Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

ChangiNOW: A Mobile Application for Efficient Taxi Allocation at Airports

Afian Anwar[1], Mikhail Volkov[1], and Daniela Rus[1]

Abstract— **We present an application that uses a predictive queueing model to efficiently allocate taxis. The 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 demand is highest. We propose a service model that enables our system to be deployed on a smartphone platform to participating taxi drivers. We present the theoretical details which underpin our prediction engine and corroborate 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.**

I. Introduction

We introduce 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 to efficiently allocate taxis across the airport's four terminals.

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 [1]. The airport has four terminals - Terminal One, Two, Three and a Budget Terminal. 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 available, 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). 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 (a) predict the expected waiting times at each terminal, and (b) 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 taxi's waiting time while ensuring that all the passengers that arrive at Changi Airport find a taxi waiting for them.

Fig. 1: 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.

The main contributions of this paper are:

- data mining algorithms to find the average waiting time, arrival rate, departure rate and queue length of taxis waiting at any given terminal,
- a first study on quantifying the imbalance of taxi supply at terminals of airports,
- a queuing model and an automated planning system that can be used to send taxis to an airport terminal when demand is high, and
- lastly, a direct comparison between simulated taxi and passenger waiting times in the current system versus one that uses ChangiNOW

*Support for this research has been provided by the by the Future Urban Mobility project of the Singapore-MIT Alliance for Research and Technology (SMART), SMART Innovation Center Explorer Grant No.015824-119 and ONR grant N00014-09-11051. We are grateful for this support.

[1]A. Anwar, M. Volkov and D. Rus are with the MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, MA 02139, United States {afian, mikhail, daniela} at csail.mit.edu

Fig. 2: 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.

A. 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 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. This problem was first posed by Kendall in [3]. Previous work [4], [5], [6], [7], 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 [8].

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 [9], [10]. Most proposed solutions to this problem involve minimizing some cost function subject to performance constraints. For example, [11] 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 [11], 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 IV, 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 [12], real time taxi trajectories were used to monitor taxi availability at taxi stands in Singapore while [13] visualized the real time spatial distribution of available taxis in Wu Han, China. Similarly, [14] 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, ChangiNOW 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 2).

B. Paper Outline

Section II introduces the problem setup, defines notation and states assumptions. We describe the data we use for this study in Section III. In Section IV, 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 Section V, 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.

II. Problem Statement

In this section 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 τ 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.

A. Service Model

Let us consider a scenario where every taxi in Singapore has a smartphone with our ChangiNOW app installed (Figure 3). 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) 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

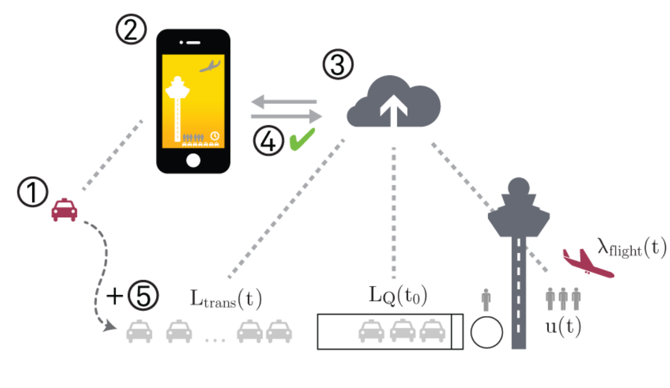


Fig. 3: Stages of the ChangiNOW service model: (1) taxi makes query, (2) server performs calculations, (3) server responds to taxi with optimal suggestion, (4) taxi makes acknowledgment, (5) server updates information.

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 τ 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. 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.

B. 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.

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 τ 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 λ.

III. Data

Our queuing model described in Section IV 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.

A. 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 $b_1, b_2 ... b_n$ that represent the physical queuing area at airport terminal T (Figure 4).

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

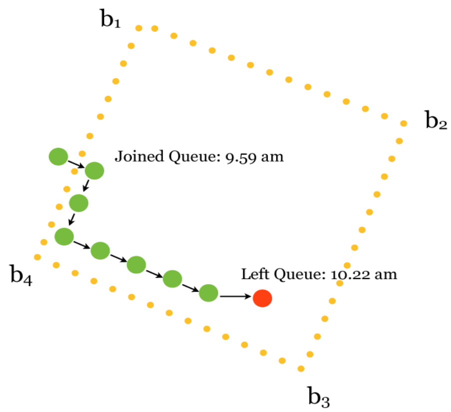


Fig. 4: Bounding box representing the terminal taxi queueing area. Each red (BUSY) or green (FREE) circle represents a taxi's state as it waited in the queueing area

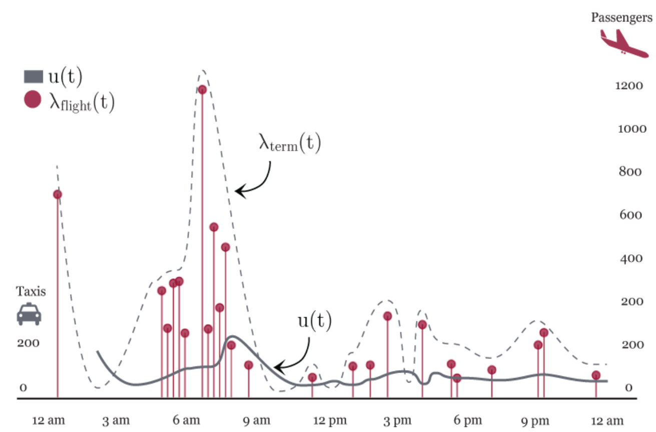


Fig. 5: Estimating derived taxi demand $u(t)$ from passenger arrival function $\lambda_{flight}(t)$

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.

B. 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 λ_{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 5). 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 λ_{flight} does not correspond to any given discrete time interval. To overcome this, we smooth the time series λ_{flight} using a 1×5 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

$$\mu(t) = f\lambda_{term}(t-k) \tag{1}$$

IV. Queueing Model and Prediction System

The taxi 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 $L_{trans}(t)$ that are in transit to each terminal. Further, we know the maximum queue capacity L_{max} and an estimate of the travel time τ to each terminal, as described in Section II-A.

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 $L_{trans}(t)$ 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

$$L_v(t) = L_q(t) + L_{trans}(t) \tag{2}$$

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.

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 $t+\tau$ $L_{trans}(t)$ 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

$$\bar{\mu}_\tau = \frac{1}{\tau}\int_t^{t+\tau} \mu(x)\,\mathrm{d}x \tag{3}$$

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 $L_{trans}(t)$ 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

$$\begin{aligned} E[L_q] &= L_q(t) + L_{trans}(t) - \tau\bar{\mu}_\tau \\ &= L_v(t) - \tau\bar{\mu}_\tau \end{aligned} \quad (4)$$

This gives us a quantitative statement for our first result.

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

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

2) 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 some times these results are inadequate.

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

(i) $L_v(t) < L_{max}$: This implies $E[L_q] < L_{max}$, 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) $E[L_q] \gg L_{max}$: 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 ε precision. Note that $L_v(t) \gg L_{max}$ does not necessarily imply that $E[L_q] \gg L_{max}$ since $\tau\bar{\mu}_\tau$ may be large.

(iii) $E[L_q] \approx L_{max}$: This is the main case of interest. Depending on the service rate $\bar{\mu}_\tau$ 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* $\Pr[\text{entry}] = \Pr[L_q(t+\tau) < L_{max}] =$

$$\int_t^{t+\tau} \bar{\mu}_\tau e^{-\bar{\mu}_\tau x} \frac{(\bar{\mu}_\tau x)^{(L_v(t)-L_{max})}}{(L_v(t)-L_{max})!}\,\mathrm{d}x \ . \quad (5)$$

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

3) 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* $E[W] =$

$$\min W^* \text{ s.t. } \int_{t+\tau}^{t+\tau+W^*} \mu(x)\,\mathrm{d}x \geq L_q(t+\tau) \ . \quad (6)$$

Proof: Define the waiting time service rate $\bar{\mu}_W$ as the average service rate while the taxi is waiting in the queue, given by

$$\bar{\mu}_W = \mu^* \text{ s.t. } \mu^* = \frac{1}{W^*}\int_{t+\tau}^{t+\tau+\frac{L_q(t+\tau)}{\mu^*}} \mu(x)\,\mathrm{d}x \ . \quad (7)$$

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

$$\frac{1}{L_q(t+\tau)}\int_{t+\tau}^{t+\tau+W^*} \mu(x)\,\mathrm{d}x = 1$$

and then multiplying across:

$$\int_{t+\tau}^{t+\tau+W^*} \mu(x)\,\mathrm{d}x = L_q(t+\tau) \quad (8)$$

i.e. the waiting time W^* must be such that (8) 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 mimimum W^* that satisfies (8), giving (6). ■

4) 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_{max}. Second, the taxi driver can specify a waiting time certainty margin $\alpha \in [0,1]$. We define the α-certainty waiting time W_α as a time such that a taxi driver entering the terminal at time $t+\tau$ will experience a wait of less than W_α with probability α.

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

$$\int_0^{W_{max}} \bar{\mu}_w e^{-\bar{\mu}_w x} \frac{(\bar{\mu}_w x)^{L_q(t+\tau)}}{L_q(t+\tau)!}\,\mathrm{d}x \ . \quad (9)$$

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

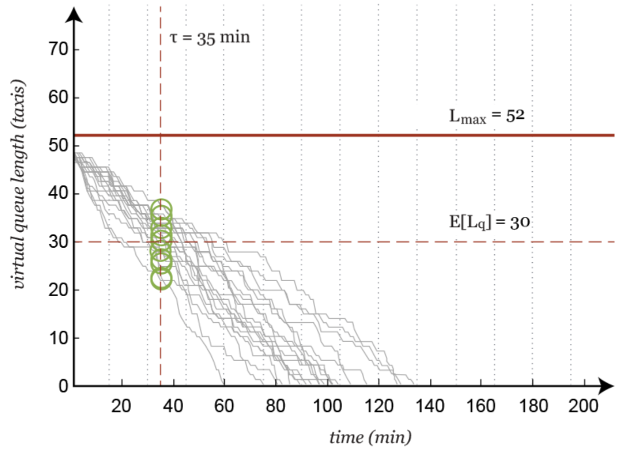


Fig. 6: When $L_v(t) < L_{max}$, all the taxis are guaranteed to enter the queue

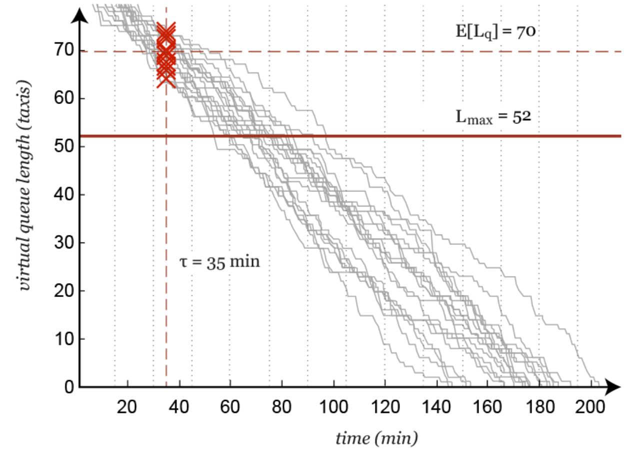


Fig. 7: When $E[L_q] \gg L_{max}$, taxis are almost certain to be rejected from the queue

$$= \min W^* \text{ s.t. } \int_0^{W^*} \bar{\mu}_w e^{-\bar{\mu}_w x} \frac{(\bar{\mu}_w x)^{L_q(t+\tau)}}{L_q(t+\tau)!} \, dx \geq \alpha \ . \qquad (10)$$

In (10) choose the smallest possible W_{max} such that the probability computed through the integral is greater than α.

V. Experiments and Results

In this section, we conduct several experiments using a simulation environment in MATLAB. We run two kinds of experiments - individual terminal simulations and a large scale urban simulation. Verifying the correctness of the results of individual terminal simulations before running a large scale urban 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.

A. Preliminary Simulations

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

In Figures 6, 7, 8 we plot time on the x-axis against the virtual queue length on the y-axis using 3 different initial queue length conditions. 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 IV-.2, 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).

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

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 7).

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

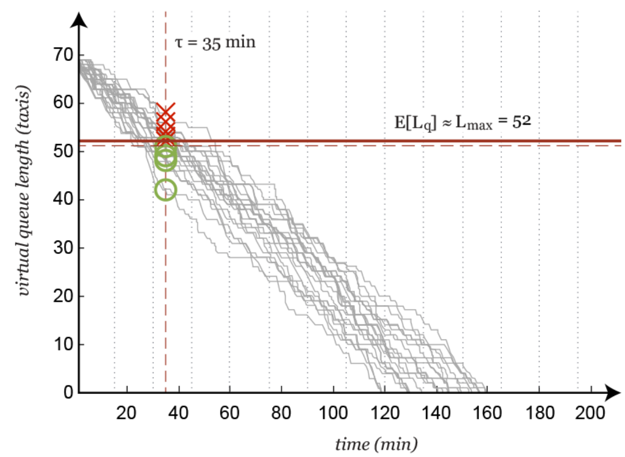


Fig. 8: When $E[L_q] \approx L_{max}$, some taxis are able to enter, while others are rejected from the queue

Figure 8 demonstrates why a simple expected queue length prediction is not enough. When $E[Lq] \approx L_{max}$, the number of taxis that entered the queue is split almost 50/50, so a definitive answer is not possible.

B. 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 8, 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 9).

As expected, when $E[L_q] \ll L_{max}$ (Case 1), every taxi is able to enter the queue and so $Pr[\text{entry}] = 1$. As $E[L_q]$ approaches L_{max}, $0 < \Pr[\text{entry}] < 1$ due to the stochastic nature of passenger arrivals at the front of the queue (Case 3). As we increase $E[L_q]$ past L_{max}, Pr[entry] drops to 0 (Case 2).

We validate Theorem 2 in simulation by adjusting L_q and L_{trans} so that Pr[entry] = **0.65**. The simulation results

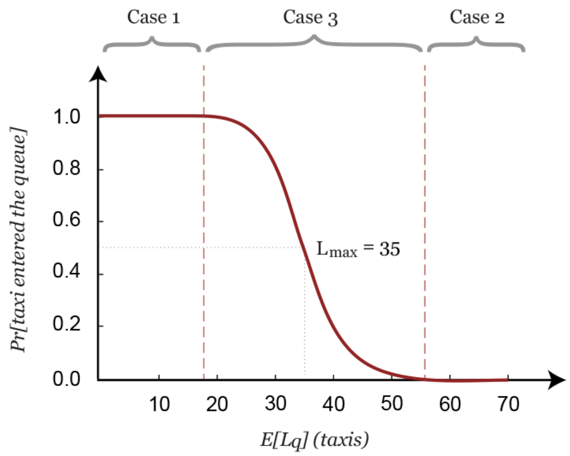


Fig. 9: This graph highlights the area of uncertainty (middle section in between the vertical dashed lines) when $0 < \Pr[\text{taxi entered the queue}] < 1$ effect due to $E[L_q] \approx L_{max}$. The plot shows the expected queue length on the x-axis against the probability of a taxi entering the queue on the y-axis. The vertical dashed lines indicate the certainty (either 0 or 1) cutoff at an accuracy of 3 decimal places.

(100,000 runs) are as follows:

$$\text{no. taxis entered} = 65,154/100,000 = \mathbf{0.65}$$

C. 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:

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

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

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

D. Maximum Waiting Time and α-certainty Simulations

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

$$\begin{aligned} \Pr[\text{entry}] &\approx 0.76 \\ \text{no. taxis entered} &= 75,431/100,000 \\ \text{Group A: avg. } \Pr[W < 40] &= \mathbf{0.18} \\ \text{Group B: avg. } W_\alpha, \alpha = \mathbf{0.9} &= 57 \text{ min} \end{aligned}$$

Tested in simulation:

$$\begin{aligned} \text{no. Group A with } W < W_{max} &= 13,695/75,431 = \mathbf{0.18} \\ \text{no. Group B with } W < W_\alpha &= 70,243/75,431 = \mathbf{0.93} \end{aligned}$$

E. Large Scale Urban Simulation

We test our rebalancing policy with a simulation environment comprising of 500 taxis, and 5 nodes, 4 representing each terminal at Changi Airport and the last, downtown Singapore. In our simulation, 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 $\lambda_{taxi_i}(t)$. Both $\mu_i(t)$ and $\lambda_{taxi_i}(t)$ 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:

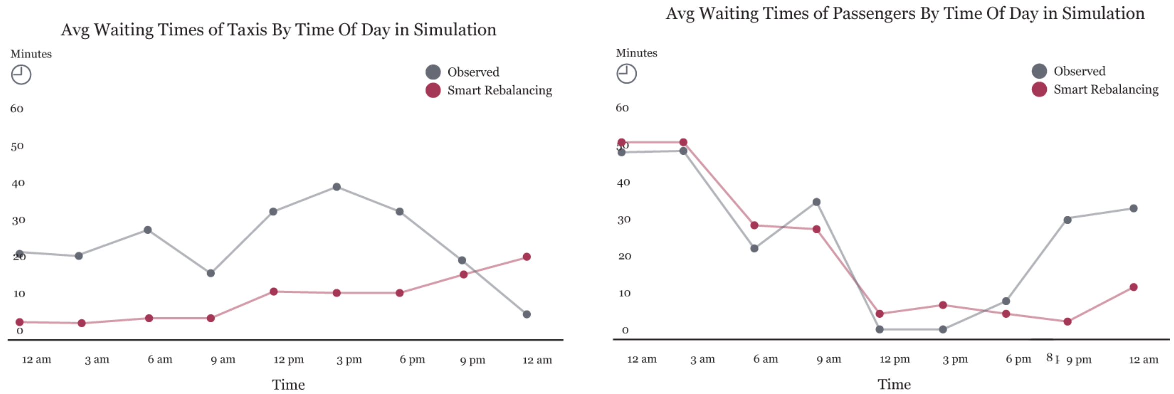
Observed Policy: P_{obs} is based on empirical taxi data. It represents the "ground truth" travel behavior of taxis that visit Changi Airport. To obtain it, we take the proportion of taxis entering terminal i at time t and smooth it using a 1x5 Gaussian kernel in time. This gives us the distribution $\alpha_i(t)$.

Smart Rebalancing Policy: In P_{smart}, taxis at each node i (including the terminal nodes) query our ChangiNOW server, which returns an answer, $DEST_j$ that tells the taxi where to go based on the projected waiting times each taxi would encounter and w_{max}, the maximum amount of time each taxi is prepared to wait. If there are no better alternatives, our server returns $DEST_{j=i}$, 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_{obs}, destinations are based on historical patterns while for P_{smart}, taxis are routed to the terminal with the shortest predicted waiting time.

For each policy, we plot the waiting time of taxis (Figure 10a) and passengers (Figure 10b) 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



(a) Taxi waiting times

(b) Customer waiting times

Fig. 10: Comparison of taxi waiting times under Observed and Smart Rebalancing policies.

simulation and real world contexts would use known flight and passenger manifest data provided by the airport.

VI. Conclusions

The contributions of this paper 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.

This 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 expect that the methods and algorithms described in this paper can be applied outside Singapore.

Acknowledgments

The authors would like to thank Amedeo Odoni for his valuable discussions, encouragement and advice.

References

[1] C. A. Group, "Changi airport - facts and statistics," March 2012.
[2] "Airport council international monthly traffic statistics," July 2012.
[3] D. Kendall, "Some problems in the theory of queues," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 151–185, 1951.
[4] R. Larson and A. Odoni, *Urban operations research*, ch. 4.1. No. Monograph, 1981.
[5] B. Kashyap, "The double-ended queue with bulk service and limited waiting space," *Operations Research*, pp. 822–834, 1966.
[6] M. Sasieni, "Double queues and impatient customers with an application to inventory theory," *Operations Research*, pp. 771–781, 1961.
[7] G. L. Curry, A. D. Vany, and R. M. Feldman, "A queueing model of airport passenger departures by taxi: Competition with a public transportation mode," *Transportation Research*, vol. 12, no. 2, pp. 115 – 120, 1978.
[8] B. Conolly, P. Parthasarathy, and N. Selvaraju, "Double-ended queues with impatience," *Computers and Operations Research*, vol. 29, no. 14, pp. 2053 – 2072, 2002.
[9] G. Berbeglia, J. Cordeau, and G. Laporte, "Dynamic pickup and delivery problems," *European Journal of Operational Research*, vol. 202, no. 1, pp. 8–15, 2010.
[10] S. Parragh, K. Doerner, and R. Hartl, "A survey on pickup and delivery problems," *Journal für Betriebswirtschaft*, vol. 58, no. 2, pp. 81–117, 2008.
[11] M. Pavone, S. Smith, E. Frazzoli, and D. Rus, "Load balancing for mobility-on-demand systems," *Robotics: Science and Systems, Los Angeles, CA*, 2011.
[12] S. K. Wei Wu, Wee Siong Ng, "To taxi or not to taxi? - enabling personalised and real-time transportation decisions for mobile users," in *2012 IEEE 13th International Conference on Mobile Data Management*, 2012.
[13] Y. Y. Ke Hu, Zhangguang He, "Taxi-viewer: Around the corner taxis are!," in *?2010 Symposia and Workshops on Ubiquitious and Trusted Computing*, 2010.
[14] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 99–108, ACM, 2010.