Journal of Advanced Transportation, Vol. 40, No. 2, pp. 203-220
www.advanced-transport.com

The Optimal Dispatching of Taxis under Congestion: a Rolling Horizon Approach

K.I. Wong
Michael G.H. Bell

Taxis make an important contribution to transport in many parts of the world, offering demand-responsive, door-to-door transport. In larger cities, taxis may be hailed on-street or taken from taxi ranks. Elsewhere, taxis are usually ordered by phone. The objective of a taxi dispatcher is to maximize the efficiency of fleet utilization. While the spatial and temporal distribution of taxi requests has in general a high degree of predictability, real time traffic congestion information can be collected and disseminated to taxis by communication technologies. The efficiency of taxi dispatching may be significantly improved through the anticipation of future requests and traffic conditions. A rolling horizon approach to the optimisation of taxi dispatching is formulated, which takes the stochastic and dynamic nature of the problem into account. Numerical experiments are presented to illustrate the performances of the heuristics, taking the time dependency of travel times and passenger arrivals into account.

Key Words: Taxi; Vehicle dispatching; Rolling horizon; Traffic congestion; Heuristics

1. Introduction

In many towns and cities, taxis are requested by telephone and then dispatched by a control centre. Taxis are in radio contact with their control centres from where they receive instructions about which job to go to next (and perhaps also the job after that). In some systems, the jobs are queued on a display in front of the driver, with the driver reporting

K.I. Wong, Department of Transportation Technology and Management, National Chiao Tung University, Taiwan, R.O.C.
Michael G H Bell, Department of Civil and Environmental Engineering, Imperial College London, London, United Kingdom
Received: August 2004; Accepted: August 2005

back to the control centre upon the completion of each job. This paper is concerned with taxi dispatching algorithms.

Perhaps the simplest and most practical algorithm is to send the taxi that can reach each request first. An existing taxi dispatching system has been studied recently by Lee *et al.* [2003]. A common practice adopted in dispatching systems using the Global Positioning System (GPS), employed by some taxi operators, is the assignment of the taxi nearest the request. As an alternative dispatching criterion, they propose assigning the closest taxi in travel time terms, as determined by real-time traffic conditions.

In this paper, we assume the travel time and taxi requests arise in a time-dependent environment. Taxis are assigned jobs by a central dispatcher with the objective of minimizing passenger waiting time. This objective can be regarded as synonymous with the maximisation of fleet utilisation. Dispatching the taxi that can reach a request first may, however, result in a sub-optimal assignment of taxis. It is conceivable that the taxi that can reach the request first might actually be better assigned to a subsequent request.

The principal of looking ahead to the next request could clearly be extended to looking ahead to the next two, three or more requests, probably with rapidly diminishing returns. An efficient method is required for finding the assignment of the next N taxis that minimises total expected customer waiting time based on a time-dependent probability for the origin and destination of the request and a time-dependent headway between requests. This method can then be incorporated in a rolling horizon approach to taxi assignment, with the anticipation of future requests and traffic conditions.

In practice, finding the taxi that minimises the expected total wait over a rolling horizon is computationally very demanding, and for large zoning systems and fleets not possible in real time. Consequently a heuristic has also been devised which for each taxi compares the wait that would be experienced by the current requester if it were assigned now with the expected wait of the next requester if instead it were assigned to the next request. The benefits of a dispatching policy that looks ahead and the heuristic just outlined are tested in simulation.

2. Literature Review

The vehicle routing problem (VRP) has been studied for decades, and different approaches and algorithms have been developed to solve specific problems [Toth and Vigo, 2002]. The taxi dispatching problem is a form of VRP [Ghiani *et al.*, 2003]. The traditional VRP seeks the optimal pick up and drop off schedules for a fleet of vehicles with finite capacities subject to time constraints of various kinds. Requests are generally known in advance, but dynamic requests can be treated by inserting these into the existing routes. The taxi dispatching problem differs from the traditional VRP in a number of important ways. Customer requests are dynamic, the pickup time is usually "as soon as possible" and the delivery to the drop off point should be without deviation. The taxi dispatching problem can be classified as strongly dynamic, compared with other problems like dial-a-ride or courier services. Larsen [2000] gives an overview of the dynamic vehicle routing problem (DVRP).

Most solution algorithms for real-time VRPs do not take historical information into account. For some of the emergency services, like fire and ambulance, which are also highly dynamic, the quality of a priori information is generally poor in terms of the locations of future requests. In the case of taxi dispatching, however, the spatial distribution of requests can be expected to follow a regular daily pattern, subject to stochastic variations [Gendreau *et al*., 1996]. This information about the future can be used to improve the efficiency of dispatching through a look ahead capability. The essentials of heuristics with a look ahead capability are discussed in Ghiani *et al.* [2003].

Another concern with VRPs is variation in congestion and travel time, which has received little attention. Nearly all models proposed in the literature adopt the assumption of constant travel time. Two reasons for this were pointed out by Fleischmann *et al*. [2004]. It adds complexity to the problem formulation which may require essential structural modifications in the algorithms, and the estimation and data collection for time-varying travel times are difficult so the reliability of the results are open to question. One example considering traffic congestion which is stochastic and time-dependent throughout the day is studied by Fu [2002], who investigated the dial-a-ride problem (DARP). With the uncertainty in travel time, the time windows for pickup and delivery are formulated probabilistically with pre-specified threshold of reliability. Although the travel time is stochastic and time-varying, the requests are known in advance so their problem is not fully dynamic. In real time systems, planning decisions are frequent and the network condition and congestion level can always be updated at the moment of planning, which minimizes the offset of the decision from the optimal.

On the assessment of road traffic and congestion due to taxi movements, research on taxi modelling has recently been carried out by Yang and Wong [1998]. They modelled the equilibrium of street-hailed taxi services. They characterized taxi movements for a given and fixed customer O-D demand pattern in an uncongested road network. The model was further reformulated and extended to incorporate congestion effects and customer demand elasticity [Wong *et al*., 2001], for which network equilibrium with movements of vacant and occupied taxis is described. Relationships between taxis and customer waiting times, and the relationship between customer demand and taxi supply are further constrained and satisfied. Furthermore, the potential applications of the model have been demonstrated by several case studies of the urban area of Hong Kong [Yang *et al*., 2001, 2002]. In a more recent study, Wong *et al*. [2005] developed a model for the bilateral micro-searching behaviour of taxi drivers and customers. The absorbing Markov chain approach is employed to formulate the taxi movement within the network, where the local searching behaviour of taxis is specified by a logit model, and the O-D demand of passengers is also estimated by a logit model with a choice of taxi meeting point.

Taxi dispatching as considered in this paper is a form of real-time dispatching problem. Customer requests are dynamic, the requested pickup time is usually "as soon as possible", and the delivery to the drop off point should be without deviation. The problem is therefore highly dynamic. The vehicle routing problem is a generalization of the travelling salesman problem, a well known combinatorial optimization problem, and consequently is NP-hard. Any optimal solution method to these problems may only be practical for small sized cases. As real-time VRP problems are more difficult to solve, given the dynamic nature of the problem, exact algorithms are not yet capable of handling typically large practical problems. This justifies the use of heuristics in real-time environments. In a previous work Bell *et al*. [2005] considered the problem in an environment where travel time and speed are assumed to be constant. In this paper, the model is extended and the time varying traffic condition for taxi dispatching will be considered with the look ahead approach.

3. Methodology

Assume a customer request with a defined pick-up point and delivery point is generated at a particular time. Let I and J be the set of origin and destination zones, and K be the taxi fleet. Further let

$a(k)$	=	Scheduled delivery time of taxi k
$d(k)$	=	Scheduled delivery point of taxi k
$p_{ij}(t)$	=	Probability of a taxi trip from i to j at time t
$h(t)$	=	Expected headway between requests at time t
$c(i, j, t)$	=	Travel time from i to j at the departure time t

The spatial and temporal distribution of request, $p_{ij}(t)$ and $h(t)$, is assumed to be time-dependent following a regular and predictable pattern. The travel time in the network, $c(i, j, t)$, which varies through the day due to traffic congestion, can be modelled as a function of the departure time t. It could be predicted by a time dependent shortest path algorithm.

Taxi assignment without look-ahead

In the case of taxi assignment without look ahead, a simple dispatching rule is to assign the taxi which minimises the wait time of the current requester. Suppose a request with pick up point i and delivery point j is generated at time t. The customer wait time if taxi k is assigned is

$$w(t, a(k), d(k), i) = \max(0, a(k) - t) + c(d(k), i, t + \max(0, a(k) - t)),\quad k = 1, 2, ..., K \tag{1}$$

where the travel time c is evaluated for the moment when the vehicle is immediately available to the passenger. The dispatcher assigns the taxi that can reach the requester first, so

$$k^* = \arg\min_k (w(t, a(k), d(k), i)),\quad k = 1, 2, ..., K \tag{2}$$

If the taxi arrived at its scheduled delivery point before this job is assigned to it, we assume it remained idle at that location. The idle time for taxi k^* is therefore

$$l(t, a(k^*)) = \max(0, t - a(k^*)) \quad (3)$$

In practice, many taxi dispatchers would reposition their idle taxis to be closer to where future requested pickups are likely to arise.

The scheduled delivery time and delivery point of taxi k^* are then updated as follows

$$a(k^*) = t + w(t, a(k^*), d(k^*), i) + c(i, j, t + w(t, a(k^*), d(k^*), i)) \quad (4)$$

$$d(k^*) = j \quad (5)$$

where c in Eq. (4) is calculated for the time that passenger is reached by the taxi. While this dispatching strategy is likely to be reasonably efficient, and is equivalent in some respects to a greedy algorithm, a better assignment could be made if the dispatcher were to anticipate future requests and allow this to modify the current assignment.

Taxi assignment with look-ahead

Assume the dispatcher knows the probabilistic request profile (i.e. the probability the next request will correspond to a particular OD pair) and the average headway between requests. The expected time of the next request is $t+h(t)$, where $h(t)$ is the time dependent headway between requests. The probability that this request has pickup point i and delivery point j is $p_{ij}(t+h(t))$. The expected time of the next n^{th} request can be calculated by $t + \sum_{n=1..N} h_n \left(t + \sum_{n'=1..n-1} h_{n'}\right)$, where the time dependent headway is updated with the look head steps. For simplicity it is approximated in the first order as $t+nh(t)$ in the following calculations. Suppose the dispatcher looks ahead to the next N assignments and seeks to minimize total expected waiting time. The problem can be formulated as a Dynamic Programming (DP) problem with a finite horizon. In the terminology of dynamic programming, let the dispatching policy be $\pi = \{k_0, \ldots, k_n, \ldots, k_N\}$, where k_n is the taxi dispatched at stage n. In the N-stage problem, the expected cost of a policy π, given the initial customer pick up location i_0, is

$$J_\pi(i_0) = w(t, a(k_0), d(k_0), i_0) + \sum_{n=1..N} \psi(t + nh(t), a(k_n), d(k_n)) \quad (6)$$

where cost is measured in terms of expected waiting time, and

$$\psi(t+nh(t), a(k_n), d(k_n)) = \sum_{ij}\left(\max(0, a(k_n)-t-nh(t)) + c(d(k_n), i, t+\max(0, a(k_n)-t-nh(t)))\right)p_{ij}(t+nh(t)) \tag{7}$$

If $k_n \in \pi$ then

$$a(k_{n+1}) = t+nh(t)+w(t, a(k_n), d(k_n), i)+c(i, j, t+nh(t)+w(t, a(k_n), d(k_n), i)) \tag{8}$$

And

$$d(k_{n+1}) = j \tag{9}$$

Otherwise

$$a(k_{n+1}) = a(k_n) \tag{10}$$

And

$$d(k_{n+1}) = d(k_n) \tag{11}$$

The optimal policy is

$$J_{\pi^*}(i) = \min_{\pi}(J_{\pi}(i)) \tag{12}$$

Note that as a consequence of (8) and (9), which update the delivery time and point for the assigned taxi, the expected wait for the n^{th} request, $\psi(t+nh(t), a(k_n), d(k_n))$, depends on the preceding *n-1* assignments and corresponding requests. Consequently the calculation of $J_\pi(i)$ is computationally very demanding. If the taxi fleet consists of *K* taxis then the total number of possible assignments is K^{N+1}. For each assignment after the first, all possible requests need to be considered. Consequently computational complexity is $O(K^{N+1}OD^N)$, where *OD* is the number of origin-destination pairs.

Given its recursive structure, Bellman's decomposition may be applied [see Bertsekas, 1995]. However, finding the taxi that minimises the expected total wait over the rolling horizon remains computationally very demanding, and for large fleets is impractical in real time (Bellman's "curse of dimensionality"). Consequently, a heuristic has been devised which for each taxi compares the wait that would be experienced by the current request if it were assigned now with the expected wait of the next request if instead it were assigned to the next request.

A heuristic algorithm

The concept of this heuristic is to search for a taxi which minimizes the difference between the wait incurred if the taxi is assigned to the current request at point i and α times the expected wait if the taxi is assigned instead to the next request.

H1 (α):

$$k^* = \arg\min_k \Big(w(t, a(k), d(k), i) - \alpha \sum_{ij} w(t + h(t), a(k), d(k), i) p_{ij}(t + h(t)) \Big), \quad k = 1, 2, ..., K \tag{13}$$

where α is a parameter which discounts future waiting time. Experiments show that the best value of α lies between 0.5 and 0.75. Each taxi is looked at once, so the computational complexity is only O(K* OD). As the following simulation results demonstrate, this heuristic is able to reduce total waiting time considerably, particularly during periods when the demand for taxis is high.

4. Simulation Experiments

In order to demonstrate the effectiveness of the proposed methodology, a simulation is used to verify the improvement offered by looking ahead periods of different durations. Monte Carlo simulation is adopted to generate the customer requests, while a taxi dispatcher assigns taxis using the proposed methods.

In this section, the effects of various factors on the performance of the algorithm are examined. We assume a network in the shape of a strip with 9 zones as shown in Figure 1. The travel time and trip rates between each pair of zones are assumed to be time dependent as in Figure 2, from 7am to 9pm for a total of 14 hours. The travel distances between zones are displayed in Table 1. Travel times between zones are assumed to be linear with the distances and inversely proportional to the travel speed, which is time-varying and set at 40 km/hr in the off peak period but drops to 20 km/hr during the morning and evening peak periods. The time varying travel time can be formulated as a function of the departure time and obtained by a time-dependent shortest path algorithm. In this example, it is calculated by summing the quotients of the travel distance from i to j divided by the piecewise linearized travel speed from time t (with a step of 1 minute say) until the vehicle reaches j. As a result, the travel time within zones varies between 7.5 and 15 minutes (for 5 kilometres), and between 45 and 90 minutes (for 30 kilometres) from zone 1 to zone 9, depending on the departure time of the vehicle.

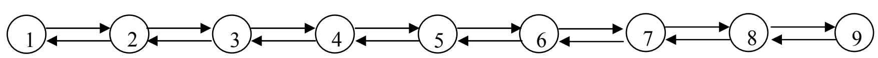

Figure 1. The example network

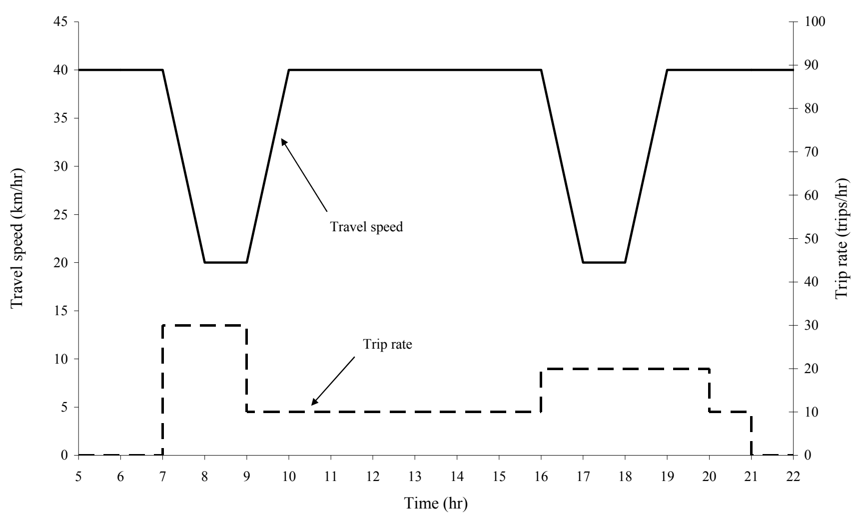

Figure 2. Time-dependent travel speed and trip rates profile

Table 1. Travel distance table (km)

Zone	1	2	3	4	5	6	7	8	9
1	5	8	11	14	17	20	23	26	30
2	8	5	8	11	14	17	20	23	26
3	11	8	5	8	11	14	17	20	23
4	14	11	8	5	8	11	14	17	20
5	17	14	11	8	5	8	11	14	17
6	20	17	14	11	8	5	8	11	14
7	23	20	17	14	11	8	5	8	11
8	26	23	20	17	14	11	8	5	8
9	30	26	23	20	17	14	11	8	5

The rate of trip generation varies with the time, between 10 and 30 requests per hour. The trip rate tables are inversely proportional to the travel distance, and we considered two directions of flow. One is right to left (Table 2a), in which there is more traffic going from zone 9 to zone 1 than from zone 1 to zone 9; and the other is left to right (Table 2b), in which more traffic is moving from zone 1 to zone 9, a reversal compared to Table 2a. The zones in the table correspond to those in Figure 1. An element in the table represents the possibility of a trip that a particular OD pair corresponds to, and the sum of all elements in the table equals 100. We assumed two scenarios: (P1) constant trip rate with traffic from right to left (see Table 2a) throughout the day and (P2) varying trip rate pattern, with right to left (see Table 2a) from 7am to 2pm and left to right (see Table 2b) from 2pm to 9pm. P1 characterizes the simple case where the OD pattern is fixed, while P2 illustrates the case with morning peak and evening peak where the flow pattern is tidal but in opposite directions.

Taking the trip rate table as the means, the requests are generated by Monte Carlo simulation. Trip generation follows a Poisson process with a time-varying headway calculated from Figure 2, and a total of 220 trips are considered. For each request, the taxi dispatcher assigns a taxi. Once assigned, the taxi cannot be substituted with another. The simulation is performed with 5 different seed numbers and the results are averaged to account for the effect of the seeds. For the same comparison between heuristics, the demand sequences are kept the same for each of the seed numbers.

Table 2a. Trip rate table (%): right to left [insert Table 2b here]

Zone	1	2	3	4	5	6	7	8	9
1	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48
2	1.31	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48
3	1.15	1.31	1.48	1.48	1.48	1.48	1.48	1.48	1.48
4	0.99	1.15	1.31	1.48	1.48	1.48	1.48	1.48	1.48
5	0.82	0.99	1.15	1.31	1.48	1.48	1.48	1.48	1.48
6	0.66	0.82	0.99	1.15	1.31	1.48	1.48	1.48	1.48
7	0.49	0.66	0.82	0.99	1.15	1.31	1.48	1.48	1.48
8	0.33	0.49	0.66	0.82	0.99	1.15	1.31	1.48	1.48
9	0.16	0.33	0.49	0.66	0.82	0.99	1.15	1.31	1.48

Table 2b. Trip rate table (%): left to right

Zone	1	2	3	4	5	6	7	8	9
1	1.48	1.31	1.15	0.99	0.82	0.66	0.49	0.33	0.16
2	1.48	1.48	1.31	1.15	0.99	0.82	0.66	0.49	0.33
3	1.48	1.48	1.48	1.31	1.15	0.99	0.82	0.66	0.49
4	1.48	1.48	1.48	1.48	1.31	1.15	0.99	0.82	0.66
5	1.48	1.48	1.48	1.48	1.48	1.31	1.15	0.99	0.82
6	1.48	1.48	1.48	1.48	1.48	1.48	1.31	1.15	0.99
7	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.31	1.15
8	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.31
9	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48	1.48

The dispatching algorithms are tested for different demand intensities. When the demand is relatively low, one would expect all requests to be met while the waiting times could depend on the dispatching policy. For heavy demand, in the peak hour, the system may not be able to keep up with the demand, so delay may increase continuously. The results of typical simulation runs for different taxi fleet sizes are shown in Table 3 for P1 and Table 4 for P2. Tables 3a and 4a display the average customer wait in minutes and Tables 3b and 4b show the average taxi idle time in minutes. It can be seen that the customer wait is generally decreasing with the number of taxis in operation for all the algorithms, as the taxis would be less busy and have a higher availability. The customer waiting time is very high when the number of taxis is less than about 10.

Table 3a. Average customer waiting times (minutes) for P1

Number of taxis	Customers per taxi	Without look ahead	1-step look ahead	2-step look ahead	H1, $\alpha = 0.5$	H1, $\alpha = 0.6$	H1, $\alpha = 0.7$
6	36.7	323.9	297.8	281.2	295.2	286.4	281.7
7	31.4	214.7	200.0	188.7	195.6	188.2	183.2
8	27.5	139.6	127.3	119.0	121.8	119.9	117.9
9	24.4	91.5	84.8	81.0	82.4	81.7	80.1
10	22.0	72.3	66.4	64.0	66.2	65.7	64.3
11	20.0	60.0	55.7	53.9	55.3	54.7	54.4
12	18.3	51.9	48.2	46.1	47.1	47.1	46.9
13	16.9	45.7	42.0	40.2	42.2	41.9	41.3
14	15.7	40.3	37.0	35.6	37.3	36.7	36.8
15	14.7	36.3	33.7	32.4	33.6	33.4	33.8
16	13.8	33.2	30.6	29.9	30.8	31.0	31.0
17	12.9	30.8	28.8	28.1	29.0	28.9	29.5
18	12.2	28.6	26.7	25.9	26.7	27.2	27.8
19	11.6	26.8	25.0	24.3	25.3	25.7	26.3

Table 3b. Average taxi idle times (minutes) for P1

Number of taxis	Customers per taxi	Without look ahead	1-step look ahead	2-step look ahead	H1, $\alpha = 0.5$	H1, $\alpha = 0.6$	H1, $\alpha = 0.7$
6	36.7	0.3	0.2	0.3	0.3	0.3	0.3
7	31.4	0.4	0.4	0.4	0.4	0.4	0.4
8	27.5	0.5	0.6	0.7	0.6	0.7	0.8
9	24.4	1.9	2.1	2.5	2.4	2.6	2.9
10	22.0	4.9	5.1	5.4	5.5	5.7	5.9
11	20.0	8.1	8.3	8.5	8.8	9.0	9.1
12	18.3	11.1	11.4	11.7	12.2	12.0	12.2
13	16.9	14.3	14.6	14.9	15.1	15.2	15.3
14	15.7	17.5	17.9	17.9	18.2	18.3	18.4
15	14.7	20.7	20.9	21.1	21.5	21.2	21.8
16	13.8	23.7	24.0	24.2	24.5	24.8	25.1
17	12.9	26.3	27.1	27.3	27.6	28.1	28.3
18	12.2	29.8	30.2	30.3	31.1	31.1	31.2
19	11.6	33.1	33.3	33.5	34.3	34.5	34.9

The waiting time decreases initially and then tends to be flat with increasing numbers of taxis, as more taxis improve the service and the waiting time approaches the quickest response time (the intrazonal travel time). In almost all cases, the 2-step look ahead outperforms the 1-step look ahead and the 1-step look ahead outperforms the case without any look ahead. The performance of the heuristic depends on alpha and the demand level in number of customers per taxi. When the taxi fleet is small and the system is very busy, the heuristic with alpha value of 0.7 performs the best of all the algorithms tested. With the case of 10 taxis, it saved more than 10% of the customer wait for both P1 and P2 when compared with the algorithm without look ahead. However, when the number of taxis is greater than about 15, it shows no advantage over 1-step look ahead. The heuristic with alpha of 0.5 or 0.6 is better than 1-step look ahead for those cases with 13 taxis or less, but becomes less effective when the system is less busy. When compared to the heuristic with alpha of 0.7, it tends to be more stable and closer to the look ahead results for lower demand intensities. The heuristics with larger α values generally perform well when the demand level is high, but it generates higher waiting time when the demand level is low, when compared to the without look-ahead case. This could be explained by the fact that the strategy with a higher α value tends to reserve the system capacity for future demand.

The average taxi idle time for P1 is shown in Table 3b. It is interesting to note that for all cases of taxi fleet the heuristics with alpha value of 0.7 generates the highest taxi idle wait compared to other heuristics, no matter the average passenger wait is higher or lower than that of other heuristics. The taxi idle time has a strong relation with the vacant taxi mileages, from the conservation of total taxi time, i.e., total taxi time is equal to the sum of occupied taxi time, vacant taxi time and taxi idle time [Wong *et al*., 2001]. Since total taxi time and occupied taxi time are fixed for each scenario, longer taxi idle time implies shorter vacant taxi hours or mileages in general, without taking into account the effect of congestion over the day. This confirms that the saving in passenger wait by the heuristics comes from the better assignment of vehicles and the decrease in the vacant mileages of taxis.

Similar conclusions can be drawn for the problem P2, as shown on Table 4. This suggests that the performances of the heuristic depend on the chosen discount factor (alpha), and the best alpha depends on the demand intensity. This dependency is worth further investigation.

Table 4a. Average customer waiting times (minutes) for P2

Number of taxis	Customers per taxi	Without look ahead	1-step look ahead	2-step look ahead	H1, $\alpha = 0.5$	H1, $\alpha = 0.6$	H1, $\alpha = 0.7$
6	36.7	314.2	288.0	287.7	282.9	273.6	269.2
7	31.4	205.1	193.8	193.0	184.6	178.0	172.5
8	27.5	131.1	121.2	121.1	113.6	110.9	108.6
9	24.4	84.9	78.7	78.5	76.9	75.3	74.0
10	22.0	66.0	60.5	60.5	59.4	59.6	58.9
11	20.0	53.1	49.2	49.1	48.3	47.9	48.2
12	18.3	44.9	41.2	41.1	40.2	40.3	40.7
13	16.9	38.5	35.5	35.5	35.0	34.7	35.4
14	15.7	33.4	30.7	30.7	30.3	30.6	31.0
15	14.7	29.6	27.5	27.5	27.8	27.9	28.0
16	13.8	27.2	25.2	25.2	25.5	26.1	25.8
17	12.9	25.2	23.6	23.6	23.7	24.0	24.5
18	12.2	22.9	21.7	21.8	22.0	22.0	22.8
19	11.6	21.3	20.2	20.2	20.7	20.9	21.4

Table 4b. Average taxi idle times (minutes) for P2

Number of taxis	Customers per taxi	Without look ahead	1-step look ahead	2-step look ahead	H1, $\alpha = 0.5$	H1, $\alpha = 0.6$	H1, $\alpha = 0.7$
6	36.7	0.3	0.2	0.2	0.3	0.3	0.3
7	31.4	0.4	0.4	0.4	0.4	0.4	0.4
8	27.5	0.5	0.6	0.6	0.7	0.8	0.9
9	24.4	2.1	2.3	2.3	2.7	2.9	3.2
10	22.0	5.1	5.5	5.5	6.0	6.1	6.3
11	20.0	8.5	8.7	8.7	9.2	9.4	9.6
12	18.3	11.7	11.8	11.9	12.5	12.5	12.6
13	16.9	14.9	14.9	14.9	15.6	15.8	15.9
14	15.7	18.0	18.5	18.5	19.0	18.9	19.1
15	14.7	21.5	21.6	21.4	21.9	22.2	22.4
16	13.8	24.5	24.9	25.0	25.2	25.7	25.7
17	12.9	27.7	27.8	27.8	28.6	28.9	28.8
18	12.2	31.0	31.5	31.5	31.7	32.2	32.2
19	11.6	34.2	35.1	35.1	35.3	35.9	35.7

The efficiency of dispatching algorithms determines the response time to requests and/or the number of taxis needed.

Different dispatching rules can give different distributions of customer waits, even if their average values are the same. To investigate this, we select a case where the taxi fleet has size 8. The cumulative frequencies of customer waiting times for three algorithms are shown in Figure 3. The cumulative frequency for the algorithm without look ahead increases approximately linearly from zero to about 120, and then increases rapidly to its maximum value. By contrast, for the 2-step look ahead and the heuristic algorithms, the cumulative frequency starts to increase more rapidly around 70. The gap between the two cumulative frequency curves equals the saving in passenger waiting time due to looking ahead. It also shows that the 2-step look ahead and the heuristic algorithms are similar in equity terms as the cumulative distribution functions are parallel. The corresponding statistics of maximum, minimum, average and standard deviation of the passenger waiting time for the three algorithms are shown in Table 5. Both the 2-step look ahead and the heuristic have smaller waiting time standard deviations. While the heuristic shows a smaller average and standard deviation compared with the 2-step look ahead method, it has a higher maximum.

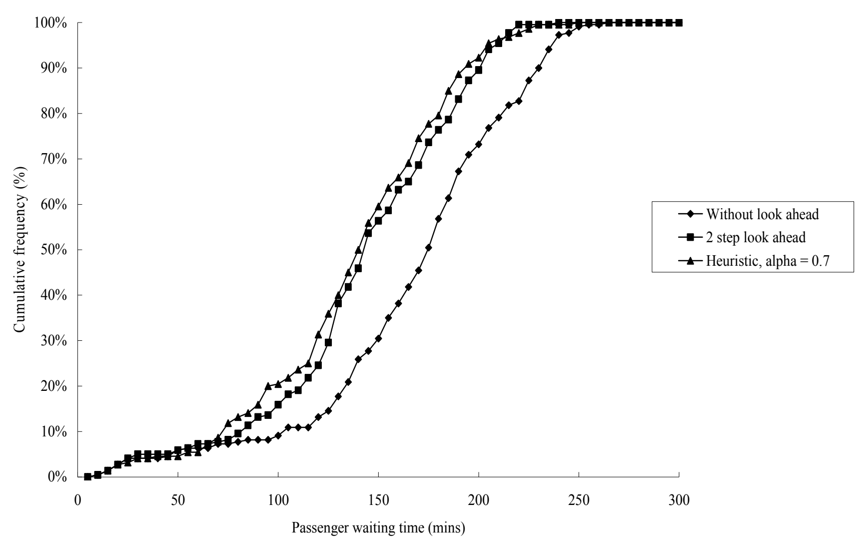

Figure 3. Cumulative frequency of customer waiting times for different algorithms

Table 5. Statistical values of various dispatching algorithms

	Without look ahead	2-step look ahead	Heuristic, $\alpha = 0.7$
Maximum	265.0	239.8	246.3
Minimum	7.5	7.5	7.5
Average	166.7	141.9	137.1
Standard Deviation	52.9	47.7	47.5

5. Conclusions

The real-time dispatching of taxis in response to incoming requests is studied, with the objective of minimising the total wait experienced by requesters over time. Several dispatching rules are examined. Given the distribution of future requests, which may be obtained from historical trip origin and destination data, a rolling horizon approach is formulated. As finding the taxi that minimises the expected total wait over a rolling horizon is computationally very demanding, a heuristic has been devised. A simulation model is developed and the properties of several dispatching algorithms are explored, taking into account the time dependency of travel times and passenger arrivals. The results of both the rolling horizon and the heuristic approaches are promising, but depend on the demand intensity. The performance of the rolling horizon approach improves with the number of look ahead stages, but the "curse of dimensionality" is the bottleneck to practical applications. The heuristic presented in this paper looks particularly promising when demand is high, but further investigation is required.

A limitation of the proposed heuristic algorithm is that we did not make use of the free time between the arrivals of calls to do further calculations. Some studies suggested the use of meta-heuristics in that spare period to further optimize the decisions, while some simple heuristic rules are used in the first instance of request arrival to reduce the computing requirement. This type of "double-horizon" heuristic can be considered in further research.

Acknowledgement

The second author gratefully acknowledges the financial support of the Fellowship scheme of the Croucher Foundation of Hong Kong. We would also like to thank the two anonymous referees for their suggestions and helpful comments on an earlier version of the paper.

References

Bell, M.G.H., Wong, K.I. and Nicholson, A.J. 2005. A rolling horizon approach to the optimal dispatching of taxis. In H.S. Mahmassani (ed.) *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction*, Elsevier Science Ltd., pp 629-648. (ISBN: 0-08-044680-9)

Bertsekas, D. P. 1995. *Dynamic Programming and Optimal Control*. Vol. 1. Athena Scientific. Belmont, Massachusetts, U.S.A.

Fleischmann, B., Gietz, M. and Gnutzmann, S. 2004. Time-varying travel times in vehicle routing. *Transportation Science* 38(2): 160-173.

Fu, L. 2002. Scheduling dial-a-ride paratransit under time-varying, stochastic congestion. *Transportation Research* 36B: 485-506.

Gendreau, M., Laporte, G. and Seguin, R. 1996. Stochastic vehicle routing. *European Journal of Operational Research* 88**:** 3-12.

Ghiani, G., Guerriero, F., Laporte, G. and Musmanno, R. 2003. Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research* 151: 1-11.

Larsen, A. 2000. *The dynamic vehicle routing problem*. Ph.D. thesis, Department of Mathematical Modelling (IMM) at the Technical University of Denmark (DTU). Available on the Internet.

Lee, D.H., Wang, H., Cheu, R.L. and Teo, S.H. 2003. A taxi dispatch system based on current demands and real-time traffic conditions. *82nd Transportation Research Board Annual Meeting*, 12 – 16 Jan 2003, Washington DC, USA.

Toth, P. and Vigo, D. 2002. *The vehicle routing problem*. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia.

Wong, K.I., Wong, S.C. and Yang, H. 2001. Modeling urban taxi services in congested road networks with elastic demand. *Transportation Research* 35B: 819-842.

Wong, K.I., Wong, S.C., Bell, M.G.H. and Yang, H. 2005. Modeling the bilateral micro-searching behavior for urban taxi services using the absorbing Markov chain approach. *Journal of Advanced Transportation* 39(1): 81-104.

Yang, H. and Wong, S.C. 1998. A network equilibrium model of urban taxi services. *Transportation Research* 32B: 235-246.

Yang, H., Wong, K.I. and Wong, S.C. 2001. Modeling urban taxi services in road networks: progress, problem and prospect. *Journal of Advanced Transportation* 35: 237-258.

Yang, H., Wong, S.C. and Wong, K.I. 2002. Demand-supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research* 36B: 799-819.