$x = c \frac{\lambda(x)E[S^2]}{2(1-\rho(x))} + \frac{c\lambda(x)E[S^2]}{2(1-\rho(x))^2} \quad (3.45)$$

3.14 Equilibrium with Nonpreemptive Priorities

The case with nonpreemptive priorities follows closely the reasoning in paragraph 3.13, with the substitution of the appropriate equations. First, we re-write the general expression for calculating the total cost of using the facility.

$$x_i = c_i \bar{W}_{qi}(x) + \sum_{k=1}^m \left(\lambda_k(x_k) c_k \frac{d\bar{W}_{qk}(x)}{d\lambda_i(x_i)} \right), \quad \forall i \quad (3.46)$$

Next we re-write (3.25), (3.26), (3.27), (3.30), and (3.31) as follows

$$\bar{W}_{qk}(x) = \begin{cases} \frac{\bar{W}_0(x)}{(1-a_{k-1}(x))(1-a_k(x))} & , k \leq m \\ \infty & , k > m \end{cases} \quad (3.47)$$

$$a_k(x) \equiv \sum_{i=1}^k \rho_i(x) \quad (3.48)$$

$$a_0 \equiv 0 \quad (3.49)$$

$$\bar{W}_0(x) = \sum_{i=1}^m \frac{\rho_i(x)E[S_i^2]}{2E[S_i]} \quad (3.50)$$

For the case of negative exponential service times, (3.50) reduces to

$$\bar{W}_0(x) = \sum_{i=1}^m \frac{\rho_i(x)}{\mu_i} \quad (3.51)$$

For the M/M/1 queue, we can combine (3.47) and (3.51) for an expression for the expected waiting time in queue for a user of type k

$$\bar{W}_{qk}(x) = \begin{cases} \frac{\sum_{i=1}^m \frac{\rho_i(x)}{\mu_i}}{(1-a_{k-1}(x))(1-a_k(x))} & , k \leq m \\ \infty & , k > m \end{cases} \quad (3.52)$$

Differentiating with respect to $\lambda_i(x_i)$

$$\frac{d\bar{W}_{qk}(x)}{d\lambda_i(x_i)} = \frac{\frac{(1-a_{k-1}(x))(1-a_k(x))}{\mu_i^2} - \bar{W}_0(x) \frac{d[(1-a_{k-1}(x))(1-a_k(x))]}{d\lambda_i(x_i)}}{[(1-a_{k-1}(x))(1-a_k(x))]^2} \quad (3.53)$$

Substituting (3.51), (3.52) and (3.53) into (3.46) we can calculate the total cost of using the facility for the M/M/1 queue.

For the M/G/1 queue, the same logic can be applied to obtain the following expressions.

$$\bar{W}_{qk}(x) = \begin{cases} \frac{\sum_{i=1}^m \frac{\rho_i(x)E[S_i^2]}{2E[S_i]}}{(1-a_{k-1}(x))(1-a_k(x))} & , k \leq m \\ \infty & , k > m \end{cases} \quad (3.54)$$

Again, differentiating with respect to λ_i for the case where $k \leq m$,

$$\frac{d\bar{W}_{qk}(x)}{d\lambda_i(x_i)} = \frac{\left(\frac{E[S_i^2]}{2\mu_i E[S_i]}\right) (1-a_{k-1}(x))(1-a_k(x)) - \bar{W}_0(x) \frac{d[(1-a_{k-1}(x))(1-a_k(x))]}{d\lambda_i(x_i)}}{[(1-a_{k-1}(x))(1-a_k(x))]^2} \quad (3.55)$$

Substituting (3.50), (3.54) and (3.55) into (3.46) we can calculate the total cost of using the facility for the M/G/1 queue

Chapter 4

Applications

4.1 Examples with One Type of User

In this section, sample results are presented to highlight some of the characteristics of the different queueing models. First, we demonstrate a simple example with only one type of user in an M/G/1 system. The issue of priorities naturally becomes irrelevant when there is only one type of user, so either model will yield the same results. We have chosen parameter values typical of an airport setting (see Table 4-1), but as mentioned earlier the models can be applied to any type of congested facility. The examples use a linear demand function, but the models work equally well with any monotonically decreasing function, as is demonstrated in section 4.2

Parameter	Value
Available service rate (hr^{-1})	120
Standard deviation of expected service time (seconds)	10
Cost of delay (\$/hr)	2000
Demand with no delay cost (hr^{-1})	115
Elasticity of demand ((users/hr)/\$100)	3

Table 4-1: Parameter Settings for One User Type Example

Figure 3 below shows the demand function using the data in Table 2 with the calculated equilibrium demand indicated by a circle. Recall that users take the total cost,

which includes the internal cost (due to waiting) and the external cost (the congestion fee), into account when deciding whether to use the facility.

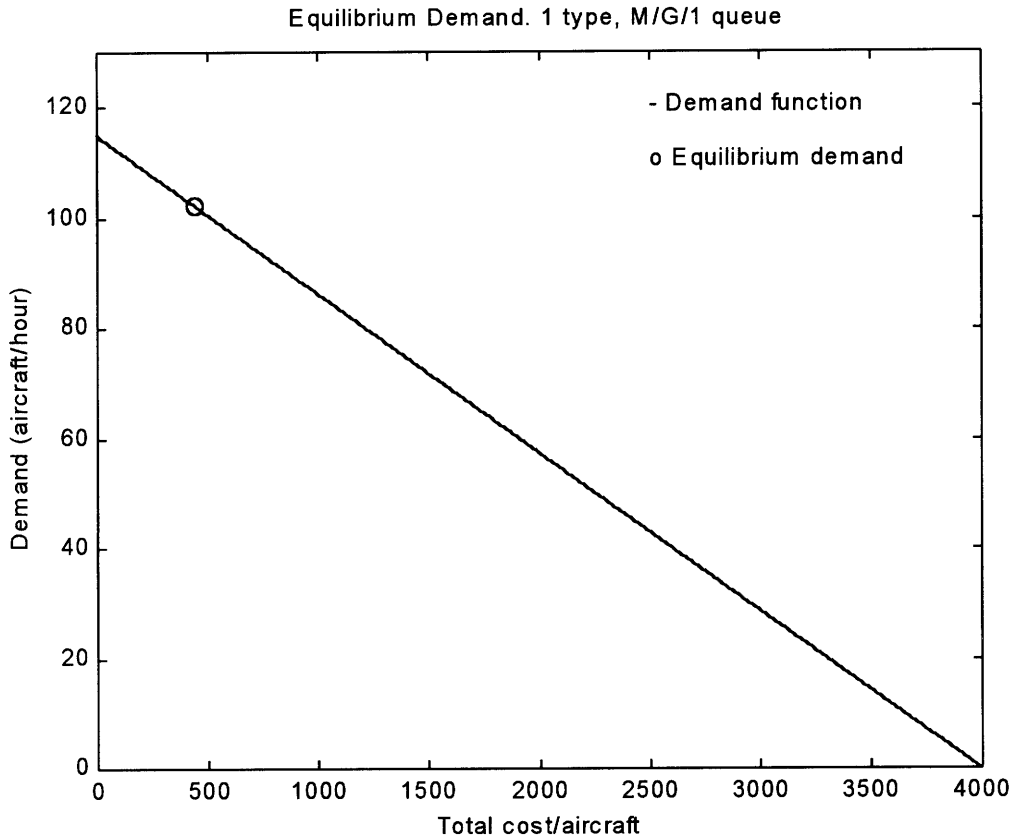


Figure 4-1: M/G/1 Equilibrium Demand for 1 Type of User

The next example is intended to illustrate the effects of varied demand elasticity. The zero cost demand is set fixed at 115 users/hour, and the slope of the demand function is varied from being flat (inelastic demand) to twenty users/hour per \$100. Figures 4-2 and 4-3 show the equilibrium demand level and the equilibrium congestion fee as a function of elasticity. In the figures it can be seen that as elasticity increases, making the users more price-sensitive, the equilibrium demand and the equilibrium congestion fee both decrease.

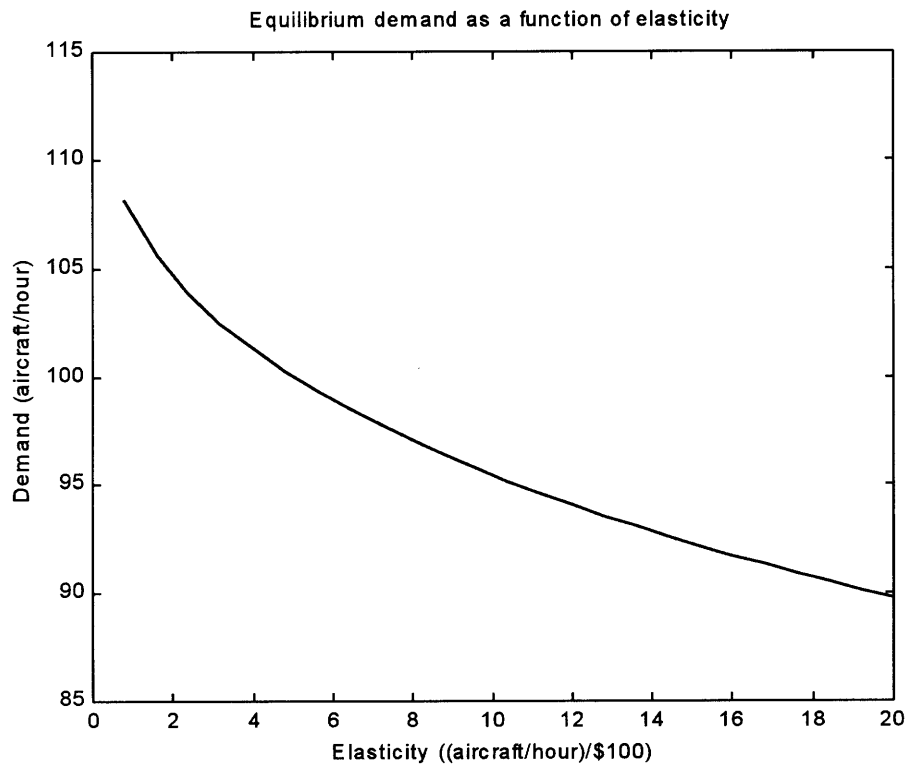


Figure 4-2: Equilibrium Demand Level

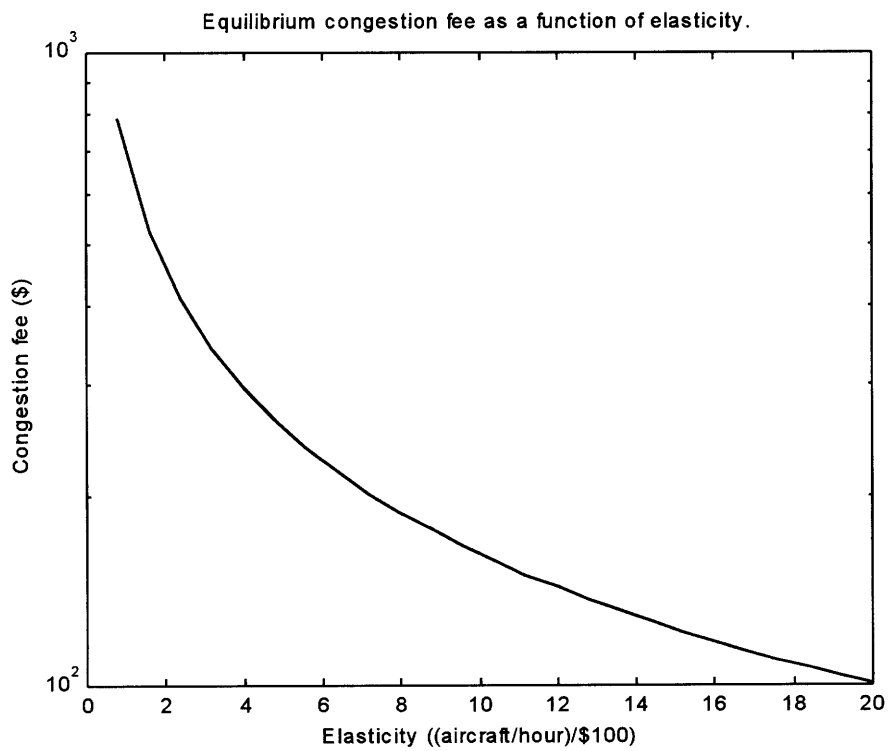


Figure 4-3: Equilibrium Congestion Fee

Figures 4-4 and 4-5 are intended to show more clearly how the equilibrium demand level and the equilibrium congestion fee vary together with the elasticity of the users. Figure 4-4 is a 2-D plot of how demand and external cost increase as elasticity decreases, and Figure 4-5 is a 3-D representation of the same data. The demand levels are given in users per hour, and the congestion fee in dollars. The numbers on the curve indicate the elasticities in users per hour/\$100.

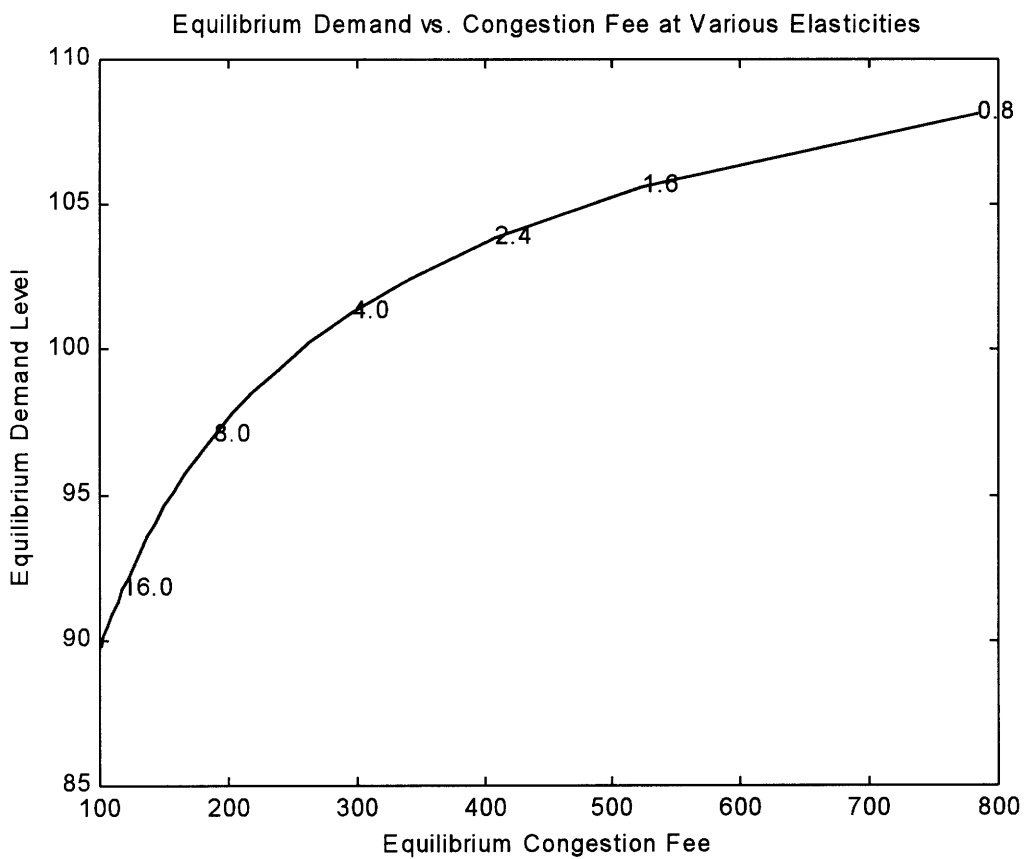


Figure 4-4: Demand and Congestion Fee as a Function of Elasticity (2D)

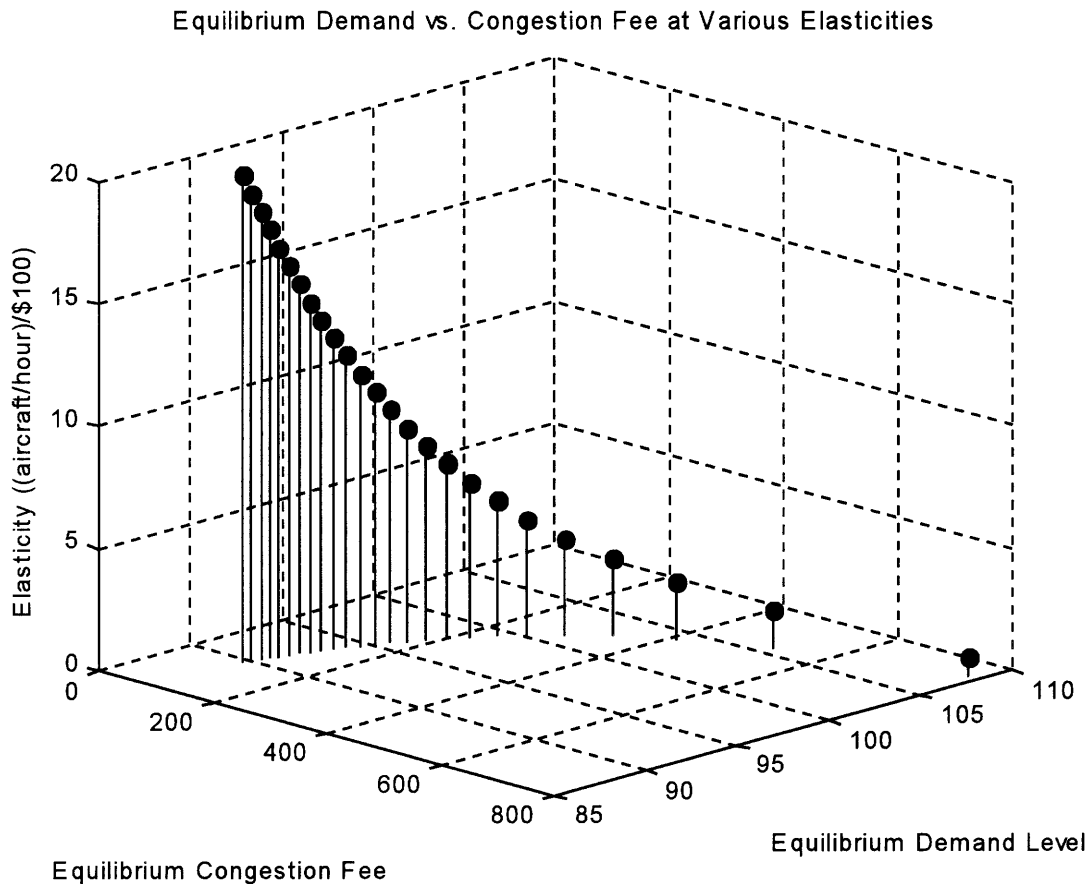


Figure 4-5: Demand and Congestion Fee as a Function of Elasticity (3D)

4.2 Examples with Multiple Types of Users

Here, we use several examples with multiple types of users to illustrate some of the differences between the queueing models. For simplicity, the examples use only three types of users, but the models can be extended to include any number of types of users.

The parameters used in the following examples are listed in Table 4-2.

Parameter	Type 1	Type 2	Type 3
Service rate (hr^{-1})	80	90	100
Standard deviation of service time (seconds)	10	10	10
Cost of delay (\$/hr)	2,500	1,000	400
Demand function	$\lambda_1 = 40 - 0.001x$	$\lambda_2 = 50 - 0.003x$	$\lambda_3 = 60 - 0.01x$
(x = total cost)	$-1 * 10^{-5} x^2$	$-2 * 10^{-5} x^2$	$-8 * 10^{-5} x^2$

Table 4-2: Parameter Settings for Multiple Types of Users Examples

The given standard deviations in Table 3 are only used in the M/G/1 models, since the negative exponential service times in the M/M/1 model intrinsically have $1/\mu^2$ variance. Figure 4-6 shows the different demand functions for the three types of users. In order to make the examples easier to follow, the three types could be viewed as three different types of aircraft; large commercial jets (Type 1), small commercial jets (Type 2), and private aircraft (Type 3). For instance, large commercial jets are likely to be less price-sensitive than other aircraft. It should be noted that it is very difficult to estimate the demand functions of aircraft, and that the numbers used in these examples are hypothetical.

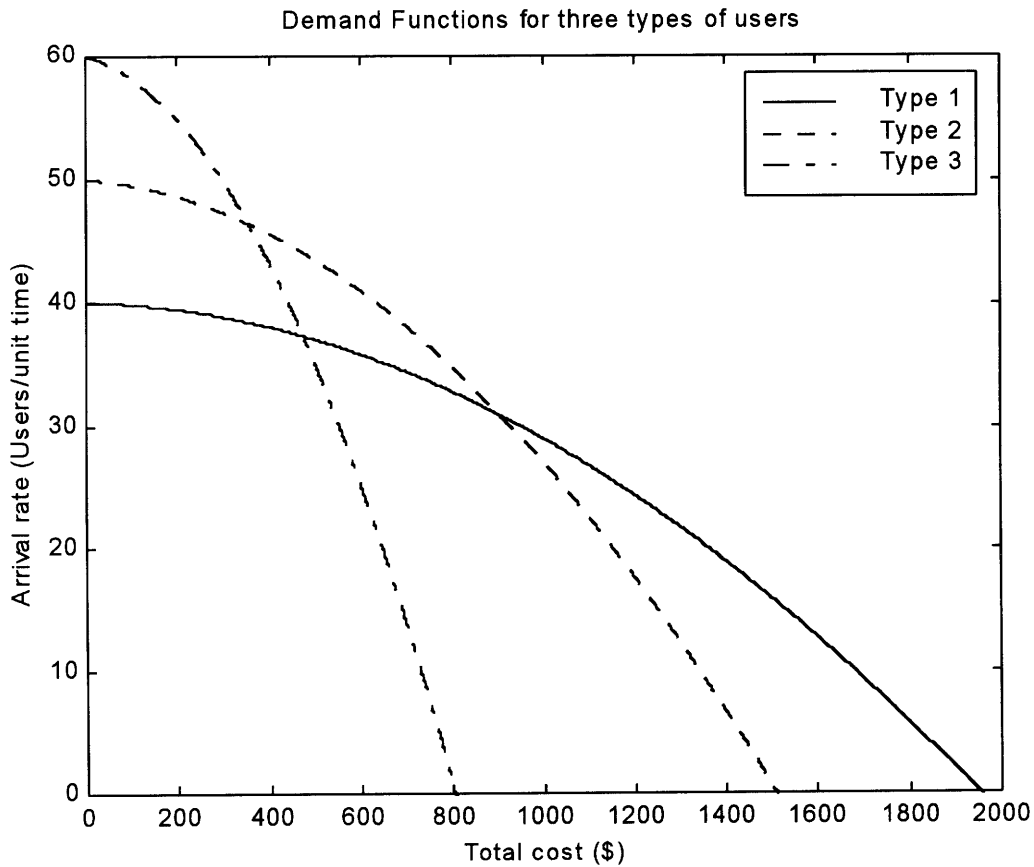


Figure 4-6: Demand Functions for Multiple Types of Users Examples

4.2.1 Fee versus No Fee Comparison

Perhaps the most important question to be asked is: “How will equilibrium demand levels, costs and delays change when a congestion fee is introduced?” The idea of marginal cost congestion pricing would be of little use if we were not to see significant decreases in delay. Figure 4-7 shows how the demand levels change in an M/G/1 system without priorities, when the optimal congestion fee is introduced using the data in Table 4-2.

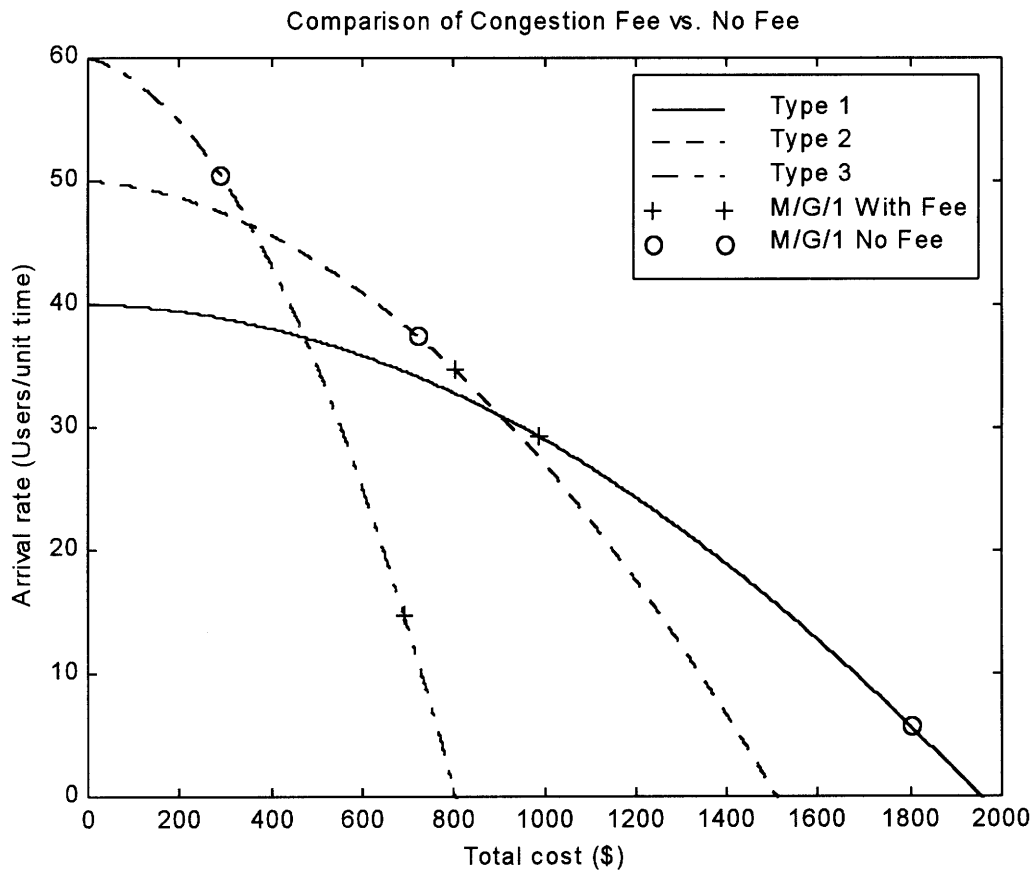


Figure 4-7: Comparison of Congestion Fee vs. No Fee

As can be seen in Figure 4-7, the different types of users react differently to the congestion fee. The Type 1 equilibrium demand level increases significantly, while it decreases for the other two types. This should not be surprising, but is explained by the

cost structure for this example and the changes in delays according to Table 4-3 below. The cost of delay associated with Type 1 is much higher than for the other types (\$2,500/hr vs. \$1,000/hr and \$400/hr), which makes Type 1 users much more sensitive to delays. Thus, in this example, the increase in cost due to the congestion fee is more than offset by the reduction in delay costs, and more Type 1 users choose to use the facility. Conversely, the reduced cost of delay for Type 2 and Type 3 users is not enough to compensate for the congestion fee. Thus, these users experience a net increase in total cost and their demand levels decrease.

Variable	Without congestion fee	With congestion fee
Expected Delay/User	43 min 15 sec	3 min 15 sec
Utilization (rho)	99.22%	89.89%
Type 1 Congestion Fee	\$0	\$853.47
Type 2 Congestion Fee	\$0	\$750.44
Type 3 Congestion Fee	\$0	\$669.58
Type 1 Delay Cost/User	\$1,802.20	\$135.03
Type 2 Delay Cost/User	\$720.88	\$54.12
Type 3 Delay Cost/User	\$288.35	\$21.65

Table 4-3: Comparison of No Congestion Fee vs. Optimal Congestion Fee

The overall demand decreased by slightly less than 10%, and the average delays decreased from 43 to 3 min! This indicates that introducing a congestion fee is a very powerful tool for decreasing congestion.

4.2.2 M/M/1 versus M/G/1 Comparison

In this section, we make a comparison between the M/M/1 and the M/G/1 queueing models. First, we note that the results from the M/M/1 model are the same as the M/G/1 model when the variance is set to $1/\mu^2$. In this example, the standard deviation of the expected service time is 10 seconds ($=1/360$ hours), which is less than $1/\mu=1/120$ hours. Thus, since increased variance translates to increased delays (eq. 3.22), our M/G/1 queue should yield lower costs at a given demand level. That, in turn, should translate to lower equilibrium costs and higher equilibrium demand levels. Figure 4-8 below shows that this is indeed the case.

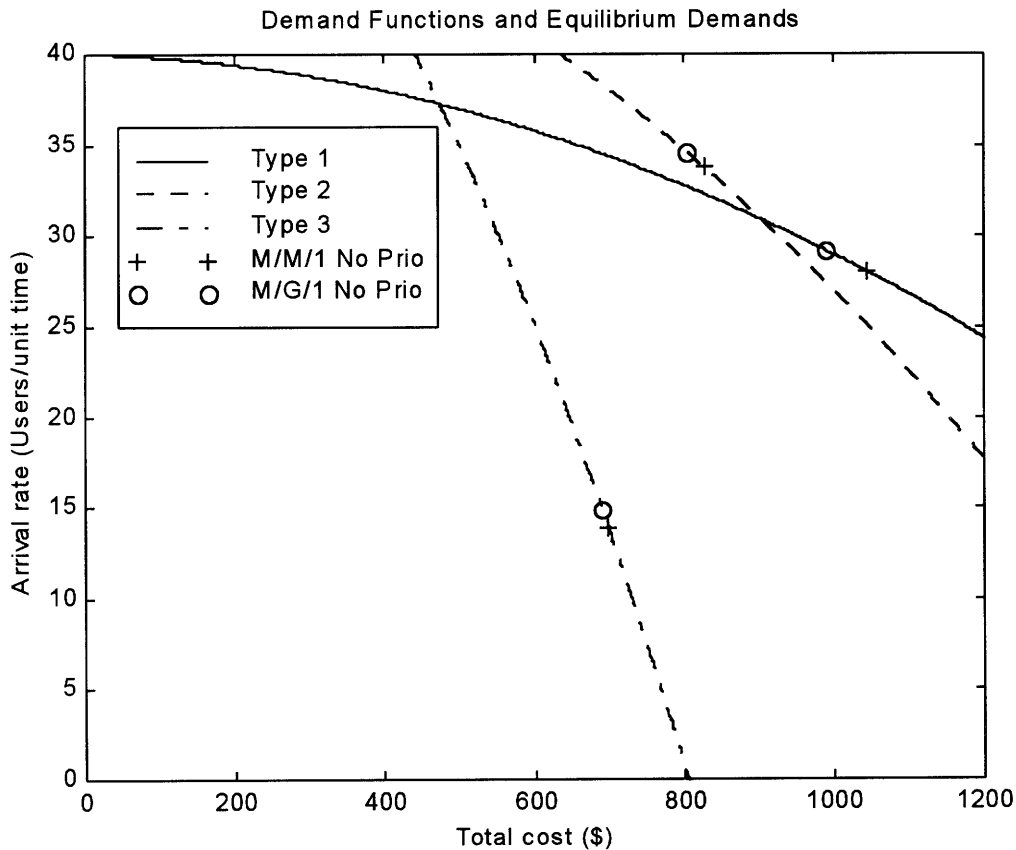


Figure 4-8: Comparison of M/M/1 and M/G/1 Models

4.2.3 Nonpreemptive Priorities versus No Priorities Comparison

This is an example using the same parameters as the previous example, but this time the effects of introducing nonpreemptive priorities are illustrated. Figure 4-9 shows how the equilibrium demand levels in the M/G/1 queues are affected by the introduction of priorities. In this specific example, the equilibrium demand levels of the Type 1 and Type 3 users increase noticeably, while the Type 2 demand level decreases slightly. In general, there does not seem to be a simple relationship governing which types of users benefit from priorities and which types do not. In appendix A, there is an example with only two types of users intended to illustrate the complexity of the problem.

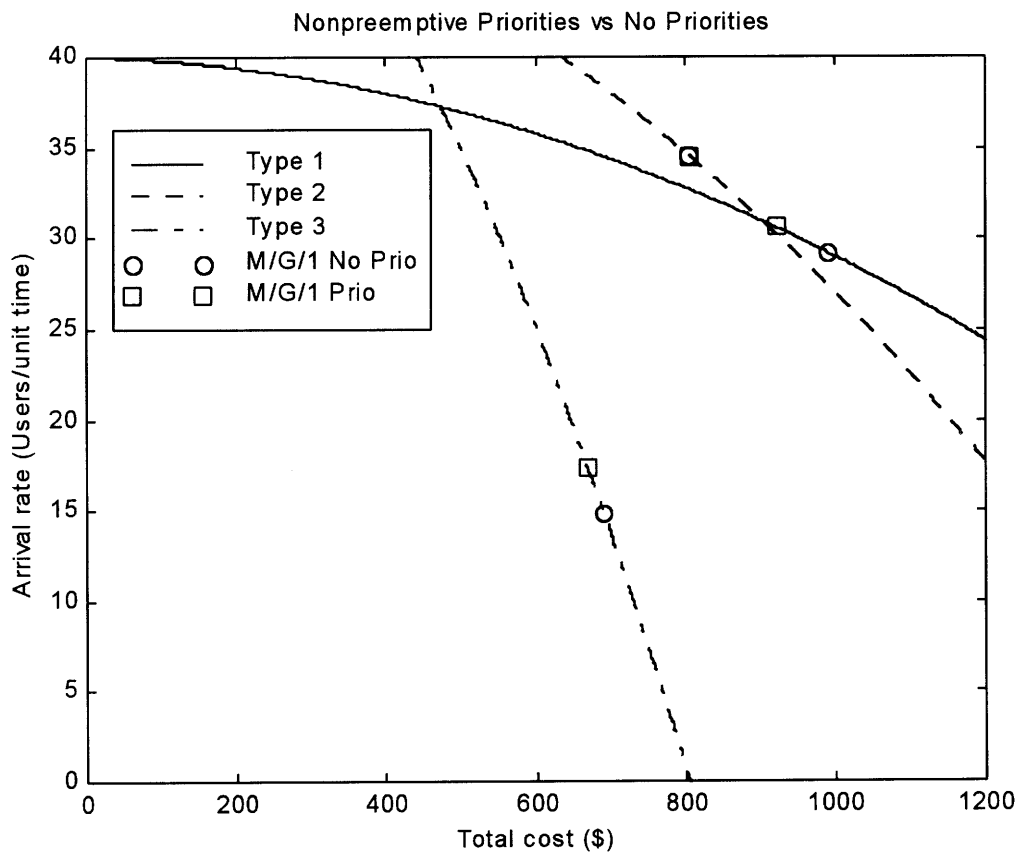


Figure 4-9: Nonpreemptive Priorities vs. No Priorities

4.2.4 Ordering of Types with Nonpreemptive Priorities, Inelastic Demand

It is a well-known result of queueing theory that the expected waiting time in a pure FCFS (First Come, First Served) system can be decreased using priorities [LARS 82]. This decrease can be achieved by giving priority to the users with the shortest expected service time. More generally, the optimal strategy to minimize the expected cost of waiting (internal cost) to all users is to always give priority to the user in the queue with the maximum ratio of (unit time cost)/(expected service time), $r_i = c_i / (1/\mu_i) = c_i * \mu_i$. An intuitive argument for this strategy is that it is desirable to maximize the rate at which expected cost leaves the system. The ratio, r_i , is proportional to the cost per unit time of having a user in the system, and inversely proportional to the time the server spends servicing that user. Thus, by ordering the types of users according to their cost/service time ratios (the higher the value, the higher the priority), the rate at which expected cost leaves the system is maximized.

This result is valid for minimizing the internal cost, but the question we are interested in is whether it can be extended to include the external cost. More specifically, is the strategy still optimal for minimizing the total cost? It turns out that the strategy does not necessarily minimize the external cost, and therefore the total cost is not minimized either. Table 4-4 shows the parameters for an example that illustrates how the total cost is not minimized by ordering the types according to their r_i ratios.

Type	Lambda (Arrival Rate)	Mu (Service Rate)	Sigma (Standard Deviation)	Cost of Delay (\$/hour)	Cost/Service Time Ratio (r_i)
H (1)	20	80	10	2000	160000
M (2)	30	60	10	2500	150000
L (3)	3	40	10	3000	120000

Table 4-4: Parameters for Ordering Example

The example has three types of users with fixed arrival rates (perfectly inelastic demand). The three types are denoted by H (high), M (medium), and L (low), respectively, corresponding to their relative cost/service time ratios. The internal and external costs were calculated for all six ordering permutations, and the results are shown in Table 4-5 and Figure 4-10. The order denotes what priority each type receives, e.g. LHM means that Type L (3) has the highest priority, followed by Type H (1) and Type M (2). It can be seen in Table 4-5 that the internal cost is minimized by order HML as expected, but that the external cost and the total cost are minimized by MHL.

Order	Internal Cost	External Cost	Total Cost
LHM	\$4,870	\$27,785	\$32,654
LMH	\$5,073	\$29,374	\$34,447
MLH	\$5,017	\$28,123	\$33,140
MHL	\$4,643	\$24,342	\$28,985
HML	\$4,528	\$24,594	\$29,122
HLM	\$4,851	\$27,618	\$32,469

Table 4-5: Equilibrium Costs with Different Priority Orders

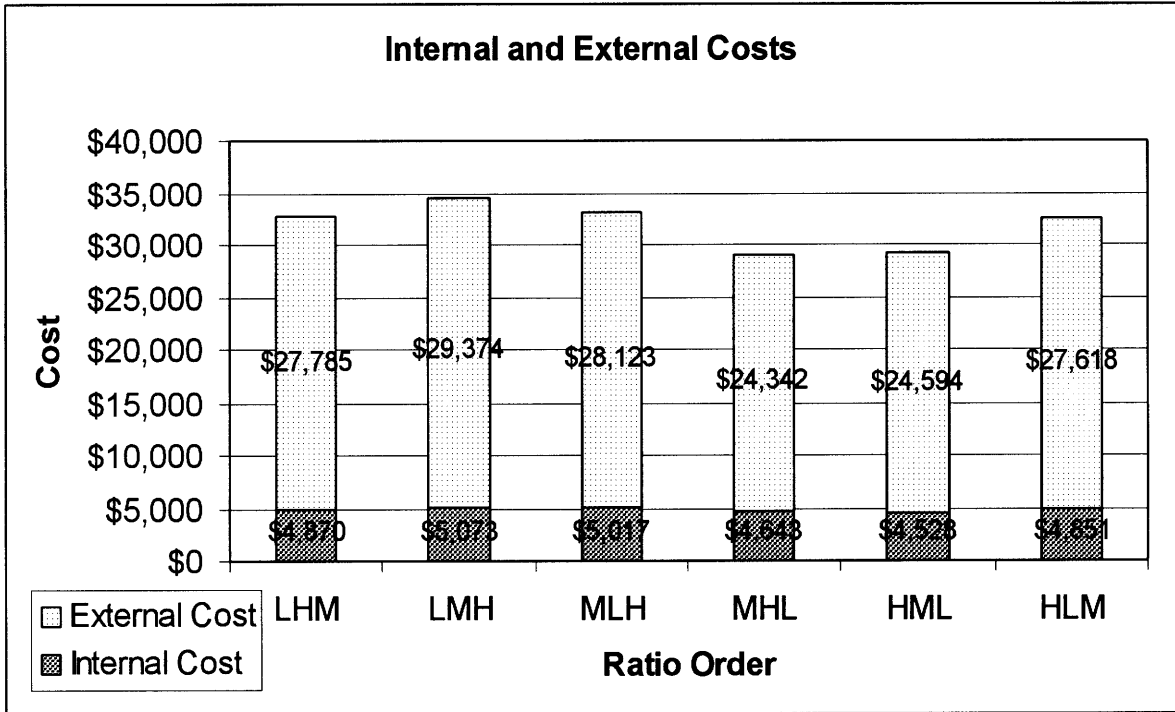


Figure 4-10: Internal and External Costs with Different Priority Orders

Chapter 5

Conclusions and Discussion

In this thesis, we have utilized Little's Law to formulate stochastic steady-state bottleneck models to calculate economically optimal equilibrium congestion fees. The four models use Little's Law to obtain closed form expressions for optimal steady-state congestion fees in M/M/1 and M/G/1 systems with and without nonpreemptive priorities. We have shown that the models can be used to calculate equilibrium congestion fees and demand levels for any monotonically decreasing demand function. Furthermore, we have given examples that emphasize the characteristics of the models, and the key differences between them. We have demonstrated how marginal cost congestion pricing can decrease delays significantly in a socially optimal way with only a relatively low decrease in utilization of the facility. In one example, a 10% decrease in demand led to a more than 90% decrease in delay.

If the concept of marginal cost pricing has been shown to optimize social welfare, why has it not yet been universally accepted? One of the reasons is that, in reality, it is not quite as simple as above to calculate the marginal cost. For example, an airport is a dynamic system; demand varies throughout the day and service capacity varies as the weather changes, which makes the steady-state assumption questionable. The models are thus best suited for systems where demand stays approximately constant for a longer

period of time. Dividing up the day into periods and averaging demand and predicted service capacity (according to a weather forecast) over that period is not as trouble-free as it may at first seem for the following reasons:

- Utilization could temporarily be greater than the capacity of the airport, a condition for which no steady-state expressions exist.
- The system time-constant can be on the order of days when the airport is operating near capacity and thus the system will never get sufficiently close to steady-state for the assumption to be reasonable.
- The delay is a highly non-linear and concave function of the demand, so averaging the demand leads to an inherent underestimation of the delay.
- Different time periods are not independent since it is likely that there is a queue left from a peak period into the next time period.

Another reason why marginal cost pricing has not been more widely implemented in the air transport industry, is that there is a strong resistance to change in many parts of the aviation community. For instance, one effect of marginal cost pricing would doubtlessly be that small private aircraft would experience such a tremendous hike in landing fees that they would no longer afford them. One could argue that these small planes should not use the runways during peak hours if the benefit they derive from such use is not so great that they are willing to compensate for the social cost that they cause due to increased delays. However, the fact remains that it is politically very difficult to change the present system.

In addition, if marginal cost pricing is introduced in a highly congested system, the fees would almost invariably be higher than they were before. This may lead to a

significant change in demand, which means that an accurate demand function is needed to predict where the equilibrium will be. In the air transport industry, for instance, it is very difficult to make an estimate of how airlines will react to changes in the landing fees. The demand function may even be discrete (e.g. an airline may decide to terminate its entire service to an airport), or highly dependent on network effects (e.g. if an airline terminates its Stockholm – New York route, it may terminate some of its NY - Boston flights as well). This makes changes in demand at a particular airport extremely unpredictable. Thus, it would be desirable to investigate the sensitivity of the models presented here to numerous variations of the demand functions.

This thesis raises several other questions for future research. It would be beneficial to extend the models to include multi-server queues that more closely resemble many real-world applications. A more rigorous method to determine when the steady-state assumption breaks down is also needed to ascertain under what conditions the use of our models is appropriate.

References

- [ARNO 93] Arnott, R., A. de Palma, and R. Lindsey. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand." American Economic Review 83 (1993): 161-179
- [CARL 70] Carlin, Alan and R. E. Park. "Marginal Cost Pricing of Airport Runway Capacity." American Economic Review 60 (1970): 310-318.
- [CH'N 83] Ch'ng, E. "Aeronautical Charges And The Pricing Of Runways." S.M. Thesis, MIT, 1983.
- [COBH 54] Cobham, A. "Priority Assignment in Waiting Line Problems." Operations Research 2 (1952): 70-76; correction *ibid.* 3:547
- [DANI 95] Daniel, Joseph I. "Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues." Econometrica 63.2 (1995): 327-370
- [ECKE 72] Eckert, Ross D. Airports and Congestion: A Problem of Misplaced Subsidies. Washington D.C.: American Enterprise Institute for Public Policy Research, 1972.
- [GROS 85] Gross, Donald, and Carl M. Harris. Fundamentals Of Queueing Theory. 2nd ed. New York: Wiley, 1985.
- [HEND 85] Henderson, J. V. Economic Theory and the Cities. 2nd ed. New York: Academic Press, 1985
- [KOOP 72] Koopman, B. "Air-Terminal Queues under Time-Dependent Conditions." Operations Research 20 (1972): 1089-1114
- [LARS 81] Larson, R. C., and Amadeo R. Odoni. Urban Operations Research. Englewood Cliffs, NJ: Prentice-Hall, 1981.
- [LITT 61] Little, J. D. C. "A Proof of the Queueing Formula $L = \lambda W$." Operations Research 9 (1961): 383-387

- [ODON 76] Odoni, Amedeo R. and Joseph Vittek. Airport Quotas and Peak Hour Pricing: Theory and Practice, U.S. Dept. of Transportation, Federal Aviation Administration, Rpt. FAA-AVP-77-5, May 1976.
- [PULL 72] Pulling, Ronald E. Economics of Capacity (Delay Costs) Peak-Load Pricing. Federal Aviation Administration, Draft Report, 1972.
- [REED 92] Reed, A. "Grandfather is Well and Living in Europe." Air Transport World 29 (1992): 65-67
- [ROSE 75] Rose, Kenneth, and G.B. Hamilton. "An Investigation of the Effects of Airport Landing Fees on Airport Congestion." Transportation Research Forum Proceedings 16.1 (1975): 175-186.
- [VICK 69] Vickrey, William. "Congestion Theory and Transport Investment." American Economic Review Proceedings 59 (1969): 251-260

Appendix A

Examination of What Types of Users Benefit From Priorities

A factorial design was set up to examine what factors determine which types of users benefit from priorities and which types do not. An M/M/1 queue was used with two types of users, and the cost of delay and the demand elasticity were varied. Table A-1 below shows how the parameter settings were changed as well as the resulting changes in demand. Each row is a performed run with and without nonpreemptive priorities. The second and third column show the cost of delay per unit time for Type 1 and Type 2 users, respectively. The next two columns indicated the relative demand elasticities. The last two columns show the change in demand when nonpreemptive priorities are introduced. A positive number indicates that the type of user benefited from priorities, (i.e. demand increased).

We were looking for relationships of the form: “Users with relatively high cost of delay and low demand elasticity benefit from being high priority users,” or simply “users with high demand elasticity are generally at a disadvantage when priorities are introduced.” Multiple linear regressions were performed but the results proved inconclusive. Unfortunately it seems that there is no clear relationship between the cost structure or the demand elasticities, and what types of users benefit from priorities.

RUN	Cost(1)	Cost(2)	Elast(1)	Elast(2)	Demand(1)	Demand(2)
1	1200	1200	High	High	0.3561	-0.2968
2	1200	500	High	High	2.0484	-0.0667
3	500	1200	High	High	-1.3395	-0.1821
4	500	500	High	High	0.1632	-0.1412
5	1200	1200	High	Low	0.0686	-0.0378
6	1200	500	High	Low	1.628	-0.1974
7	500	1200	High	Low	-1.1969	0.1728
8	500	500	High	Low	0.0319	-0.0197
9	1200	1200	Low	High	0.2404	-0.2186
10	1200	500	Low	High	0.6135	1.4697
11	500	1200	Low	High	-0.2405	-1.4446
12	500	500	Low	High	0.1013	-0.0942
13	1200	1200	Low	Low	0.1394	-0.1133
14	1200	500	Low	Low	1.2661	-0.0254
15	500	1200	Low	Low	-0.9472	-0.0351
16	500	500	Low	Low	0.061	-0.0557

Table A-1: Factorial Design Parameters