Bike Sharing Systems: Does demand forecasting yield a better service?

Selin Ataç[a], Nikola Obrenović[b], Michel Bierlaire[a]

[a] *EPFL Transport and Mobility Laboratory*
[b] *University of Novi Sad, Faculty of Technical Sciences, Department of Power, Electronic and Telecommunication Engineering*

Abstract

Although the idea of vehicle sharing systems (VSSs) emerged back in 1940s, sustaining such a system became simpler with the improvements in technology in the past decade. In addition to that, people have become more concerned about environmental effects and try to find solutions on reducing emission and energy consumption. On the other hand, VSSs require effort to make them profitable. In this paper, we focus on the two of the operational level challenges, which are the demand forecasting and routing for the rebalancing operations, in a one-way station-based bike sharing system (BSS). Since the data collection is exhaustive and costly, we would like to find the answer to whether it is worth to collect data and develop demand prediction models. In order to do that, we create a simulation of a city BSS in operation during the day. Then, using a mathematical model from the literature, we assess the rebalancing costs under two scenarios: one where we assume the perfect demand forecast, and the other where the future demand is unknown. By this way, we determine the trade-off between the lost demand and the rebalancing cost under the mentioned scenarios, and assess the benefit of forecasting the demand. Lastly, we present a case study on the Swiss BSS named PubliBike.

Keywords: Transportation, Bike sharing systems, Forecasting, Rebalancing operations

---

1. Introduction

The idea of sharing vehicles arose in the late 1940s with cars. The first known CSS, Selbstfahrergemeinschaft, was initiated in Zurich, Switzerland, in 1948 [1]. This attempt was more because of convenience than of profitability. The idea came up again in the early 1970s. After the first attempts that took place in Amsterdam and France, which were mostly based on the economic and environmental reasons, profit-based companies saw the opportunity in these systems and invested money [1, 2]. Consequently, not only user convenience but also the profitability of the company also became an issue to discuss. These conflicting objectives attracted the research community's attention and the focus on sharing systems increased substantially.

A VSS has several kinds of configurations. The type of trips can be either return or one-way. In the former, the user is supposed to drop the vehicle off to the pick-up station after the usage. The latter, on the other hand, allows the user to park anywhere designed in the city. One-way trips are much more flexible from the user perspective, however, it brings the problem of imbalance. The vehicles tend to be accumulated in the stations or areas which are mainly trip destinations, which often causes a lack of vehicles in the frequent trip origins.

To overcome imbalance, the operators introduce rebalancing operations which relocate the vehicles from stations with high availability and low demand to the stations with high demand and low availability to satisfy user demand at a higher level. The rebalancing operations might take place at a different time of the day in different VSSs. Static rebalancing is done when there is the least demand, generally during the night, every day. Dynamic rebalancing, on the other hand, is flexible and executed throughout the day. Moreover, in some VSSs the rebalancing is done by the staff relocating the vehicles or by staff using relatively big vehicles, e.g. trucks, which carry the vehicles.

In order to be able to rebalance vehicles which improves the level of service, the operators aim to forecast the demand for the following time steps. This lets the operator to anticipate the demand in order to perform rebalancing in due time. This close relationship between the demand forecasting and rebalancing operations leads us to further examine the trade-off between them. In this work, we aim to provide a framework that incorporates these two aspects of a VSS and analyze the added value of demand forecasting.

The rest of the paper is organized as follows: Section 2 reviews the existing literature on rebalancing operations and demand forecasting in particular. Then, the proposed methodology is discussed in Section 3 and the results on a case study are presented in Section 4. Finally, Section 5 concludes the paper.

2. State of the art

In this section, we talk about the studies that belong to our context. A brief literature review on the rebalancing operations is presented in Section 2.1 and on the demand forecasting in Section 2.2.

2.1. Literature on rebalancing operations

The most of the research was based on static rebalancing operations. One of them was addressed by [3] under the setting of station-based BSSs. The stations had docks for bikes, which implies that they were capacitated. The research question is how to apply static rebalancing operations in the defined system. To do that two mixed integer linear programming (MILP) formulations are developed, which are extensions to one-commodity pickup and delivery traveling salesman problem (1-PDTSP). Valid inequalities and dominance rules are proposed for these formulations. The objective function is minimizing the operating costs. In addition to that, it takes the user satisfaction into account as well as the loading and unloading times. The authors in [4] present a solution method for the same problem using a constraint programming (CP) approach.

Another approach for the static rebalancing with capacitated vehicles was introduced in [5]. To solve this problem the authors present four MILP formulations. These models are strengthened by introducing valid inequalities. Furthermore, because of the exponential number of constraints, tailor-made branch and cut algorithms are developed. Although the approach taken to the problem was not innovative, introduction of different formulations and allowing relocation vehicles to do both pick up and drop off at stations in one route contributed the literature.

As opposed to operator-based rebalancing operations, it is also possible to keep the balance of the system using user-based operations. In order to encourage users to rebalance the system, dynamic pricing in which one's trip characteristics, such as origin, destination, time of the day, and trip duration, are also taken into account while determining the price can be used. Such a work was conducted in [6], on a one-way CSS. This system uses pricing incentives to encourage users to rebalance the system. The level of service (LOS), which is the ratio of number of served users and the number of vehicle requests, is used as a measure to evaluate the performance of the system. The authors propose a simulation-based optimization approach, where the optimization module determines the optimal thresholds of available vehicles at which the incentives should be introduced, while the simulation module evaluates the stochastic user response to the offered incentives and how the CSS operates.

The pure user-based rebalancing strategies might not be enough to keep the balance of the system at the desired level and still need the operator-based operations. In [7], the authors aim to combine the dynamic pricing and dynamic online rebalancing operations in a station-based BSS. A predictive model is used to see the future demand. For the routing algorithm, they use a time-expanded network (as in [8]) and develop a mixed integer quadratic program (MIQP). This solution is provided only for the single-truck and a heuristic is proposed for the multi-truck case.

We see from the literature that the rebalancing operations are essential to keep the balance regardless of the kind of vehicles used in the system. Mathematical models, simulation, heuristics and metaheuristics are utilized to propose a solution to rebalancing operations. The reader may refer to more thorough literature surveys in [9, 10].

2.2. Literature on demand forecasting

One of the research questions related to demand forecasting was to identify the important factors of the corresponding VSS. In works [11, 12], the authors consider a station-based BSS. In [11], the authors aim to understand the bike count's behavior by including the weather conditions. Two models are developed based on Poisson Regression Model (PRM) and Negative Binomial Regression Model (NBRM). In [12], the authors use the data related to weather conditions, temporal variables, station attributes, socio-economic characteristics of the users to analyze the number of pick-up and drop-offs. Since traditional linear regression assume independence between the observations, they use the random intercept multilevel modeling to identify the factors affecting the ridership.

The regression models were not the only methodology used to forecast the demand. For example, in [3, 4], the authors include a Markov chain structure to estimate the number of vehicles per station in the steady state. On the other hand, the authors in [13] start by creating a data set which identifies the user arrivals and departures per station, as well as rebalancing operations. Authors test assumptions about the factors that influence customer arrivals and departures and rebalancing refill and removal. They apply a methodology for analyzing such systems using behavioral models, in particular, autoregressive moving average (ARMA) models. Finally, they present the first empirical analysis of system rebalancing by the operator focused on understanding the factors creating such imbalances, using an approach consisting of a binary logit model, for identifying stations that need rebalancing, and a linear regression model for the amount of rebalancing. This analysis can help in creating plans for rebalancing well in advance, as well as in creating incentive mechanisms for customers to rebalance bikes.

As well as analyzing the system and optimizing parameters, the data collection is also an important part to manage VSSs. As an example, [14] conducted a stated preference survey in Beijing, China, for a station-based BSS, in which users can take conventional or electric bikes. Then, a multinomial logit model is developed to model mode-choice. On the other hand, we did not find any works which provide the value of demand forecasting in VSSs. In other words, the upper bound on the cost of demand forecasting operations, such as collecting data and processing, is not discussed in the literature. This work puts the first and simplistic attempt to find an answer to this question.

3. Methodology

The proposed framework includes two main modules: simulation and optimization. The former simulates the events taking place in the VSS during a day and the latter takes care of the optimization of the rebalancing operations by minimizing the total operational cost. After initialization of the parameters, the simulator simulates the day and passes the final configuration of the day to the optimization module. Then, the difference between the number of bikes at the station at the end of the previous day and the number of bikes which is desired at the beginning of the next day, demand of a station, is calculated and given as an input to the optimization module. The desired initial configuration of the next day is passed to the simulation module and the simulation of the next day is triggered. These two modules feed each other in terms of information (Figure 1).

To mimic the real-world, a discrete event simulator is developed. The events of the simulator are designed so that the simulator is adaptable to any kind of VSS. Sections 3.1 and 3.2 give the details on the simulator and optimization modules, respectively.

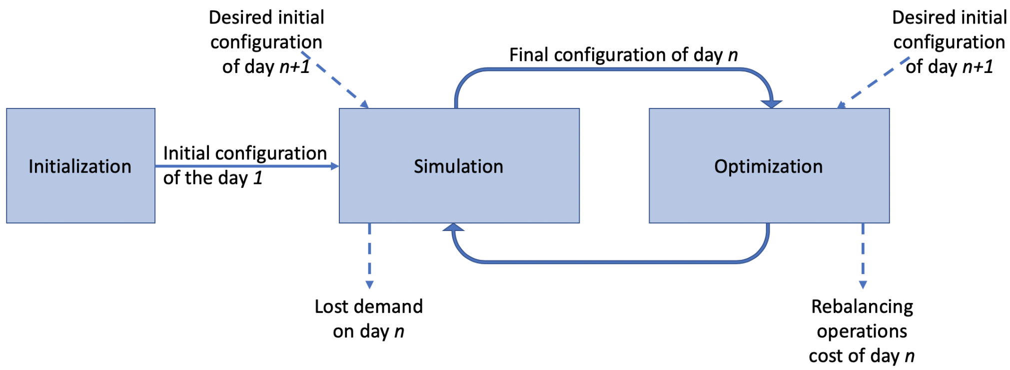


Figure 1: The framework for VSS simulation and rebalancing optimization

3.1. Simulation

The simulation module consists of four event types, namely `REQUEST`, `PICKUP`, `DROPOFF`, and `COMPLETED`. The number of people in the system (ns) and the number of people using a vehicle (nu) at that time are recorded as indicators. The triggered events are added to the event list which is in chronological order. Appendix A provides further details on the events and Table A.1 summarizes the event types, the triggered event(s) by each event and the change in event queue status.

The state variable of the system is time, denoted as t, and the time horizon is T. The time is not discretized but drawn for each event according to the Poisson distribution. Within $[0, T]$, O-D pair requests arrive to the system. After T, the events in the system are served and no more O-D pair requests are generated. We denote the number of stations by N. C_i, for $i = 1..N$, represents the capacity of station i. The distance from station i to station j with mode k is denoted as c_{ij}^k, where $i = 1..N$, $j = 1..N$, and $k = \{$'walking', 'bicycle', 'car'$\}$.

The time horizon is divided into P number of time windows, each denoted TW_p, where $p = 1..P$. This differentiation makes the simulator flexible at the temporal level to test different behaviors during the day, such as rush hours and specific event times. Therefore, the O-D pair request rates are also specific to these time windows. λ_p provides the information on the rate of requests for time window p, where $p = 1..P$.

Two different cases are compared: unknown and known demand. For the unknown demand case, we assume that no information regarding the next day's O-D pair request is known. Therefore, the vehicles are distributed equally every day. For the known demand case, the distribution of the bikes are done according to next day's O-D demand behavior regardless of the time of the request. The difference between the total number of pick-ups and drop-offs for each station is calculated, and normalized to the total number of vehicles in the system. In other words, if for a station the number of pick-ups are more than the number of drop-offs, it gets more bikes; if not, less bikes are assigned to that station. It is essential to note that the future work includes the investigation of known demand case where the spatial and temporal information on O-D pairs rather than the absolute difference of pick-ups and drop-offs are used.

3.2. Optimization

Among many mathematical formulations, we utilize one of the models, namely F1, which was introduced in [5]. They consider a station-based BSS and apply static rebalancing at the end of the day. They define the problem as a generic one and assume that the initial configuration for the next day is a parameter. Since the availability of the simulator provides the full information on O-D pair requests of the next day, we can utilize this formulation. It should be noted that, the authors report that formulation F3 provides better results than F1 in terms of computational time. However, we select F1 due to its convenience to modify the subtour elimination constraints. The reader is referred to the work [5] for the full notation of parameters and decision variables.

Given the exponential number of constraints, the model becomes intractable for large instances. The classical subtour elimination constraints that are used in F1 corresponds to Dantzig-Fulkerson-Johnson (DFJ) formulation [15]. This formulation introduces 2^{n+1} number of constraints where n is the number of stations. In [16], the authors introduce a new formulation, i.e. Miller-Tucker-Zemlin (MTZ), using additional decision variables and decrease the number of constraints to $(n+1)^2$. In order to overcome the computational burden, this work provides an extension to their formulation by utilizing the MTZ constraints, constraints (6) and (7) and uses the valid inequalities proposed by [5], constraints (12) and (13). These valid inequalities ensure that if three nodes have a total supply/demand larger than the capacity, there is no feasible solution going through them consecutively.

The modified model is given in $F1_m$. Please note that the constraint set (14) is added to prevent visiting the same node consecutively. This could have also been achieved by modifying the cost matrix.

$$(F1_M)\min \quad \sum_{i\in V}\sum_{j\in V} c_{ij}x_{ij} \tag{1}$$

$$\text{s.to} \quad \sum_{i\in V} x_{ij} = 1 \qquad \forall j \in V \setminus \{0\} \tag{2}$$

$$\sum_{i\in V} x_{ji} = 1 \qquad \forall j \in V \setminus \{0\} \tag{3}$$

$$\sum_{j\in V} x_{0j} \leq m \tag{4}$$

$$\sum_{j\in V\setminus\{0\}} x_{0j} - \sum_{j\in V\setminus\{0\}} x_{j0} = 0 \tag{5}$$

$$u_i - u_j + n * x_{ij} \leq n - 1 \qquad \forall i, j \in V \setminus \{0\} \tag{6}$$

$$1 \leq u_i \leq n \qquad \forall i \in V \tag{7}$$

$$\theta_j \geq \max\{0, q_j\} \qquad \forall j \in V \tag{8}$$

$$\theta_j \leq \min\{Q, Q + q_j\} \qquad \forall j \in V \tag{9}$$

$$\theta_j - \theta_i + M(1 - x_{ij}) \geq q_j \qquad \forall i \in V, j \in V \setminus \{0\} \tag{10}$$

$$\theta_i - \theta_j + M(1 - x_{ij}) \geq q_j \qquad \forall i \in V \setminus \{0\}, j \in V \tag{11}$$

$$x_{ij} + \sum_{h\in S(i,j)} x_{jh} \leq 1 \qquad \forall i, j \in V \setminus \{0\}, h \in S(i, j) \tag{12}$$

$$\sum_{h\in S(i,j)} x_{hi} + x_{ij} \leq 1 \qquad \forall i, j \in V \setminus \{0\}, h \in S(i, j) \tag{13}$$

$$x_{ii} = 0 \qquad \forall i \in V \tag{14}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i, j \in V \tag{15}$$

4. Computational experiments

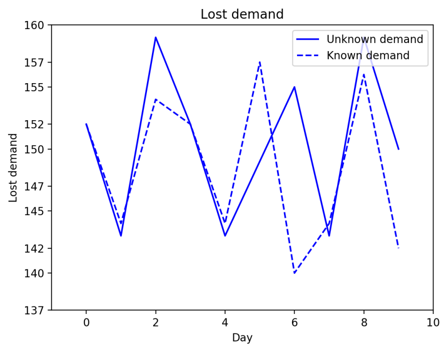
The simulation is implemented on a machine with 8 GB RAM and 2.3 GHz Intel Core i5 processor in *python* and *python* API for CPLEX 12.9 is used to solve the optimization model. The constructed environment includes the station information from Lausanne-Morges district of PubliBike BSS from Switzerland, which has a station-based, one-way configuration and deploys static rebalancing. We assume that static rebalancing is done at the end of every day.

For random demand scenario, the demand location for an O-D pair is generated randomly and accepted as a valid location for a request if it lies in 20 minutes' walk to a station. The Lausanne-Morges district has 35 stations and 180 bikes in total. Since the stations of PubliBike do not have lockers, it is possible to leave the bike regardless of the number of bikes existing in that station. Therefore, the capacity of each station is set to infinity. λ_p, for all $p = 1..P$, is equal to 20 requests per hour. This value does not have a rationale for now, however in case of availability of data from the corresponding system, it is possible to include this information with the time windows. For the scenarios which take spatial differences, i.e. difference in altitude, into account less demand is generated for the uphill trips compared to downhill ones. For Lausanne-Morges case study, spatial differences are important since the city has many uphill and downhill rides.

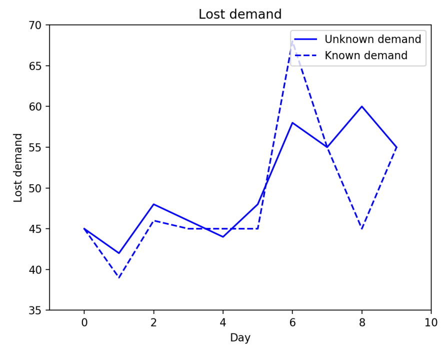
Each scenario is generated and used for both known and unknown demand cases to compare between the lost demand and rebalancing operations cost. The objective function is built with the cost value being the distance traveled by a truck carrying the bikes. The capacity of a relocating vehicle, Q, is set to 20 bikes and the number of such vehicles, m, is set to 4. Since we are interested in the evaluation of the added value of demand forecasting, the number of lost demand and the total number of O-D pair requests are presented along with the rebalancing costs. Lost demand corresponds to the number of users who opt-out because of unavailability of bikes.

The computational time burden results from the mathematical model. However, with 35 stations it is still solvable in reasonable time. We first investigate the effect of knowing the O-D demand on lost demand. For spatially randomly distributed case (Figure 2a), we see that it is difficult to differentiate between the unknown and known demand cases. On the other hand, when we take the spatial differences into account (Figure 2b), unknown demand results in losing more demand in general. The results show a tendency in decreasing the number of users who opt-out between the unknown and known demand case. It is worth to note that these results solely rely on the simulation parameters. Different scenarios might lead to different results.

From Figure 3, we see that the main cost contributor to the operational cost is the number of trucks. As the number of trucks decreases the route length of the rebalancing operations increases substantially. It can also be deduced that the demand knowledge does not necessarily yield less rebalancing cost.

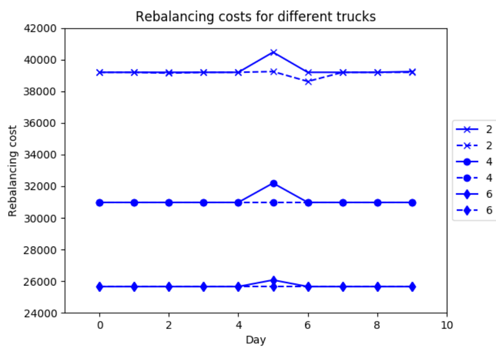


(a) Lost demand in the case of known and unknown demand when the demand is distributed spatially equally

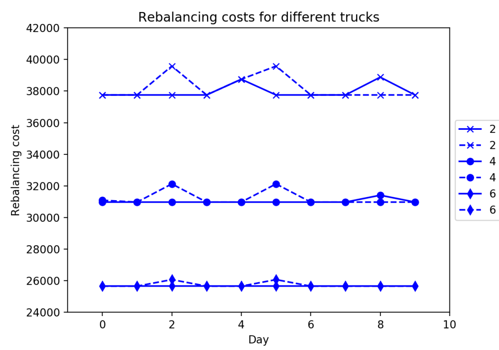


(b) Lost demand in the case of known and unknown demand when spatial differences are taken into account

Figure 2: Day vs Lost demand



(a) Rebalancing cost in the case of known/unknown demand and different number of trucks when the demand is distributed spatially equally



(b) Rebalancing cost in the case of known/unknown demand and different number of trucks when spatial differences are taken into account

**Dashed lines correspond to known and full line to unknown demand*

Figure 3: Day vs Rebalancing cost and number of trucks

We also analyzed whether there is a relation between the lost demand and the cost of rebalancing operations. The results show that there is no clear pattern between these two performance measures. All in all, in our preliminary experiments, we see a slight benefit, in terms of the less demand loss, coming from the knowledge of future demand. However, because of insufficient number of experiments, a statistical test on the significancy of this decrease cannot be applied. Therefore, the results cannot be generalized.

5. Conclusion and future work

Like in many fields of research, the demand forecasting process is one of the several challenges in the context of VSSs. Although there are many studies worked on demand forecasting, none set an upper bound on the cost of such operations. Therefore, this work presents an early attempt to determine the value of demand forecasting. Optimization and simulation modules are developed to include two aspects of the operations: supply and demand. By this way, we are able to investigate the trade off between the cost of rebalancing operations and lost demand in the case of known and unknown demand scenarios. Computational experiments are performed on a case study in Lausanne-Morges district of PubliBike, which is a BSS in Switzerland. The results show that the cost of rebalancing operations mainly depend on the number of trucks used. On the other hand, we see a slight improvement in the number of lost demand with the demand knowledge. No clear relation between the lost demand and rebalancing costs is found yet, but we will further investigate in this direction by increasing the number of experiment repetitions.

The future work includes the development of another simulation module which mimics the rebalancing operations. By this way, we will be able to see whether the results of the optimization module can be applied in real life perfectly. Moreover, the demand distribution is an important element in this framework. Following work also aims to analyze different scenarios. At this stage of the research, only the results for BSSs is presented whereas the simulator is easily adaptable to CSSs. The future work also includes the extension of the BSS simulator and adapting it to a CSS.

6. References

[1] S. Shaheen, D. Sperling, and C. Wagner, "Carsharing in europe and north america: Past, present, and future," *Transportation Quarterly*, vol. 52, 06 1998.

[2] S. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in europe, the americas, and asia: Past, present, and future," *Institute of Transportation Studies, UC Davis, Institute of Transportation Studies, Working Paper Series*, vol. 2143, 01 2010.

[3] T. Raviv, M. Tzur, and I. A. Forma, "Static repositioning in a bike-sharing system: models and solution approaches," *EURO Journal on Transportation and Logistics*, vol. 2, pp. 187–229, 2013.

[4] J. Schuijbroek, R. Hampshire, and W.-J. van Hoeve, "Inventory rebalancing and vehicle routing in bike sharing systems," *European Journal of Operational Research*, vol. 257, no. 3, pp. 992 – 1004, 2017.

[5] M. Dell'Amico, E. Hadjicostantinou, M. Iori, and S. Novellani, "The bike sharing rebalancing problem: Mathematical formulations and benchmark instances," *Omega*, vol. 45, pp. 7 – 19, 2014.

[6] M. Clemente, M. P. Fanti, G. Iacobellis, M. Nolich, and W. Ukovich, "A decision support system for user-based vehicle relocation in car sharing systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, pp. 1–14, 07 2017.

[7] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari, "Dynamic vehicle redistribution and online price incentives in shared mobility systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 1567–1578, Aug 2014.

[8] B. Boyacı, K. G. Zografos, and N. Geroliminis, "An integrated optimization-simulation framework for vehicle and personnel relocations of electric carsharing systems with reservations," *Transportation Research Part B: Methodological*, vol. 95, pp. 214 – 237, 2017.

[9] S. Illgen and M. Höck, "Literature review of the vehicle relocation problem in one-way car sharing networks," *Transportation Research Part B: Methodological*, vol. 120, pp. 193 – 204, 2019.

[10] G. Laporte, F. Meunier, and R. Calvo, "Shared mobility systems: an updated survey," *Annals of Operations Research*, vol. 271, 12 2018.

[11] H. I. Ashqar, M. Elhenawy, and H. A. Rakha, "Modeling bike counts in a bike-sharing system considering the effect of weather conditions," *Case Studies on Transport Policy*, vol. 7, no. 2, pp. 261 – 268, 2019.

[12] D. M. Scott and C. Ciuro, "What factors influence bike share ridership? an investigation of hamilton, ontario's bike share hubs," *Travel Behaviour and Society*, vol. 16, pp. 50 – 58, 2019.

[13] A. Faghih-Imani, R. Hampshire, L. Marla, and N. Eluru, "An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville," *Transportation Research Part A: Policy and Practice*, vol. 97, pp. 177 – 191, 2017.

[14] A. A. Campbell, C. R. Cherry, M. S. Ryerson, and X. Yang, "Factors influencing the choice of shared bicycles and shared electric bikes in beijing," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 399 – 414, 2016.

[15] G. Dantzig, R. Fulkerson, and S. Johnson, "Solution of a large-scale traveling-salesman problem," *Journal of the operations research society of America*, vol. 2, no. 4, pp. 393–410, 1954.

[16] C. E. Miller, A. W. Tucker, and R. A. Zemlin, "Integer programming formulation of traveling salesman problems," *J. ACM*, vol. 7, p. 326–329, Oct. 1960.

Appendix A. Simulation: Events

The `REQUEST` event is generated at the beginning of the simulation for each O-D pair request. These `REQUEST` events form the initial event list which is kept in chronological order and traversed one by one. By this way, the simulator keeps track of the time. As soon as a `REQUEST` event is observed in the event list ns is increased by 1 and a `PICKUP` event is generated if there is a station with a positive number of vehicles within walking distance. If there are no vehicles at the stations within walking distance, the user opts-out, and the ns is decreased by 1.

When a PICKUP event appears in the event list, a DROPOFF event, which consists the desired drop-off station information, is generated if there is at least one vehicle in that station and nu is increased by 1. Otherwise, the user opts-out and leaves the system. In this case, we decrease ns by 1.

The DROPOFF event triggers either another DROPOFF event or a COMPLETED event. The user chooses a drop-off station according to his/her destination location. However, in some cases, the user might not be able to find an available parking spot there. Then, another DROPOFF event is triggered and the user tries the next closest available station. If there is at least one vehicle available at the corresponding drop-off station, then a COMPLETED event is generated and nu is decreased by 1. The COMPLETED event is removed from the queue as soon as the user reaches the destination point which also makes him/her leave the system, i.e. ns is decreased by 1.

Table A.1: Event types

Event	**Triggered Event**	**Queue**
`Sim Start`	`REQUEST`, `Sim End`	-
`REQUEST`	`REQUEST` (if $t < T$), `PICKUP` (if there is an available station within 20 min walk)	$ns = ns + 1$ -
`PICKUP`	`DROPOFF` (if there are available vehicles)	$nu = nu + 1$
`DROPOFF`	`DROPOFF` (if no parking available), `COMPLETED`	- $nu = nu - 1$
`COMPLETED`		$ns = ns - 1$
`Sim End`		-