ChangiNOW: A Mobile Application for Efficient Taxi Allocation at Airports

by

Afian Anwar
B.A., Cornell University (2005)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of Master of Science in Transportation at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY September 2013

© Massachusetts Institute of Technology 2013. All rights reserved.

Author

Department of Civil and Environmental Engineering

August 10, 2013

Certified by

Daniela Rus
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

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

Accepted by

Heidi Nepf
Chair, Departmental Committee for Graduate Students
ChangiNOW: A Mobile Application for Efficient Taxi Allocation at Airports

by

Afian Anwar

B.A., Cornell University (2005)

Submitted to the Department of Civil and Environmental Engineering on August 10, 2013, in partial fulfillment of the requirements for the degree of Master of Science in Transportation

Abstract

The important role that taxis play in bringing passengers from an airport terminal to their final destination is often overlooked in airport operations and design. Due to varying flight arrival patterns at different terminals, taxi drivers are often unsure which terminal they should queue at. In this thesis, we present ChangiNOW, a mobile app that uses a predictive queueing model to efficiently allocate taxis. The ChangiNOW system uses observed taxi and flight data at each of the four terminals of Singapore’s Changi Airport to estimate the expected waiting time and queue length for taxis arriving at these terminals, and then sends taxis to terminals where waiting time is shortest. The app communicates this information to taxi drivers in a visually intuitive and appealing way, motivating them to service those terminals with the highest taxi demand. We present the theoretical details that underpin our prediction engine and validate our theory with several targeted numerical simulations. Finally, we evaluate the performance of this system in large-scale experiments and show that our system achieves a significant improvement in both passenger and taxi waiting time.

Thesis Supervisor: Daniela Rus
Title: Professor of Electrical Engineering and Computer Science

Thesis Supervisor: Amedeo Odoni
Title: Professor of Aeronautics and Astronautics
Professor of Civil and Environmental Engineering
Acknowledgments

I am tremendously grateful to my advisors Daniela Rus and Amedeo Odoni. Without their advice, guidance and encouragement this thesis would not be possible. I am very lucky to have met them.

I am also very privileged to have been admitted to the Masters of Science in Transportation (MST) program of the Department of Civil and Environmental Engineering (Course 1). The MST is an amazing interdisciplinary program and exposed me to many tools and concepts I use almost every day. I am particularly in debt to the instructors of 1.200 (Transportation Systems Analysis: Performance and Optimization), a class that gave me the necessary academic background to understand and apply operations research principles and methods to my research. Kris Kipp, the graduate administrator at Course 1 was an incredible resource and helped me navigate the administrative mazes at MIT.

My lab mates at the Distributed Robotics Lab of the Computer Science and Artificial Intelligence Laboratory (CSAIL) made graduate student life enjoyable and I will look back at my stint in CSAIL with great fondness. It never ceases to amaze me how MIT produces such brilliant and accomplished people that are down to earth and always willing to lend their brain power to solve difficult problems, research or otherwise. In particular, I would like to thank Mikhail Volkov my coauthor in the ChangiNOW research paper for the brilliant ideas and deep mathematical expertise he shared with me. Without him, the theory sections of this thesis would be incomplete.

I would also like to thank members of the Singapore-MIT Alliance for Research and Technology (SMART) for funding my research, providing access to data and supporting my research projects in Singapore. Howard Califano from the SMART Innovation Center believed in me and helped initiate the commercialization of my research with two large proof of concept grants.

Lastly I would like to thank my fiancée, Cheryl Ann, for saying yes.
## Contents

1 **Introduction** 10  
1.1 Motivations and Goals 10  
1.2 Contributions 13  
1.3 Relation to Previous Work 14  
1.4 Thesis Organization 16  

2 **Related Work** 17  
2.1 Double Ended Queueing 17  
2.2 Mobility on Demand 18  
2.3 Taxi System Optimization 19  

3 **Problem Setup** 22  
3.1 Service Model 22  
3.2 Assumptions 23  

4 **Data** 26  
4.1 Taxi Data Analysis 26  
4.2 Estimating Passenger Arrivals 27  
4.3 Changi Airport Case Study 29  

5 **Queueing Model and Prediction System** 33  
5.1 Non Equilibrium Queueing Model 33  
5.1.1 Is the queue expected to be free? 34  
5.2 Bounds and Guarantees 35  

6
5.2.1 How sure are we? .................................. 35
5.2.2 What is the waiting time? .......................... 36
5.3 Behavioral Parameters ................................. 37

6 Experiments and Results ............................... 39
6.1 Simulations for a Single Terminal .................. 40
  6.1.1 Entry Simulation (Case 3) ......................... 42
  6.1.2 Waiting Time Simulations ......................... 43
  6.1.3 Maximum Waiting Time and \( \alpha \)-certainty Simulations 44
6.2 Multi-Terminal Simulation ............................ 45

7 Conclusions and Future Work .......................... 48
  7.1 Application Design ................................ 49
  7.2 Conclusions ....................................... 51
List of Figures

1-1 A typical scene at Changi Airport. Taxi drivers are motivated to pick up passengers from the airport because they receive an extra fee. However, this often results in an overabundance of taxis. .......................... 11

1-2 Electronic signboard on the highway leading to Changi Airport showing the number of taxis at each terminal, along with the number of flights arriving in the next half hour. ................................. 12

3-1 Stages of the ChangiNOW service model: (1) taxi makes query, (2) server performs calculations, (3) server responds to taxi with optimal suggestion, (4) taxi makes acknowledgment, (5) server updates information. .................................................. 24

4-1 Bounding box representing the terminal taxi queueing area. Each red (BUSY) or green (FREE) circle represents a taxi’s state as it waited in the queueing area. ............................................. 27

4-2 Estimating derived taxi demand $u(t)$ from passenger arrival function $\lambda_{flight}(t)$. ................................................................. 28

4-3 Mean taxi waiting times for all terminals at Changi Airport over the course of a weekday. Different colored bars represent average waiting times at different terminals. The shortest waiting times can be observed at about 5 pm. .................................................. 29

4-4 Taxi arrivals at Changi Airport Terminal 3 over the course of a weekday. 30

4-5 Taxi departures at Changi Airport Terminal 3 over the course of a weekday. ................................................................. 31
4-6 Taxi queue lengths at Changi Airport Terminal 3 over the course of a weekday. .................................................. 32
4-7 Mean taxi waiting times at Changi Airport Terminal 3 over the course of a weekday. .................................................. 32
6-1 This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $L_v(t) < L_{\text{max}}$ (Case 1), all the taxis are guaranteed to enter the queue. .................................................. 41
6-2 This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $E[L_q] \gg L_{\text{max}}$ (Case 2), taxis are almost certain to be rejected from the queue. .................................................. 41
6-3 This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $E[L_q] \approx L_{\text{max}}$ (Case 3), some taxis are able to enter, while others are rejected from the queue. .................................................. 42
6-4 This graph highlights the area of uncertainty (middle section between the vertical dashed lines) when $0 < \Pr[\text{taxi entered the queue}] < 1$ due to $E[L_q] \approx L_{\text{max}}$. The plot shows the expected queue length on the x-axis against the probability of a taxi entering the queue on the y-axis. The vertical dashed lines indicate the certainty (either 0 or 1) cutoff at an accuracy of 3 decimal places. .................................................. 43
6-5 Comparison of taxi waiting times under Observed and Smart Rebalancing policies. .................................................. 45
6-6 Comparison of passenger waiting times under Observed and Smart Rebalancing policies. .................................................. 46
7-1 Mock up of the ChangiNOW app showing real time (left) and historical (right) modes. .................................................. 50
Chapter 1

Introduction

Queues are pervasive in everyday life, particularly those involving transportation. We see them when vehicles stop at a traffic light, when congestion builds up on a busy highway, or when people wait in line to board a taxi. Quantifying the dynamics of queues therefore has important and broad applications not just for transportation, but also in a variety of diverse fields including computer science, telecommunications and the optimization of factories, shops and hospitals [12]. Double ended queues such as those found at taxi stands, are especially important because they appear frequently in parallel processing, database concurrency control, flexible manufacturing systems and communication protocols [22]. This thesis aims to build a queueing model that accurately predicts the observed performance metrics of taxis queuing at Singapore's Changi Airport, and use it as part of a system that efficiently allocates taxis across the airports terminals.

1.1 Motivations and Goals

Singapore is an island city-state in South East Asia, home to more than 5 million people. While most commuters use public transit or drive, about 10% of commuters use taxis on a regular basis [3] because it is both convenient and inexpensive. The city is well served by a fleet of about 25,000 taxis but like any mobility on demand system, there are times where there are too many taxis and no passengers and vice
versa. Nowhere is this more apparent (Figure 1-1) than at Changi International Airport, where taxi drivers, incentivised by surcharges, often wait for hours to pick up recently arrived passengers.

Changi International Airport is the main point of disembarkation for tourists arriving in Singapore and serves more than 100 airlines operating 6,100 weekly flights to some 210 cities worldwide [11]. The airport has four terminals - Terminal One, Two, Three and a Budget Terminal¹. In total, Changi Airport handles more than 50 million passengers annually, making it the 18th busiest airport worldwide by passenger traffic [2]. Each terminal has one taxi queue of fixed capacity, where taxis wait in line to pick up passengers leaving the terminal. Although public transit options are

¹The Budget Terminal has since been closed to make way for a new Terminal Four in 2017
The main method by which travelers get to and from the airport is by taxi. However, like any mobility on demand system, there are times where there are too many taxis and no passengers and vice versa. When too many taxis wait at the airport, it reduces the number of taxis available to service the rest of the city and reduces the income of taxi drivers waiting in queue because they could be more productively finding fares elsewhere. When too few taxis are available, this results in travelers having to wait in line for long periods of time.

Changi Airport has tried to address this problem by putting up roadside electronic signboards just outside the airport that show the number of flights arriving at each terminal in the next hour, together with the number of taxis in queue (Figure 1-2). But this does not tell the taxi driver what he really wants to know - how long he
would have to ultimately wait at a certain terminal to pick up a passenger. Ideally, this information should be provided to the driver before time is invested to get to the airport, so that he can decide if it is worthwhile for him to head to the airport or not.

Instead of relying on roadside signage, we propose ChangiNOW, a mobile application that uses real time flight and taxi arrival information to

- Predict the expected waiting times at each terminal, and
- Direct taxis to the airport when these waiting times are short.

Essentially, we want to create a system that sets an upper bound on a taxis waiting time while ensuring that all the passengers that arrive at Changi Airport find a taxi waiting for them. ChangiNOW uses recorded flight passenger manifest data to estimate taxi demand. We then cross reference this with the GPS logs of some 15000 taxis in Singapore to build a queueing model that explains the interaction between recently disembarked passengers and the taxis that serve them. We test this model in simulation and show that might reduce taxi waiting times by one half and passenger waiting times by one third.

1.2 Contributions

The main contributions of this thesis are:

- **Data mining algorithms to find taxi queueing metrics at any given terminal.** In order to build our queuing model and predict how time varying passenger and taxi arrival rates affect taxi waiting time, we need to analyze our data and observe the interaction between passengers arriving at the airport and the taxis that pick them up. We provide data mining algorithms to easily derive the taxi arrival rate, departure rate, queue length and average waiting time at each terminal, and show how these metrics vary over time on a given weekday.
• A queuing model and an automated planning system that can be used to send taxis to an airport terminal when demand is high. Traditional queueing models applied to the taxi queuing problem have emphasized steady state solutions. In reality, taxi queues are seldom in equilibrium, and such steady state metrics do not provide realistic approximations of actual system performance. To address this, our ChangiNOW system uses real time flight and taxi arrival information to build a novel queueing model that takes into account the rate at which taxis arrive and depart as well as the potential demand coming from people that have just landed at the terminal to predict how long a specific taxi will have to wait at each terminal at the airport. We extend well known results in queueing theory to prove the correctness of our model and validate the results in targeted simulations.

• A direct comparison between simulated taxi and passenger waiting times in the current system versus one that uses ChangiNOW. In the last part of this thesis, we implement the ChangiNOW system in simulation and demonstrate how a system in which every taxi driver uses the ChangiNOW app and heads to the terminal with the shortest taxi waiting time is able to effect a 51% improvement in taxi waiting time and a 31% improvement in passenger waiting time.

1.3 Relation to Previous Work

An important aim of this thesis is to show how the taxi allocation problem can be formally modeled and solved using queueing theory. This section elaborates on this claim, and highlights our motivation for improving existing state of the art. Chapter 2 provides a comprehensive survey of related work.

Our problem of allocating taxis efficiently across Changi Airport's four terminals can be viewed as two subproblems. The first is a queuing problem - how do we find the expected waiting times and queue lengths of taxis in a double-ended queueing system. This problem was first posed by Kendall in [15]. Previous work [16, 13, 23,
have emphasized obtaining steady state solutions. However, in many real world applications, such steady state measures of system performance are not realistic for systems that are essentially non equilibrium or in situations where the system operates up to some specified time [7].

The second is one of rebalancing, where we view terminals at Changi Airport as nodes and taxis as autonomous robots in a networked, mobility on demand system [5, 19]. Most proposed solutions to this problem involve minimizing some cost function subject to performance constraints. For example, [20] developed a provably optimal rebalancing policy for a set of 50 randomly distributed nodes, that minimized the number of empty vehicle (rebalancing) trips while guaranteeing service levels.

Unlike [20], we do not aim to minimize the number of rebalancing trips. The cost of sending an empty taxi from one terminal to another is small and can be safely disregarded because the terminals are near one another. Instead, we are trying to reduce the amount of time each taxi driver spends waiting for passengers. Our research is motivated by concern that taxi drivers, encouraged by airport pickup surcharges are not only spending too much time at the airport, but are also waiting in queue at the wrong terminals. Secondly, our queuing model, elaborated in 3.1, is more realistic because it allows for taxi and passenger arrival rates to vary over the course of the day. More generally, there has been significant interest in using real time and historical data to optimize taxi operations. In [24], real time taxi trajectories were used to monitor taxi availability at taxi stands in Singapore while [14] visualized the real time spatial distribution of available taxis in Wu Han, China. Similarly, [25] introduced a recommendation system that directs taxi drivers in Beijing to zones of high taxi demand, thereby increasing the likelihood that they find a passenger quickly.

Rather than attempting to match taxi demand and supply within a city, Changi-NOW tries to solve the specific problem of directing taxis to a terminal at Changi Airport when demand at that terminal is high. Traditional systems use hot spot analysis to generate density maps that show how popular pick up and drop off “transactions” within the city vary by time of day. In our case, such standard methods fail because there is only one designated taxi stand per terminal at Changi Airport. Sending a
taxi to a specific terminal when “transaction” volume is high may not be optimal if many taxis are ahead of it in queue (Figure 1-1).

1.4 Thesis Organization

This thesis is organized into five chapters. Chapter 2 provides a comprehensive survey of related work. Chapter 3 introduces the problem setup, defines notation and states assumptions. We describe the data we use for this study in Chapter 4. In Chapter 5, we explain how we use arriving taxi and passenger information to predict how long each taxi will wait at an airport terminal and derive useful bounds and guarantees. Finally in Chapter 6, we use simulation to show how a system in which every taxi driver uses the ChangiNOW app and heads to the terminal with the shortest taxi waiting time is able to effect a 51% improvement in taxi waiting time and a 31% improvement in passenger waiting time.
Chapter 2

Related Work

This thesis builds on important prior work in queueing theory, mobility on demand and taxi fleet optimization.

Our problem of allocating taxis efficiently across Changi Airports four terminals can be viewed as two subproblems. The first is a queuing problem i.e. finding the expected waiting times and queue lengths of taxis in a system with two queues, one of taxis, the other of passengers where both taxis and passengers arrive randomly but depart only if there is a taxi or passenger waiting. The second is one of rebalancing, where we view terminals at Changi Airport as nodes and taxis as autonomous robots in a networked, mobility on demand system. This thesis presents solutions to both these problems with the aim of optimizing the distribution of taxis at an airports terminals.

2.1 Double Ended Queueing

Queueing theory is the mathematical study of waiting lines, or queues [4]. It is considered a branch of operations research that explores the relationship between demand on a service system and the delays suffered by the users of that system, and plays a central role in the analysis of and planning for urban services [16]. In its simplest form, customers arrive and are served by a single server according to a Poisson process.
A taxi queue is a type of double-ended queue where we can consider taxis waiting at a taxi stand as customers served by arriving passengers. This problem was first posed by Kendall in [15]. The analysis of such queues, particularly the expectation and frequency-distribution of waiting times, is important because it enables us to understand the relationship between taxi supply and the level of quality of service experienced by passengers arriving at the taxi stand [9].

Most of the literature [13, 23] have emphasized obtaining steady state solutions. However, in many real world applications, such steady state measures of system performance are not realistic for systems that are essentially non-equilibrium or in situations where the system operates up to some specified time [7]. This is particularly true for an airport terminal taxi queue which, due to daily flight schedules, experience fluctuating taxi and passenger arrivals over the course of the day. The queueing model we introduce in Chapter 3.1 provides a more realistic approximation of actual system performance because it allows for taxi and passenger arrival rates to vary.

2.2 Mobility on Demand

The mobility on demand (MOD) problem is well known and has been described in [5, 19]. In an MOD system, customers arrives at designated stations and are transported to others, either by driving themselves, or by being driven by an employed driver. In one-way mobility on demand systems such as a city bike sharing program, customers do not have to return to the same stations from which they picked up their vehicles. Due to the unidirectional nature of most daily commutes (trips originate from the city periphery and end in the central business district during morning peak period, and vice versa in the evening peak period), bike sharing stations become unbalanced [8].

A taxi service is a type of MOD system. Customers wait at designated stations, commonly known as taxi stands, for taxis to arrive and deliver them to their destination. In Singapore, particularly in the central business district or around train
stations, shopping centers, hospitals and airport terminals, taxis are only allowed to pickup and deliver passengers to designated taxi stands. As in any MOD system, taxis accumulate at popular destinations and deplete at less popular ones leading to an unbalanced system.

Most proposed solutions to the rebalancing problem involve minimizing some cost function subject to performance constraints. For example, [20] developed a provably optimal rebalancing policy for a set of 50 randomly distributed nodes, that minimized the number of empty vehicle (rebalancing) trips while guaranteeing service levels.

Unlike [20], we do not aim to minimize the number of rebalancing trips. The cost of sending an empty taxi from one terminal to another is small and can be safely disregarded because the terminals are near one another. Instead, our strategy is to reduce the amount of time each taxi driver spends waiting for passengers. Our research is motivated by concern that taxi drivers, encouraged by airport pickup surcharges are not only spending too much time at the airport, but are also waiting in queue at the wrong terminals.

2.3 Taxi System Optimization

In many cities, there are simply not enough taxis to meet peak demand. Several startups, most notably San Francisco based Uber, have tried to fix this problem by turning “black car limousine” livery services into taxis [17]. The Uber smartphone app allows limousine drivers to become taxi drivers during their down time by connecting them directly to nearby customers. Customers use Uber by creating an account, registering a credit card and searching for the nearest Uber taxi. Once a match is made, the customer waits for the taxi and can even see the taxi drive to the pickup location in real time. Billing is automated via the information provided and tip is included. Unlike a taxi company, Uber does not actually own cars or employ drivers. Its a booking service that takes a commission for facilitating the transaction.

In others, the taxi industry is informally organized and relies heavily on rudimentary radio dispatch systems. For example in Kuala Lumpur, a “one-to-many” broad-
cast system is used to match passengers with taxis. According to MyTaxi, a startup that recently introduced an app that allows customers to bypass inefficient and fragmented taxi booking systems, only one in four passenger booking requests using the traditional dispatch system are successfully fulfilled [18]. The MyTaxi smartphone app replaces the driver’s existing Citizen Band (CB) radio as his primary dispatch tool. The app tracks the taxi’s location in real time and enables customers to quickly find nearby taxis without having to call multiple taxi companies. In Mexico City, Taxi Beat [1] works with both taxi companies and private limousine drivers to supplement traditional phone dispatch services. Similar to MyTaxi, customers can use the Taxi Beat iPhone app to select a taxi and driver that suits their needs, and watch in real time as the taxi makes its way to the caller. The latest version of their app enables social networking functions, allowing one to share ones route and rate ones driver.

In contrast to strategies that aim to increase taxi supply or improve the taxi booking experience, there has been significant interest in using real time and historical data to optimize taxi operations. The emergence of Big Data has revolutionized transportation science, allowing researchers to mine very large urban data sets to test analytical models and observe how people move within a city. Cabsense, a mobile app developed by Sense Networks, a New York City based startup, applies machine learning algorithms to tens of millions of observed pickup and drop-off points of New York City taxis to help users find the best street corners to catch a taxi at a given time [6]. Using similar methods [21] developed an inference engine to predict the number of vacant taxis in Lisbon, Portugal.

The practical nature of the problem has also attracted the attention of university and corporate research labs. Wei Wu and his colleagues at the A*STAR Institute for Infocomm Research mined real time taxi trajectories to monitor taxi availability at taxi stands in Singapore [24] while Yang Yue et al demonstrated an online system [25] to help people visualize the real time spatial distribution of available taxis in Wu Han, China. Similarly, Jing Yuan and colleagues from Microsoft Research Asia developed a recommendation system [14] that directs taxi drivers in Beijing to zones of high taxi demand, thereby increasing the likelihood that they find a passenger
quickly.

Rather than attempting to match taxi demand and supply within a city, Changi-
NOW tries to solve the specific problem of directing taxis to a terminal at Changi
Airport when demand at that terminal is high. Each of the systems described earlier
use some variation of hot spot analysis to generate density maps that show how pop-
ular pick up and drop off "transactions" within the city vary by time of day. In our
case, such standard methods fail because there is only one designated taxi stand per
terminal at Changi Airport. Sending a taxi to a specific terminal when "transaction"
volume there is high may not be optimal, particularly if there are many taxis already
in queue.
Chapter 3

Problem Setup

In this chapter we formulate the problem, define notation, state assumptions and propose an asynchronous service model for an end-user application that accurately predicts the expected waiting time for taxis queueing at the airport.

Suppose at time $t$ a taxi is heading to the airport. We predict how long its waiting time $w$ will be when he arrives at an airport terminal taxi queue $\tau$ minutes later. We explain how $w$ is derived, by considering an M/M/C, $C = 1$ queueing model where a single queue of taxis en route to Changi airport is being serviced by customers arriving at each terminal. We then count the number of taxis ahead of it in queue and estimate how long it will take all of these taxis ahead of him to find passengers.

3.1 Service Model

Let us consider a scenario where every taxi in Singapore has a smartphone with our ChangiNOW app installed (Figure 3-1). When a taxi driver loads the app, he sees a list of terminals with real time taxi queue lengths and the number of people that will arrive at the terminal in the next one hour. We now formally describe the ChangiNOW service model (Figure 3-1).

1. A taxi that plans to make a trip to Changi Airport that wants to know which terminal it should head to and how long it would need to wait simply uses the app to query our ChangiNOW server

22
2. The server checks the flight manifest for each incoming flight to find $\mu(t)$, the rate at which people arrive at the taxi stand. Since the number of arriving passengers that eventually take a taxi varies from flight to flight, e.g. passengers on long haul international flights being more likely to take a taxi than those on short haul regional flights, this function is necessarily an estimate. It also checks $L_{trans}(t)$, the number of taxis en route to each terminal that will arrive before the current requesting taxi does $\tau$ minutes later. This quantity is known because every taxi that heads to the airport needs to check in with our system.

3. The server processes the data and tells the taxi driver the predicted waiting time, the probability of entering the queue and a bounded estimate of the wait at the terminal with the shortest waiting time. If the taxi driver decides that the waiting time is short enough and decides to head to the airport.

4. He accepts the server’s recommendation and

5. His taxi is immediately added to $L_{trans}$ for the terminal he chose.

Because each transaction is atomic (i.e. the state of the queue is updated sequentially after each query to the ChangiNOW server), we only need to show that our system works for a taxi going to a single terminal in order to prove that it works for many taxis considering multiple terminals.

### 3.2 Assumptions

In this section, we describe the main assumptions that define the scope of the ChangiNOW prediction system.

We have data from by flight passenger manifests. This data tells us how many passengers arrived at a Changi Airport terminal at discrete times throughout the day. From this known flight arrival data, we interpolate the customer terminal arrival rate $\lambda_{term}(t)$. From the terminal arrival rate we then estimate the taxi customer arrival rate (service rate) $\mu(t)$. We note that $\mu(t)$ varies with time.
Figure 3-1: Stages of the ChangiNOW service model: (1) taxi makes query, (2) server performs calculations, (3) server responds to taxi with optimal suggestion, (4) taxi makes acknowledgment, (5) server updates information.

We have real-time taxi queue length $L_q(t)$ for each Changi Airport terminal. We also have known and fixed maximum taxi queue capacity $L_{max}$ as well as the estimated travel time to any given terminal $\tau$ from the GPS coordinates at time $t$ of a taxi that queried the ChangiNOW server.

Assumption 1 – Commitment: Taxis that utilize the ChangiNOW system are committed to go to the terminal to which they are assigned. This assumption implies that a taxi arrives at the terminal with probability 1. Note that this says nothing about whether the taxi actually enters the queue.

Assumption 2 – Order: Taxis do not overtake each other on the way to the terminal. This assumption implies that all the taxis that are in transit and ahead of the querying taxi eventually make it into the queue before the querying taxi. Note that if these taxis do not enter the queue because the queue is full, then this can only work in favor of the querying taxi, never against, since as a result there can now only be fewer taxis in the queue in front of it. For the purposes of deriving strong results
in our analysis, we assume that all taxis in front of the querying taxi will actually join the queue.

We need to assume both commitment and order because our estimate of a taxi’s wait time \( w \) is a function of how many taxis arrive before him in queue. If we relaxed either of these constraints (i.e. taxis are allowed to renege and leave the queue, or overtake each other), then our prediction for \( w \) cannot hold. Both assumptions allow us to be absolutely certain of how many taxis are heading to each terminal at the airport and so we can do away with the notion of a taxi arrival rate \( \lambda \).
Chapter 4

Data

Our queuing model described in Section 3 uses two pieces of data as input. 1) The rate of arriving taxis at each terminal and 2) the number of passengers that arrive at each terminal’s taxi stand. In the simulation that we have developed, we obtain the first from the ChangiNOW system when taxi drivers indicate their intention to head to the airport and the second from historical flight arrival data. Our dataset consists of one month of taxi journeys in Singapore. The dataset we used contains millions of taxi records, where each record contains the time-stamp, GPS coordinates, driver number, etc. as well as the operational status of the taxi. Records are logged at short intervals and allow us to track taxi journeys over the course of the month. The flight manifest data provides us with the flight id, the number of passengers arriving on each flight and the actual time the flight landed. By cross-referencing the flight ids with airline schedule data available online, we were able to determine the terminal at which the flight landed.

4.1 Taxi Data Analysis

To extract taxi trips that were made by taxis picking up passengers at the Changi Airport, we first define a Bounding Box $B_T$ composed of vertices $b_1, b_2, \ldots, b_n$ that represent the physical queuing area at airport terminal $T$ (Figure 4-1).

Next, by examining raw taxi data, we select those taxis that passed through this
queueing area and find out when each taxi entered and left with a passenger. The operational status of a taxi lets us know if it is empty and looking for passengers (FREE) or occupied (BUSY). By measuring the entering and exit times of each taxi, we can easily derive the taxi arrival rate, departure rate, queue length and average waiting time at a particular terminal.

4.2 Estimating Passenger Arrivals

In this section we address how we estimate the unknown arrival rate of passengers to the taxi terminals using known flight arrival information from Changi Airport. We are given $\lambda_{flight}$, a time series from passenger flight manifests shared by the airport that tells us how many passengers arrive at each terminal in discrete 15 minute intervals (Figure 4-2). We assume that because of the remote location of the airport, taxi demand is driven entirely by arriving passengers.

The first challenge we encounter is that $\lambda_{flight}$ does not correspond to any given
discrete time interval. To overcome this, we smooth the time series $\lambda_{flight}$ using a $1 \times 5$ Gaussian filter. Using a 15-minute discretization this results in a one hour sliding window smoothing. We interpolate the smoothed data to yield an arrival rate $\lambda_{term}(t)$.

The second challenge is the difficulty in estimating the time from landing to arrival at a taxi stand. This depends on several factors including gate location, the number of available immigration counters and baggage delays. To realistically model this, we shift $\lambda_{term}(t)$ by some constant delay time $k$ minutes, to get $\lambda_{term}(t - k)$. From observed data we find that $k = 30$ to be a reasonable approximation for this delay.

Lastly, our data set does not differentiate between connecting passengers and those whose final destination is Singapore. Further, not all passengers will take a taxi. To account for this we scale $\lambda_{term}(t - k)$ by $f$, the ratio of the total number of people that arrived on flights to the number of taxis that departed the terminal over the course of the day, to obtain $\mu(t)$, the arrival rate of passengers to a taxi stand. The
final approximation for the customer arrival rate is given by

\[ \mu(t) = \int \lambda_{\text{term}}(t - k) \]

(4.1)

4.3 Changi Airport Case Study

The data lends support to our hypothesis that there is sometimes an imbalance of taxi supply across Changi Airports four terminals, most noticeably when the number of recently arrived flights is concentrated at a single terminal. When we observe how average taxi waiting times vary across terminals on a certain day (Figure 4-3), we see that taxis entering the queue at 9 am at Terminals B, 1 and 2 had to wait an average of 67, 36 and 26 minutes respectively to pick up a passenger while taxis at Terminal 3 only had to wait 2 minutes on average, no doubt helped by the some 1500 passengers that landed at the terminal the hour before.

Avg Waiting Times of Taxis By Time of Day on Tuesday 1 Aug 2010

Figure 4-3: Mean taxi waiting times for all terminals at Changi Airport over the course of a weekday. Different colored bars represent average waiting times at different terminals. The shortest waiting times can be observed at about 5 pm.
Number of Taxis Arriving at Terminal Three on Tuesday 1 Aug 2010

Taxis
Passengers

Figure 4-4: Taxi arrivals at Changi Airport Terminal 3 over the course of a weekday.

When we look at how average waiting times at a certain terminal vary over the course of the day, we notice that there is a lot of waste. For example, at Terminal 3 (Figure 4-7), we see that some taxis entering the queue in the early hours of the morning have to wait as long as 100 minutes before they find a passenger, with average waiting times of more than an hour not being uncommon.

Examining the queue length (Figure 4-6) and rate at which taxis enter and depart the queue at each terminal over the course of the day gives us insight into how passenger demand for taxis varies over time. Take Terminal 3, Changi’s newest terminal and home base for many of Singapore Airline’s long distance international flights. Many of these flights arrive around 6 am in the morning, just in time for business travelers to begin their day. Taxi drivers are aware of this, and so the number of taxis waiting for passengers at Terminal 3 steadily increases and reaches a peak of 86 at around 8 am. After that however, the supply of new taxis slows while passenger demand for taxis holds steady and within an hour the queue of taxis drops to zero. By looking at the plots of taxis entering (Figure 4-4) and leaving (Figure 4-5) the queue, we
can deduce that taxi demand has not abated since taxis enter and immediately leave with a passenger. We can therefore infer that during this period there exists a steady queue of frustrated people waiting for taxis.

In this section we examined ground truth taxi queueing statistics at Changi Airport and found occasional but routine imbalances in taxi supply. Changi Airport is about to start construction on a new Terminal Four to replace the aging Budget Terminal and has just announced plans [10] to build a fifth terminal, Terminal Five, within the next decade in order to meet growing demand for aviation in Asia. This will raise the airport's capacity to 85 million passengers a year, up from 50 million currently. With this increase in passenger arrivals, demand for taxis at the airport can only grow and given the inelastic supply of taxis in Singapore, there will be a pressing need to allocate taxi resources more efficiently. In the next Chapter, we explain how this can be done by sending taxis to the terminals that need them the most.
Taxi Queue Lengths at Terminal Three on Tuesday 1 Aug 2010

Figure 4-6: Taxi queue lengths at Changi Airport Terminal 3 over the course of a weekday.

Mean Taxi Waiting Times at Terminal Three on Tuesday 1 Aug 2010

Figure 4-7: Mean taxi waiting times at Changi Airport Terminal 3 over the course of a weekday.
Chapter 5

Queueing Model and Prediction System

In this section, we present the mathematical model underlying ChangiNOW, and show how we extend well known results from Queuing Theory to efficiently allocate taxis among Changi Airports four terminals. In order for the ChangiNOW system to efficiently direct taxis to the terminal with the shortest waiting time, it needs to know for each querying taxi and each terminal:

(i) Which queues are expected to have available space (i.e. free) by the time the taxi arrives at the airport.

(ii) The probability with which a queue that is expected to be free will actually be free.

(iii) The expected waiting time for a taxi if it enters the queue.

5.1 Non Equilibrium Queueing Model

Traditional queuing models give the expected queue lengths and wait times from the steady state probability distribution of a continuous time Markov chain. This assumes that the arrival and departure rates of the queue are constant and that sufficient time
has passed for the queueing system to reach equilibrium. However in our case, the arrival and departure rates of taxis change quickly over the course of the day so the taxi queue does not get a chance to reach equilibrium and standard methods are thus inadequate.

Instead, let us assume that at each terminals taxi stand, taxis are serviced in the order in which they arrive by arriving passengers, one at a time, i.e. the queue discipline is first-come-first-served with one server. A taxi some distance away from the airport makes a request to the ChangiNOW server at time $t$. We know the queue length $L_q(t)$ at each terminal, and we know the number of taxis $L_{\text{trans}}(t)$ that are in transit to each terminal. Further, we know the maximum queue capacity $L_{\text{max}}$ and an estimate of the travel time $\tau$ to each terminal, as described in Section 3.1.

Assumption 1 tells us that if a taxi is in transit to the terminal, then it is guaranteed to arrive at the terminal and join the taxi queue. Assumption 2 tells us that all taxis that are in transit are guaranteed to arrive before the taxi that is making the query. Thus by Assumptions 1 and 2, we know that $L_{\text{trans}}(t)$ taxis will join the queue at the terminal by time $t + \tau$. We define the virtual queue $L_v(t)$ at a terminal at time $t$ to be projection of all the current taxis in transit onto the real taxi queue at the terminal, given by

$$L_v(t) = L_q(t) + L_{\text{trans}}(t)$$  \hspace{1cm} (5.1)

Note that although the length of the actual taxi queue $L_q(t)$ must at all times not exceed the maximum queue capacity, there is no such constraint on the size of the virtual queue $L_v(t)$. The virtual queue is essentially a projection to the size of the real queue to that time when the querying taxi arrives at the terminal.

5.1.1 Is the queue expected to be free?

Before deciding which terminal the taxi is to be deployed to, we must ensure that there will be space in the taxi queue.

By Assumptions 1 and 2, at estimated time of arrival $t + \tau$, $L_{\text{trans}}(t)$ taxis will join
the queue at back of the terminal. Meanwhile, a number of taxis will leave the queue with a passenger, according to the service rate \( \mu(t) \) over the time interval \([t, t + \tau]\). If we define \( \bar{\mu}_\tau \) as the average service rate over this time interval, given by

\[
\bar{\mu}_\tau = \frac{1}{\tau} \int_t^{t+\tau} \mu(x) \, dx
\]  

then we can say \( \tau \bar{\mu}_\tau \) taxis are expected to leave the taxi queue by time \( t + \tau \). Thus, the taxi queue \( L_q(t + \tau) \) will grow by \( L_{\text{trans}}(t) \) and is expected to shrink by \( \tau \bar{\mu}_\tau \). We define the expected queue length at time \( t + \tau \) as \( E[L_q] \), given by

\[
E[L_q] = L_q(t) + L_{\text{trans}}(t) - \tau \bar{\mu}_\tau
= L_v(t) - \tau \bar{\mu}_\tau
\]  

This gives us a quantitative statement for our first result.

**Theorem 1.** The queue is expected to be free if and only if \( E[L_q] < L_{\text{max}} \).

The proof is simply the formal statement of the definitions above.

5.2 Bounds and Guarantees

5.2.1 How sure are we?

Note, that since \( \mu(x) \) is the rate parameter for a Poisson process, we can compute the expected number of taxis that will leave the queue over any time period. Often we can satisfy ourselves with expected value results, but sometimes these results are inadequate.

Consider the following 3 cases for a terminal queue with any reasonable bounded service rate \( \mu(t) \).

(i) \( L_v(t) < L_{\text{max}} \): This implies \( E[L_q] < L_{\text{max}} \), since \( E[L_q] = L_v(t) - \tau \bar{\mu}_\tau \) and \( \tau \bar{\mu}_\tau \geq 0 \). Thus we expect the queue to be free, and in-fact it will be free with
probability 1, since by Assumption 2 there is no possibility of any other taxis overtaking the querying taxi.

(ii) $E[L_q] \gg L_{\text{max}}$: With many taxis in transit, we are almost sure there will be no space in the queue. We are not completely certain, because unlike case (1), the service rate is a Poisson process, but we are almost certain, to some $\varepsilon$ precision. Note that $L_v(t) \gg L_{\text{max}}$ does not necessarily imply that $E[L_q] \gg L_{\text{max}}$ since $\tau \mu_v$ may be large.

(iii) $E[L_q] \approx L_{\text{max}}$: This is the main case of interest. Depending on the service rate $\mu_v$ and our own specifications, our understanding of "approximately equal" will change. In this case, a binary quantitative result is not sufficient.

To afford taxi drivers the possibility to customize their ChangiNOW service, the driver specifies the minimum acceptable entry probability $\Pr[\text{entry}]$.

**Theorem 2.** The queue is expected to be free with probability $\Pr[\text{entry}] = \Pr[L_q(t + \tau) < L_{\text{max}}] = \int_t^{t+\tau} \mu_v e^{-\mu_v x} \left( \frac{L_v(t) - L_{\text{max}}}{L_v(t) - L_{\text{max}}} \right)^{L_v(t) - L_{\text{max}}} \, dx$. (5.4)

**Proof.** The probability that the queue will be free is equal to $\Pr[L_q(t + \tau) < L_{\text{max}}]$ (i.e., at least $L_q(t + \tau) - L_{\text{max}} + 1$ taxis will have left the terminal with a passenger during the time $\tau$).

5.2.2 What is the waiting time?

The other crucial parameter that determines a driver's decision to commit to the back of a taxi queue is how long he expects it will take for him to pick up a customer.

Define waiting time $W$ as the length of time from when a taxi enters the queue to when it leaves with a customer.

**Theorem 3.** The expected waiting time $E[W] =$
\[
\min W^* \text{ s.t. } \int_{t+\tau}^{t+\tau+W^*} \mu(x) \, dx \geq L_q(t + \tau).
\] (5.5)

Proof. Define the waiting time service rate \( \tilde{\mu}_W \) as the average service rate while the taxi is waiting in the queue, given by

\[
\tilde{\mu}_W = \mu^* \text{ s.t. } \mu^* = \frac{1}{W^*} \int_{t+\tau}^{t+\tau+W^*} \mu(x) \, dx.
\] (5.6)

Simplify using \( W^* = \frac{L_q(t+\tau)}{\mu^*} \) and solving for \( W^* \), first substituting \( W^* \):

\[
\frac{1}{L_q(t + \tau)} \int_{t+\tau}^{t+\tau+W^*} \mu(x) \, dx = 1
\]

and then multiplying across:

\[
\int_{t+\tau}^{t+\tau+W^*} \mu(x) \, dx = L_q(t + \tau)
\] (5.7)

i.e. the waiting time \( W^* \) must be such that (5.7) holds, implying that the taxi is serviced at time \( t + \tau + W^* \). All \( W > W^* \) are disregarded as the taxi is already serviced, thus the expected waiting time is the minimum \( W^* \) that satisfies (5.7), giving (5.5).

\[\square\]

5.3 Behavioral Parameters

The taxi makes a request at time \( t \) and the server predicts that the queue will be free with some probability and also provides an expected waiting time. So it it wise to commit to the terminal? In many cases, the decision will depend on the driver.

As well as being able to specify the entry probability \( \Pr[\text{entry}] \), we add a layer of flexibility to our model which accounts for the habits, preferences and attitudes of taxi drivers in response to the information provided by the ChangiNOW system. For example, a risk-taking but patient driver may commit to a terminal if he is 50%
certain to enter the queue, and he is also 50% certain that his waiting time will be under 30 minutes. On the other hand, a risk adverse and impatient driver may commit to the terminal only if he is 80% certain to enter the queue and 60% certain that his waiting time will be under 15 minutes.

To reflect such behavioral characteristics, we introduce two additional parameters. First, the taxi driver can specify a maximum acceptable waiting time $W_{\text{max}}$. Second, the taxi driver can specify a waiting time certainty margin $\alpha \in [0, 1]$. We define the $\alpha$-certainty waiting time $W_\alpha$ as a time such that a taxi driver entering the terminal at time $t + \tau$ will experience a wait of less than $W_\alpha$ with probability $\alpha$.

Theorem 4. The waiting time $W$ will be less than the maximum acceptable waiting time $W_{\text{max}}$ with probability $\Pr[W < W_{\text{max}}] =$

$$\int_0^{W_{\text{max}}} \mu_w e^{-\mu_w x} \frac{L_q(t+\tau)}{L_q(t + \tau)!} dx. \quad (5.8)$$

Theorem 5. The $\alpha$-certainty waiting time $W_\alpha =$

$$= \min W^* \text{ s.t. } \int_0^{W^*} \mu_w e^{-\mu_w x} \frac{L_q(t+\tau)}{L_q(t + \tau)!} dx \geq \alpha. \quad (5.9)$$

In (5.9) choose the smallest possible $W_{\text{max}}$ such that the probability computed through the integral is greater than $\alpha$. 

38
Chapter 6

Experiments and Results

In this section, we conduct several experiments using a decentralized, discrete event simulation environment in MATLAB. We run two kinds of experiments - individual terminal simulations and a large scale, multi-terminal airport simulation. The individual terminal simulations assess the validity of the analytical expressions presented in Chapter 5 by simulating hundreds of thousands of taxis entering a single terminal while the multi-terminal airport simulation tests the impact our ChangiNOW rebalancing policy has on both passenger and taxi waiting time by simulating hundreds of taxis in a multi-terminal environment over a 24 hour period. We believe that a good rebalancing policy will be able to improve both these metrics due a better matching of taxi supply and passenger demand. Verifying the correctness of the single terminal results before running the multi-terminal simulation serves as a sanity check and demonstrates the practical utility of the ChangiNOW system as a way of balancing real time taxi supply at the airport.

The simulation consists of taxis, passengers and terminals. Each terminal has a single taxi queue of fixed size where passenger arrivals are generated according to demand inferred from observed flight arrival data provided by Changi Airport (Chapter 4.3). Taxi data for the single terminal simulations is synthetic i.e. generated to meet specific needs of each experiment. In contrast, the multi-terminal simulations use historical taxi and passenger arrival data from a weekday in August 2010. We elaborate on our methodology in Chapter 6.2. As in the real world, taxis can choose
which terminal to serve and are able to switch terminals at any point in time. When a taxi picks up a passenger, the taxi is directed to a dummy node that represents the city center. For simplicity, traffic conditions in the network are treated as exogenous and are not modeled explicitly.

6.1 Simulations for a Single Terminal

In the first experiment, we verify what happens when a taxi makes a query to the ChangiNOW server to check if the queue at a particular terminal is free. Recall the 3 possible outcomes discussed in Chapter 5:

(i) The queue is certainly free \((L_v(t) < L_{max})\)

(ii) The queue is almost certainly full \((E[L_q] \gg L_{max})\)

(iii) The queue may or may not be free \((E[L_q] \approx L_{max})\)

In Figures 6-1, 6-2, 6-3 we plot time on the x-axis against the virtual queue length on the y-axis for 50 taxis using 3 different initial queue length conditions. For clarity, we use just 50 taxis but as shown in Chapter 6.1.1, our results hold for 100,000 taxis. The vertical dotted line indicates the taxi has reached the terminal after a constant travel time of \(\tau = 35\) minutes. The thick red horizontal line indicates the maximum capacity, \(L_{max}\), \((52\) taxis\) of the real queue. A green O indicates the taxi has entered the queue, and a red X indicates there it was rejected from the queue.

**Case 1: The queue is certainly free \((L_v(t) < L_{max})\)**

As indicated in 5.2.1, if the virtual queue length is less than the maximum queue capacity at the time of arrival, all taxis are guaranteed to enter the queue (Figure 6-1).

**Case 2: The queue is almost certainly full \((E[L_q] \gg L_{max})\)**

If the expected queue length at the time of arrival is much greater than the maximum queue length, the taxi is will almost certainly be unable to enter the queue (Figure 6-2).
Figure 6-1: This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $L_v(t) < L_{max}$ (Case 1), all the taxis are guaranteed to enter the queue.

Figure 6-2: This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $E[L_q] > L_{max}$ (Case 2), taxis are almost certain to be rejected from the queue.
Figure 6-3: This graph shows how the position of the taxi in the virtual queue (y-axis) varies over time (x-axis). When $E[L_q] \approx L_{max}$ (Case 3), some taxis are able to enter, while others are rejected from the queue.

**Case 3:** The queue may or may not be free ($E[L_q] \approx L_{max}$)

Figure 6-3 demonstrates why a simple expected queue length prediction is not enough. When $E[L_q] \approx L_{max}$, the number of taxis that entered the queue is split almost 50/50, so a definitive answer is not possible.

### 6.1.1 Entry Simulation (Case 3)

We consider Case 3 where $E[L_q] \approx L_{max}$ more closely. The terminal simulator was initialized with travel time $\tau = 35$ minutes, service rate $\mu(t) = 1.0$, and queue capacity $L_{max} = 35$. As in Figure 6-3, we vary $L_q$ and $L_{trans}$ so that $E[L_q]$ took values in the range [0, 70]. We plot $E[L_q]$ on the x-axis versus $Pr[entry]$ on the y-axis (Figure 6-4).

As expected, when $E[L_q] \ll L_{max}$ (Case 1), every taxi is able to enter the queue and so $Pr[entry] = 1$. As $E[L_q]$ approaches $L_{max}$, $0 < Pr[entry] < 1$ due to the stochastic nature of passenger arrivals at the front of the queue (Case 3). As we
Figure 6-4: This graph highlights the area of uncertainty (middle section in between the vertical dashed lines) when $0 < \text{Pr[taxi entered the queue]} < 1$ due to $E[L_q] \approx L_{max}$. The plot shows the expected queue length on the x-axis against the probability of a taxi entering the queue on the y-axis. The vertical dashed lines indicate the certainty (either 0 or 1) cutoff at an accuracy of 3 decimal places.

increase $E[L_q]$ past $L_{max}$, $\text{Pr[entry]}$ drops to 0 (Case 2).

We validate Theorem 2 in simulation by adjusting $L_q$ and $L_{trans}$ so that $\text{Pr[entry]} = 0.65$. The simulation results (100,000 runs) are as follows:

\[
\text{no. taxis entered} = \frac{65,154}{100,000} = 0.65
\]

6.1.2 Waiting Time Simulations

Again the terminal simulator was initialized with variable travel time $\tau = 35$ minutes and service rate $\mu(t)$. $L_Q$ and $L_{trans}$ were adjusted so that $E[L_Q]$ falls within the area
of uncertainty. The ChangiNOW server predictions are as follows:

\[
\begin{align*}
\Pr[\text{entry}] & \approx 0.76 \\
\text{avg. } E[W] & \approx 48 \text{min} \\
\text{avg. } \Pr[W < E[W]] & = 0.55
\end{align*}
\]

The simulation results (100,000 runs) are as follows:

\[
\begin{align*}
\text{no. taxis entered} & = 75,431/100,000 \\
\text{no. entered with } W < E[W] & = 41,234/75,431 = 0.55
\end{align*}
\]

### 6.1.3 Maximum Waiting Time and \( \alpha \)-certainty Simulations

The terminal simulator was initialized with variable travel time \( \tau \) and service rate \( \mu(t) \). Again, \( L_Q \) and \( L_{\text{trans}} \) were adjusted so that \( E[L_Q] \) falls within the area of uncertainty. We calibrate using both the maximum acceptable waiting time \( W_{\text{max}} \) and the certainty margin \( \alpha \). For the simulation, we designated two groups of drivers. Group A (risky) decide whether to accept the deployment based on the probability of \( W_{\text{max}} = 40 \text{ min} \). Group B (safe) decide whether to accept the deployment based on a 90% certainty waiting time (i.e. \( \alpha \)-certainty waiting time \( W_\alpha \) with \( \alpha = 0.9 \)). The ChangiNOW server predictions are as follows:

\[
\begin{align*}
\Pr[\text{entry}] & \approx 0.76 \\
\text{no. taxis entered} & = 75,431/100,000 \\
\text{Group A: avg. } \Pr[W < 40] & = 0.18 \\
\text{Group B: avg. } W_\alpha, \alpha = 0.9 & = 57 \text{ min}
\end{align*}
\]
Avg Waiting Times of Taxis By Time Of Day in Simulation

![Graph showing waiting times for taxis by time of day in simulation.](image)

**Figure 6-5**: Comparison of taxi waiting times under Observed and Smart Rebalancing policies.

Tested in simulation:

- no. Group A with $W < W_{max}$ = 13,695/75,431 = 0.18
- no. Group B with $W < W_{\alpha}$ = 70,243/75,431 = 0.93

## 6.2 Multi-Terminal Simulation

Having validated the correctness of the ChangiNOW system in Chapter 6.1 by showing that simulation results for a single terminal were consistent with that predicted by our queuing model and prediction engine in Chapter 5, we now add both complexity and scale by extending our simulation environment to include the entire Changi Airport terminal system and evaluate the ChangiNOW rebalancing policy against a naive approach.

Our simulation environment is initialized with 500 taxis, and 5 nodes, 4 represent-
Avg Waiting Times of Passengers By Time Of Day in Simulation

Figure 6-6: Comparison of passenger waiting times under Observed and Smart Rebalancing policies.

Passengers arrive stochastically at each terminal $i$ according to a time varying Poisson process with parameter $\mu_i(t)$. They are served by taxis arriving at rate $\lambda_{\text{taxi}_i}(t)$. Both $\mu_i(t)$ and $\lambda_{\text{taxi}_i}(t)$ are based on historical data. We chose to simulate 500 taxis because this was empirically sufficient to achieve stability and saw no significant changes in queuing behavior when this number was increased. We conducted experiments using two policies:

*Observed Policy*: $P_{\text{obs}}$ is based on empirical taxi data. We note the number of taxis entering the queue at each terminal for every 15 minute interval on Monday 2 August 2010, calculate the proportion of taxis entering terminal $i$ at time $t$ and smooth these values using a 1x5 Gaussian kernel in time. This gives us the rate $a_i(t)$, the ground truth queueing behavior of taxis visiting Changi Airport.

*Smart Rebalancing Policy*: In $P_{\text{smart}}$, taxis at each node $i$ (including the terminal nodes) query our ChangiNOW server, which returns an answer, $\text{DEST}_j$ that tells the
taxi where to go based on the projected waiting times each taxi would encounter and \( w_{\text{max}} \), the maximum amount of time each taxi is prepared to wait. If there are no better alternatives, our server returns \( \text{DEST}_j = i \), effectively telling the taxi to stay put (Figure 8).

We ran 5 simulations of 24 hours each. Each minute, the server updates the destination of each taxi. For \( P_{\text{obs}} \), destinations are based on historical patterns while for \( P_{\text{smart}} \), taxis are routed to the terminal with the shortest predicted waiting time.

For each policy, we plot the waiting time of taxis (Figure 6-5) and passengers (Figure 6-6) over the course of a simulation day. Each data point represents the the average waiting time of taxis and passengers that entered and left a terminal queue at each 3 hour interval.

Our results show that with the Smart Rebalancing Policy, we achieve a 51% improvement in taxi waiting time and a 31% improvement in passenger waiting time over the Observed Policy. Intuitively, we can explain the validity of our results by considering a simple example of an airport with two terminals, one with many taxis and no passengers and the other with many passengers and no taxis. With the Smart Rebalancing Policy, such situations are unlikely to persist because the ChangiNOW server would immediately send idle taxis from one terminal to pick up passengers from the other, thereby creating a better matching of taxi supply and demand so both taxis and passengers wait less. Our controlled experiments used simulated taxi and passenger arrival rates based on observed data. In actual implementation, we believe similar results can be achieved by using both real time taxi trajectories and ChangiNOW server requests in our queuing model. Passenger arrival information in both simulation and real world contexts would use known flight and passenger manifest data provided by the airport.
Chapter 7

Conclusions and Future Work

The rebalancing approach used in this thesis is promising. We developed a set of fully analytical expressions to predict performance measures for a double-ended, non stationary queue and demonstrated a system that uses our approach to match taxi supply with demand at Changi Airport.

There are a number of interesting future lines of research. Firstly, is our approach scalable? Can we generalize our queueing model and prediction engine from an airport terminal system with 4 independent taxi queues to the entire city of Singapore with hundreds of taxi queues? If so, how would we estimate demand from these taxi stands? Solving this problem has the potential to better match taxis to customers, thereby saving fuel, reducing the number of taxis required to serve the city and increasing the overall quality of service provided by the taxi system.

Secondly, could our approach be extended to airports and cities outside Singapore and if so, how would we model the behavior of taxi drivers and collect passenger demand data in these new markets? Would we need to adjust our queueing model to take into account differences in taxi fleet market structure and data collection?

Lastly, the double-ended queue neatly described by our taxi example shows up in a variety of applications and it would be interesting to explore if our methods for predicting waiting times and queue lengths apply e.g. in manufacturing and intelligent logistics contexts.
7.1 Application Design

A real world app will work slightly differently from the version introduced in Chapter 3. In actual implementation, the assumptions we made,

(a) Every taxi driver immediately follows the recommendation provided by the ChangiNOW app and is committed to go to the terminal to which they are assigned

(b) Every taxi is equipped with the ChangiNOW app

(c) Taxis arrive at the terminal in the order in which they are assigned

are not likely to hold.

Instead, the design of the ChangiNOW app will emphasize conveying taxi waiting times at Changi airports four terminals in a clear and visually intuitive way. Our app has two modes - real time and historical, both accessible through a bottom navigation bar.

In real time mode, the app regularly checks the ChangiNOW server for the latest flight arrival and passenger load information to estimate how many people will need a taxi at each terminal in the immediate future. Since not every taxi is equipped with the ChangiNOW app, it uses historical taxi arrival data to estimate how many taxis will arrive at each terminal. The app feeds both the passenger and taxi arrival estimates into our queueing model and locally computes taxi wait time predictions and guarantees, which are then displayed prominently on the main screen (Figure 7-1).

If a taxi driver decides to head to a particular terminal, he can select that terminal to get shortest path driving directions and real time traffic information. The app notifies the server that a new taxi is headed to the terminal and adds it to the "virtual queue" of taxis headed there. As explained earlier, the taxis position in the queue is based on its location (acquired through the smartphones GPS) and estimated travel time to the terminal.

In historical mode, the app allows you to see how observed passenger arrival, taxi waiting time and queue length varied over the course of each day in the past month.
Users can select statistics of interest and flip through both weekdays and weekends to see how they correlate with each other. Each mode serves a different purpose. Real time mode provides information to inform a taxi drivers decision on whether he should head to Changi Airport right now, while historical mode augments his experience and intuition on queueing at the airport with hard data, thus allowing him to plan his daily route to include a trip to the airport when demand is high.
7.2 Conclusions

The contributions of this thesis are threefold. The first is a quantitative study on the impact of passenger arrivals on taxi demand at Changi Airport, and the imbalance in taxi supply that is an immediate result of a lack of information about taxi demand at each terminal. We suggest that one way of optimizing this system would be to set up a real time control policy that limits taxis from entering a terminal’s queue when waiting times are long and redirects taxis to terminals where these waiting times are short.

The second contribution is the development of a novel queueing model and prediction engine that is used to predict the expected waiting times of taxis at each of Changi Airport’s four terminals. Unlike traditional models that require steady state assumptions, our model is non-equilibrium by nature and can handle varying arrival and departure rates to predict future queue lengths and waiting times, which we were able to verify with ground truth data from historical flight arrival and taxi records. We derive useful bounds for our predictions, which when communicated to taxi drivers will give them additional perspective to inform their decision to head to the airport.

Lastly we propose a real time taxi allocation policy that uses our prediction engine to send taxis to airport terminals where the predicted taxi waiting time is short via the ChangiNOW server. Taxi drivers can use an app to query the server and based on the taxi driver’s risk tolerance, waiting time threshold and estimated travel time to the airport, it tells the driver which terminal he should head to, if any. We tested this system in simulation, and our results show that the ChangiNOW system might able to reduce waiting times for taxis and passengers by about one-half and one-third respectively.

Our research is a first step towards a real time control system to balance the supply of taxis at Changi Airport. Providing adequate ground transportation to passengers is a problem faced by all airports worldwide, and we aim to implement our methods and algorithms in a commercial system to meet this challenge.
Bibliography


