Qualifying the benefits of ride-sharing on reducing fleet size

Shan Hao¹, Xinglu Liu², Lixin Miao¹, Wai Kin Victor Chan² and Mingyao Qi¹*

¹ Division of Logistics and Transportation Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, 518055, China
² Intelligent Transportation and Logistics Systems Laboratory, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, Guangdong, 518055, China
*Corresponding author’s e-mail: qimy@sz.tsinghua.edu.cn

Abstract. Reducing the number of operating vehicles in cities has enormous significance on mitigating traffic congestion and environment pollution. Ride-sharing is an efficient way to reduce fleet size in urban areas. In this work, we propose two integer programming models to qualify the benefits of ride-sharing on reducing fleet size. The proposed models are solved by commercial solver Gurobi. Then we conduct a series of instances based on trip records of New York City to test the proposed models. Results indicate that without delaying drop-off times, the fleet size when considering ride-sharing remains almost the same as ride-hailing service for high-density travel demand settings. Whereas the fleet size drops sharply as the demand density decreases. In addition, the number of vehicles required is reduced by nearly 30% regardless of order density under ride-sharing assumptions when a slight delay is allowed.

1. Introduction

On-demand travel request is one of the most important demands in modern cities. According to the World Bank’s forecasts, by 2050, about 5.4 billion will live in urban areas, and the number of vehicles on road will double to 2 billion[1]. These changes will lead to more severe traffic congestion and deteriorated urban environment. Litman proposed that an effective transportation system requires a significant reduction in the use and circulation of vehicles[2]. New York City Council passed regulations to limit the number of vehicles providing ride-hailing service in 2018. Thus, reducing the scale of operating vehicles in cities has enormous significance on improving traffic efficiency.

Previous studies on determining minimum fleet size have made great progress. Danzig & Fulkerson are the first researchers to model the problem of minimum number of tankers as a linear programming problem and use the simplex method to solve it[3]. Vazifeh et al. believed the size of the current taxies in New York City could be reduced by 30% through transforming individual driver decisions to a centralized operation[4]. Yao et al. used a similar approach to conclude that 128,000 shared vehicles are needed to meet travel demands of 3 million cell phone users in Shanghai, China[5].

Ride-sharing is another approach to reducing fleet size in urban areas. The concept of ride-sharing originated during the oil crisis in the 1970s[6]. The studies of Martinez & Viegas[7] and Sun[8] confirmed that sharing mode could increase average vehicle occupancy ratio and reduce urban congestion and greenhouse gas emissions. The development of information and communication technologies has opened the way to share locations and send messages in real time. Through combining multiple similar itineraries into one itinerary, ride-sharing can maximize the use of existing travel resources in society and reduce the number of vehicles in circulation[9].
In this paper, we will qualify the benefits of ride-sharing on reducing fleet size. Different from references [4] and [10], the ideal fleet size in ride-sharing will be compared with the fleet size after optimizing assignment of orders, instead of actual operation fleet size. We believe this is a more intuitive way to reflect the effects of ride-sharing.

2. Mathematical model

In this section, we will introduce two mathematical models of minimum fleet problem. The minimum fleet problem can be defined as follows: determine the minimum number of vehicles needed to serve all the travel requests specified by origin, destination and desired pick-up time.

2.1. Minimum fleet model for ride-hailing service

The model in this subsection aims to acquire the minimum number of vehicles under the situation that each vehicle can only serve the next order after serving an order. Ride-sharing is not allowed.

Each order \( r \in R \) is defined as a tuple \((T^d_r, T^p_r, l^p_r, l^d_r)\) where \( T^d_r \) represents the desired pick-up time, \( T^p_r \) the drop-off time, \( l^p_r \) the pick-up location and \( l^d_r \) the drop-off location. Here, the pick-up time \( T^d_r \) means the earliest time at which the passenger can be picked up. The drop-off time \( T^p_r \) means the estimated time of dropping off the passenger, calculated using Euclidean distance and estimated real-time speed, assuming the passenger leaves the pick-up location at \( l^p_r \) time \( T^d_r \). If the set \( R \) is extracted from a real-world dataset, \( T^d_r \) represents the actual time at which a passenger is picked up.

Figure 1(a) illustrates the construction of vehicle shareability network that enables the minimum fleet problem to be optimally solved. This is a directed network defined as \( V = (N, E) \), where node \( i \in N \) corresponds to order \( r_i \in R \) and the directed edge \((i, j) \in E\) exists if and only if

\[
(T^d_i + t_{ij}) \leq T^p_j \tag{1}
\]

\[
T^p_j - T^d_i \leq \delta \tag{2}
\]

\( t_{ij} \) in equation (1) represents the estimated travel time from the drop-off location \( l^d_i \) to the pick-up location \( l^p_j \). The existence of a link \((i, j)\) in the network corresponds to a sequence of orders that can be served by a single vehicle. \( \delta \), as a parameter in equation (2), represents the upper bound of no-load time between two consecutive orders served by a single vehicle. If no-load time of a vehicle is very long, it seems two orders that occur at distant locations or times are irrationally assigned to the vehicle. Therefore, an excessively large no-load time leads to longer travel distances, lower vehicle occupancy ratio, a lot of emissions to the environment.

Figure 1(b) indicates the optimal path cover, that is the optimal solution in minimum fleet model. In the optimal path cover, each node is covered by at most one path. The minimum fleet is equal to
least number of paths in figure 1(b), which can be transformed into the problem to include most directed edges covering all orders.

Our goal is to maximize the number of directed edges when each order is only visited once. \( y_{ij} \) is the only decision variable in the model. It is a binary decision variable. \( y_{ij} = 1 \) means the directed edge \((i, j) \in E\) is included in the optimal solution.

The objective function is given by

\[
\max \sum_{i \in N} \sum_{j \in N} y_{ij}
\]

Subject to

\[
\sum_{j \in N} y_{ij} \leq 1, \quad \forall i \in N
\]

\[
\sum_{j \in N} y_{ij} \leq 1, \quad \forall i \in N
\]

\[
y_{ij}(T_{i}^d + t_{ij} - T_{j}^p) \leq 0, \quad \forall i, j \in N
\]

\[
y_{ij}(T_{j}^p - T_{i}^d - \delta) \leq 0, \quad \forall i, j \in N
\]

\[
y_{u} = 0, \quad \forall i \in N
\]

\[
y_{u} \in \{0, , 1\}, \quad \forall i, j \in N
\]

Constraints (4) and (5) ensure that each order is included in at most one path. Orders that are not linked by directed edges will be assigned another separate vehicle to serve. Constraints (6) and (7) guarantee the rationality of directed edges, corresponding to the two conditions for the existence of directed edges. In constraints (8), self-circulation of orders is not allowed. Constraints (9) describes the value range of the decision variable.

2.2. Minimum fleet model for ride-sharing service

For ride-sharing service, each order can no longer be abstracted into a node as in figure 1. Instead, the pick-up point and the drop-off point of an order should be considered separately. The desired pick-up time and the drop-off time of an order can be taken as time windows (earliest accessible times) of the pick-up point and the drop-off point respectively.

\( R = \{r_1, r_2, ..., r_n\} \) represents a collection of orders and \( K = \{k_1, k_2, ..., k_m\} \) represents a collection of vehicles. Let \( P = \{p_1, p_2, ..., p_n\} \) denotes a collection of pick-up locations and \( D = \{d_1, d_2, ..., d_n\} \) a collection of drop-off locations. \( p(r) \) and \( d(r) \) represent the pick-up point and the drop-off point of order \( r \). The meaning of \( \delta \) is consistent with subsection 2.1.

Compared to serving only one order at a time, ride-sharing can easily cause delays in drop-off time. Let \( \varphi \) denotes maximum delay that passengers can tolerate, that is, a passenger cannot arrive at the drop-off location \( \varphi \) minutes later than ride-hailing service. The problem is defined in a directed graph \( V = (N, E) \), \( N = P \cup D \). There are 6 decision variables in the model. \( x_{ij}^{rk} \), \( y_{ij}^{r} \), \( u_{rk} \) and \( v_{k} \) are binary decision variables while \( A_{l} \) and \( L_{l} \) are continuous decision variables. \( x_{ij}^{rk} = 1 \) means vehicle \( k \) travels from location \( i \) to \( j \) carrying the passengers corresponding to order \( r \). \( y_{ij}^{r} = 1 \) indicates vehicle \( k \) travels from location \( i \) to location \( j \). \( u_{rk} = 1 \) means order \( r \) is assigned to vehicle \( k \) to serve. \( v_{k} = 1 \) indicates vehicle \( k \) has not been assigned any order. \( A_{l} \) represents the time when a vehicle arrives at location \( l \) and \( L_{l} \) represents the time when a vehicle leave location \( l \).

Our goal is to maximize the number of unused vehicles. The objective function is given by
\[ \max \sum_{k \in K} v_k \]  

Subject to

\[ \sum_{k \in K} u_{rk} = 1, \quad \forall r \in R \]  

\[ u_{rk} = \sum_{j = N} y^k_{prj}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ u_{rk} = \sum_{j = N} y^k_{jd(r)}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ \sum_{j = N} y^k_{prj} \leq \sum_{j = N} y^k_{prj}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ \sum_{j = N} y^k_{jd(r)} \leq \sum_{j = N} y^k_{jd(r)}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ \sum_{r \in N} x^p_{ij} \leq W, \quad \forall k \in K, \quad \forall i, j \in N \]  

\[ \sum_{j = N} x^p_{prj} = \sum_{j = N} x^p_{prj} - u_{rk}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ \sum_{j = N} x^d_{jd(r)} = \sum_{j = N} x^d_{jd(r)} + u_{rk}, \quad \forall r \in R, \quad \forall k \in K \]  

\[ \sum_{j = N} x^p_{ij} - \sum_{j = N} x^d_{ij} = 0, \quad \forall k \in K, \quad \forall r \in R, \quad \forall i \in P/\{p(r)\} \]  

\[ \sum_{j = N} x^d_{ij} - \sum_{j = N} x^p_{ij} = 0, \quad \forall k \in K, \quad \forall r \in R, \quad \forall i \in D/\{d(r)\} \]  

\[ x^p_{ij} \leq y^k_{ij}, \quad \forall i, j \in N, \quad \forall r \in R, \quad \forall k \in K \]  

\[ A_i - T_i - M \sum_{k \in K} \sum_{j = N} y^k_{ji} \leq 0, \quad \forall i \in P \]  

\[ T_i - A_i - M \sum_{k \in K} \sum_{j = N} y^k_{ji} \leq 0, \quad \forall i \in P \]  

\[ \sum_{k \in K} \sum_{j = N} y^k_{ji} \leq 1, \quad \forall i \in P \]  

\[ L_i \geq T_i, \quad \forall i \in N \]  

\[ L_i \geq A_i, \quad \forall i \in N \]  

\[ y^k_{ij} (A_j - L_i - t_{ij}) = 0, \quad \forall i, j \in N, \quad \forall k \in K \]
\[
A_i \leq T_i + \varphi, \quad \forall i \in N
\] (28)
\[
A_{p(r)} + t_{p(r)d(i)} \leq A_{d(r)}, \quad \forall r \in R
\] (29)
\[
L_j - L_i - \delta - M \sum_{r \in R} x_{ij}^k - M(1 - y_{ij}^k) \leq 0, \quad \forall i \in D, \forall j \in N, \forall k \in K
\] (30)
\[
\sum_{i=N} \sum_{j=N} y_{ij}^k + 1 - 2 \sum_{r \in R} u_{rk} \geq 0, \quad \forall k \in K
\] (31)
\[
\sum_{j=N} y_{ij}^k \leq 1, \quad \forall i \in N, \forall k \in K
\] (32)
\[
\sum_{j=N} y_{ji}^k \leq 1, \quad \forall i \in N, \forall k \in K
\] (33)
\[
\sum_{k \in K} y_{ij}^k \leq 1, \quad \forall i, j \in N
\] (34)
\[
y_{ij}^k = 0, \quad \forall i \in N, \forall k \in K
\] (35)
\[
v_k \times \sum_{i=N} \sum_{j=N} y_{ij}^k = 0, \quad \forall k \in K
\] (36)
\[
v_k \times \sum_{r \in R} u_{rk} = 0, \quad \forall k \in K
\] (37)

Constraints (11) indicate that each order can be served by only one vehicle. Constraints (12), (13), (14) and (15) ensure that when an order is assigned to a vehicle, the pick-up point and the drop-off point of the order are visited by the same vehicle. Constraints (16) are capacity constraints. \(W\) represents the maximum number of orders that can be served by a vehicle simultaneously. Constraints (17) and (18) reflect the changes in the number of orders on a vehicle after visiting pick-up points and drop-off points. Constraints (19) and (20) ensure that the state of each order remains unchanged after visiting pick-up points and drop-off points corresponding to other orders. Constraints (21) describe the relationship between \(x_{ij}^k\) and \(y_{ij}^k\). Constraints (22), (23) and (24) ensure that the time when a vehicle reaches the starting point of its route is equal to the earliest accessible time of the point. \(M\) is a sufficient large positive number. Constraints (25) and (26) ensure that the time when a vehicle leaves a point is the larger one of the vehicle’s arrival time and the point’s time window. Constraints (27) describe the relationship between the arrival time of a point and the departure time of its previous point. Constraints (28) require that the time a vehicle arrives at a point is no later than the latest time the customer can tolerate. In constraints (29), a drop-off point must be visited after its corresponding pick-up point. Constraints (30) ensure that the no-load time of a vehicle is less than \(\delta\), which are set for the efficient operation of vehicles. Constraints (31) guarantee each vehicle has only one starting point and one end point. Constraints (32) and (33) ensure that a vehicle cannot reach one destination from two different places, nor can it go to two different destinations from one place. Constraints (34) mean the path linking any two points can be visited by at most one vehicle. In constraints (35), self-circulation of points is not allowed. Constraints (36) and (37) describe the relationship among \(v_k, y_{ij}^k\) and \(u_{rk}\). Constraints (38) describe the value range of the decision variables.
3. Numerical experiments

In this section, we design a lot of instances based on real data and employ Gurobi to solve these instances. The results are shown in this section.

3.1. Experimental settings and computational environment

The above two models are solved by Gurobi. All computations are executed on a PC with Intel Core i5-8265U CPU (1.8 GHz) and 8 GB RAM.

The instances to be used are extracted from the dataset composed of taxi trips records originating and ending in Manhattan in January 2016. For each trip, the record reports the order ID, the Global Positioning System (GPS) coordinates of the pickup and drop-off locations, and corresponding times.

In view of complexity of the ride-sharing model and the limited computing space of the PC, instances of 20 orders are tested in the paper.

6 sets of instances are generated and their difference lies in the order density. The first set of instances is drawn from orders happening within 10 minutes on January 30, 2016 and the second set happening within 20 minutes, by that analogy, the sixth set is drawn from orders within 60 minutes.

Each set of the same order density contains 3 subsets with different parameter values. These involved parameters are listed in Table 1. The parameter $\delta$ is set to 15 minutes in 3 subsets referring to the literature[4]. These subsets adopt different values of $W$ and $\varphi$. Each subset contains 5 instances of 20 orders to ensure the stability and reliability of results.

| Parameter denotation | Parameter meaning                                           |
|----------------------|-------------------------------------------------------------|
| $\varphi$            | maximum time delay that passengers can tolerate             |
| $W$                  | maximum number of orders that can be served by a vehicle simultaneously |
| $\delta$             | upper bound of vehicle no-load time, the efficiency parameter|

3.2. The impact of ride-sharing on fleet size

| Order density | $W = 1, \varphi = 0$ | $W = 3, \varphi = 0$ | $W = 3, \varphi = 5$ |
|---------------|----------------------|----------------------|----------------------|
|               | Average fleet size   | Run time (ms)        | Average fleet size   | Run time (s)        | Average fleet size   | Run time (s)        |
| 10min         | 19.8                 | 34.8                 | 19.8                 | 3.9                 | 14.4                 | 175.6                |
| 20min         | 17.8                 | 34.2                 | 17.8                 | 5.0                 | 13.2                 | 54.1                 |
| 30min         | 16.8                 | 37.0                 | 16.8                 | 6.9                 | 12.6                 | 19.0                 |
| 40min         | 14.2                 | 34.4                 | 14.0                 | 9.2                 | 11.2                 | 111.9                |
| 50min         | 14.4                 | 35.6                 | 13.2                 | 7.8                 | 10.4                 | 135.4                |
| 60min         | 13.8                 | 34.2                 | 11.6                 | 23.5                | 9.2                  | 74.2                 |

The experimental results are showed in Table 2. $W = 1, \varphi = 0$ (situation 1) represents the situation that ride-sharing is not allowed but assignment of orders is optimized. This part of the results is benchmark for further comparison. $W = 3, \varphi = 0$ (situation 2) represents the situation that a taxi can serve at most 3 orders simultaneously and the service level is the same as the previous situation with no arrival delay. $W = 3, \varphi = 5$ (situation 3) represents the situation that ride-sharing is considered, a taxi can serve at most 3 orders simultaneously but 5 minutes’ arrival delay is allowed.
According to table 2, we can find without delaying drop-off times the fleet size when considering ride-sharing remains almost the same as ride-hailing service for high-density travel demand settings (20 orders happening during 10 minutes to 30 minutes). As the density decreases, which means there is more chance to combine some orders into a trip, the effect of ride-sharing on reducing the scale of vehicles becomes more significant. When the order time interval is equal to 60 minutes, 16% of vehicles are cut down without incurring any delay. When a slight delay ($\varphi = 5$ min) is allowed, the number of vehicles required is reduced by nearly 30% regardless of order density under ride-sharing assumptions. In terms of calculation time, run time of situation 2 is about 100 times that of situation 1 and run time of situation 3 is about 10 times that of situation 2.

4. Conclusion

In this paper, we propose two integer programming models to qualify the benefits of ride-sharing on reducing the number of vehicles. These two models are solved by Gurobi. Then we design a series of instances and perform numerical experiments to compare fleet size in ride-sharing to that in ride-hailing service. The computational experiments show that: without delaying drop-off times, the fleet size when considering ride-sharing remains almost the same as ride-hailing service for high-density travel demand settings. Whereas the fleet size drops sharply as the demand density decreases. In addition, the number of vehicles required is reduced by nearly 30% regardless of order density under ride-sharing assumptions when a slight delay is allowed.

Future study can discuss algorithms to solve large-scale minimum fleet size problem in ride-sharing, for example, exacting algorithms to quickly solve the model and heuristic algorithms to find approximate optimal solution can be taken into consideration.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under grant 71772100; Shenzhen Science and Technology Project under grant JCYJ20170412171044606; and National Key Technologies Research and Development Program under grant 2016YFB0502601.

References

[1] World Bank. (2018) Transport Overview. http://www.worldbank.org/en/topic/transport/overview.
[2] Litman T. (2007) Developing indicators for comprehensive and sustainable transport planning. Transportation Research Record, 2017: 10–15.
[3] Dantzig G, Fulkerson D. (1954) Minimizing the Number of Tankers to Meet a Fixed Schedule. Naval Research Logistics Quarterly, 1: 217–222.
[4] Vazifeh M M, Santi P, Resta G, et al. (2018) Addressing the minimum fleet problem in on-demand urban mobility. Nature, 557: 534–538.
[5] Yao Xiaorui, Wang Guan, Yang Chao. (2019) Exploring Fleet Size of Shared Autonomous Vehicles in Future City: A Case Study in Shanghai. Journal of Transportation Systems Engineering and Information Technology, 19: 85-91.
[6] Handke V, Jonuschat H. (2013) Flexible Ridesharing. Springer Berlin, 13-40.
[7] Martinez L M, Viegas J. (2017) Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. International Journal of Transportation Science and Technology, 6.
[8] Sun J. (2017) External Economic Costs of Intelligent Urban Transportation Systems: A Method to Evaluate the Externalities of Comparative Technology Adoption Pathways in the Urban Mobility Service Sector[J]. ProQuest Dissertations and Theses, 2017: 223.
[9] Machado C A S, Hue N P M de S, Berssaneti F T, et al. (2018) An overview of shared mobility. Sustainability (Switzerland), 10: 1–21.
[10] Santi P, Resta G, Szell M, et al. (2014) Quantifying the benefits of vehicle pooling with shareability networks. Proceedings of the National Academy of Sciences of the United States of America, 37: 13290–13294.