Parking Reservation Policies in One-Way Vehicle Sharing Systems

Mor Kaspi, Tal Raviv, Michal Tzur*

Department of Industrial Engineering, Tel-Aviv University, Israel

In this study, we propose improving the performance of one-way vehicle sharing systems by incorporating parking reservation policies. In particular, we study a parking space reservation policy in which, upon rental, the users are required to state their destination and the system then reserves a parking space for them until they arrive at their destinations. We measure the performance of the vehicle sharing system by the total *excess time* users spend in the system. The excess time is defined as the difference between the actual journey time and the shortest possible travel time from the desired origin to the desired destination. A Markovian model of the system is formulated. Using this model, we prove that under realistic demand rates, this policy improves the performance of the system. This result is confirmed via a simulation study of a large real system, Tel-O-Fun, the bike-sharing system in Tel-Aviv. For all the tested demand scenarios, the parking reservation policy reduces the total excess time users spend in the system, with a relative reduction varying between 14% and 34%. Through the simulation we examine additional service-oriented performance measures and demonstrate that they all improve under the parking reservation policy.

Keywords: Vehicle Sharing, Bike-Sharing, Reservations, Simulation

1. Introduction

Public transportation in modern cities is typically composed of bus systems, trams, subway and taxi services. Such a mix can satisfy most of the citizens' travel demands, but nevertheless many citizens still prefer to use private vehicles. This can be attributed to the fact that public transportation is usually limited as far as service areas (depending on the planned lines), operating hours, and service frequency. A private vehicle is available at any time and allows greater flexibility such as: out of city trips, traveling with cargo (groceries, baby carriages etc.). In addition, a private vehicle may be the choice of people who do not live within walking distance of public transportation stations (the 'last mile' problem) or who merely prefer not to travel with strangers. However, the cost of owning a private vehicle is typically higher compared to using public transportation. Moreover, the average car is used for less than an hour daily. The rest of the time it sits idle and takes up a parking space. This represents low utilization of highly needed resources (both the vehicle and the parking space), especially in city centers.

To fill the gap between the advantages of using private vehicles and the services offered by public transportation, in recent years, many cities around the world have developed station-based one-way vehicle sharing systems such as car-sharing and bike-sharing systems. One-way systems allow users to rent a vehicle at one of the system's stations scattered around the city, use it for a short time

*Corresponding author at: Department of Industrial Engineering, Tel-Aviv University, Israel. Tel.: +972-3-6407420
E-mail addresses: morkaspi@post.tau.ac.il (M. Kaspi), talraviv@eng.tau.ac.il (T.Raviv), tzur@eng.tau.ac.il (M. Tzur)

period and return it to any of the system’s stations. A station is a group of parking spaces where the vehicles are parked while not in use, in some cases next to a terminal through which the service is provided. We note that some one-way vehicle-sharing systems are operated as free-floating, i.e. a vehicle may be parked anywhere within the boundaries of the city. Such systems are not in the scope of this study. Our study focuses on station-based one-way vehicle sharing systems, although we sometimes omit, for brevity, the phrases 'station-based' and 'one-way'. For a review of the history of vehicle sharing systems, description of the various business models and prospects for the future, see Shaheen and Cohen (2007), DeMaio (2009) and Shaheen et al. (2010).

For example, a commuter who traveled in the morning using the subway and wishes to do some shopping at the end of the work day, can rent a car near her work and return it to a station located near her home. A user, who does not live within walking distance from a public transportation station, may be able to rent a bicycle at a nearby station and return it near the public transportation station.

Vehicle sharing systems typically rely on information and communication technology, such as 3G communication and GPS, to allow the operators and the users to check the availability, location and status of each vehicle in the system online. The wide spread use of smartphones increased the usability of vehicle sharing systems by making this information accessible to the system’s users any time, any place.

In recent years, there has been a rapid increase in the deployment of car sharing and bike sharing systems around the world. As of October 2012, car sharing systems were operating in 27 countries, accounting for an estimated 1,788,000 shares over 43,500 cars (Shaheen and Cohen, 2012). A notable one-way car-sharing system is Autolib, currently operating in Paris with 1,800 electric cars, 4,000 parking spaces and more than 65,000 users (Kanter, 2013). In May 2011 there were an estimated 136 bike sharing programs, with 237,000 bicycles on the streets (Shaheen and Guzman, 2011). As of February 2013, these numbers were significantly higher, with about 500 active bike sharing systems worldwide and more than 150 in planning, with over 600,000 bicycles in use in the active systems (DeMaio and Meddin, 2013).

Due to the implementation of car-sharing systems, more citizens may give up their private vehicles and switch to public transportation. Surveys conducted by car-sharing companies and cities who implemented such systems have shown that the use of car-sharing systems has not only reduced the number of vehicles on the roads but it also reduced the number of executed trips, as users changed their traveling habits (Zhao, 2010). Hence, car-sharing systems may assist in reducing city traffic congestion and improving utilization of city land resources, as the demand for wider roads and parking spaces may decrease. Furthermore, many car-sharing systems are based on electric cars which are less polluting. Bike-sharing systems are clearly also environment friendly vehicles. Although the prevailing approach is that bike-sharing systems promote the use of public transport by providing a solution for the “last mile” problem (Martens, 2007 and Shaheen et al., 2010), other studies argue that bike-sharing systems compete with public transport on the same share of users, see Midgley (2011) and Kumar et al. (2013).

A distinction should be made between round-trip vehicle-sharing systems and one-way vehicle-sharing systems. While in the former a user must return the vehicle to the same station in which it was rented, in the latter, a user may return the vehicle to any of the system's stations. Clearly, the flexibility of one-way systems makes them more useful for their users. However, this flexibility creates an intricate challenge to the operators due to the need to redistribute vehicles in the system in order to meet the demand. For this reason, until recent years, most of the car sharing systems provided only round-trip service. Metropolitan bike sharing systems, however, are typically deployed as one-way systems since physical redistribution of bicycles can be carried out more easily

Although physical redistribution would be done quite differently, there are many similarities between bike sharing systems and one-way car sharing systems. In particular, the mechanism that we propose in this study can be implemented in both types of systems in a similar manner and therefore in the sequel we refer to both as simply vehicle sharing systems. A comparison between the two with respect to systems characteristics and factors affected (transport, social-environment and personal) can be found in Efthymiou et al. (2013).

The main challenge faced by one-way vehicle sharing system operators is to satisfy demands for vehicles (upon rental) and for vacant parking spaces (upon return). The demands are typically stochastic, non-stationary and asymmetric processes. Occasionally, users may find that some stations are empty (no available vehicles) and some are full (no available parking spaces). If a vehicle is not available at the desired origin of the journey, the user may either abandon the system, possibly use other means of transportation, or she may look for an available vehicle in a neighboring station. If, on the other hand, a parking space is not available at the destination, the user is obliged to find a station with available space in order to return the vehicle to the system. Typically, the system performance is measured as a function of these two types of undesired situations. We note that although users can check online the availability of vehicles and parking spaces, this does not guarantee that their requests would be satisfied, because by the time the user arrives at her destination, the space might be already occupied.

Two main approaches can be taken to reduce the occurrences of such situations: adding more resources to the system or redirecting demands. Adding more resources may include setting up more stations, enlarging the station capacities and adding more vehicles to the system. The following studies focus on related strategic planning aspects of vehicle sharing systems such as the location of stations and the number of vehicles to be dispersed in the system: Lin and Yang (2011), Lin et al. (2011) and Correia et al. (2012) take a deterministic centralized approach, as often done in studies of traditional logistics networks (see, for example, Shu et al. 2005), while George and Xia (2011) and Shu et al. (2013) formulate stochastic models and assume users are strategic, i.e. a decentralized approach (similar to the approach we take in this paper). In addition, some operators actively redistribute the vehicles in the system by moving vehicles between stations. This mode of operation is referred to in the literature as repositioning (or redistribution), and is further classified into static repositioning and dynamic repositioning. Static repositioning operations are carried out during off-peak periods when demands are negligible (at night) in order to prepare for the demands of the following day, see Kek et al. (2009), Nair and Miller Hooks (2011), Benchimol et al. (2011), Chemla et al. (2012), Raviv et al. (2013), Angeloudis et al. (2012), Forma et al. (2013) and Erdoğan et al. (2012). Methods for calculating the inventory target levels at the beginning of the day are proposed by Raviv and Kolka (2013) and Schuijbroek et al. (2013). In most systems, repositioning operations are carried out also during peak periods, while bicycles are on the move. Such operations are referred to as dynamic repositioning, see Contardo et al. (2012) and Pessach et al. (2012). Note that repositioning is merely another way of adding resources (workforce and redistribution trucks) in order to meet the demand; this approach is not in the scope of this study.

Redirecting demands may be accomplished, for example, by introducing incentives to users, see Fricker and Gast (2013), Waserhole and Jost (2012) and Di Febbraro et al. (2012), or by establishing system regulations such as reservations, which is the approach considered in this paper. Both incentives and regulations may be seen as means of passively redistributing vehicles in the system.

Reservations are often used to coordinate supply-and-demand mismatches in various systems. They provide a control over the flow of demands and, in addition, allow a better forecast of the system's future state. Due to their ability to smooth demand peaks and reduce uncertainty, this mechanism is widely used by many service providers such as hotels, restaurants, airline companies, healthcare facilities, etc.

In this study, we focus on the reservations of parking spaces at the destinations. We consider the following policy: when a user rents a vehicle, she declares her destination station and a vacant parking space in that station is reserved for her, if one is available. That is, the reserved vacant parking space at the destination becomes unavailable for other users from the moment the reservation is made, until the user reaches her destination. This assures that when she reaches her destination, she will be able to return her vehicle. If there is no available parking space at the destination, the transaction is denied, so that the user is unable to rent the vehicle. We refer to this policy as the *Complete-Parking-Reservation* (CPR) policy. It is complete in the sense that all users are obliged to reserve their parking space. By implementing this parking reservation policy, the system can secure an ideal ride for some users. On the other hand, it holds parking spaces without utilizing them for a certain amount of time, and by doing so, possibly rejecting other potential users. Due to this tradeoff, the effectiveness of using reservations is not obvious.

Technologically, the implementation of parking reservations requires mainly software updates, which need to be adjusted to include the above-mentioned renting process, and then present only empty and non-reserved parking spaces as available. As for the hardware, specific vehicle identification is already in place in some existing systems, therefore returning a vehicle to a reserved parking space may be allowed by the system only if the vehicle is identified as the one for which the reservation was made. In some systems (mainly bike-sharing), different light colors are used to signal whether a parking space (locker) is operative or out of order. An additional color can be used to signal that a certain parking space is reserved.

The contribution of this study is as follows: for the first time, we study a parking reservation policy in one-way vehicle sharing systems and compare its performance to a base policy, entitled No-Reservation (NR) policy. We propose measuring the performance of the system by the total excess time users spend in the system due to unfulfilled renting requests or delays in returning the vehicles, where the excess time of a user is the difference between her actual and ideal (shortest) travel time. We show that implementing parking reservations as suggested in this study will improve the performance of one-way vehicle sharing systems in most realistic systems. In particular, using Markovian models, we prove our main theoretic result, stating that if the demand faced by the system is not extremely high, the CPR policy outperforms the NR policy. In other words, under the CPR policy users will spend less excess time in the system. We further demonstrate through a small example that in some cases the CPR is superior for any demand rate. Finally, the main theoretic result is reinforced by a numerical study based on a discrete event simulation of a real-world vehicle sharing system with an enhanced user behavior model. The bike sharing system in Tel-Aviv, Tel-O-Fun, was used as our case study. The behavior model assumes users are strategic, i.e., in case of an unfulfilled renting request or unavailability of a parking space, the user determines her alternative route so as to minimize the expected time she spends in the system. Through this numerical study, we also demonstrate that the CPR policy outperforms the NR policy under several additional performance measures.

The structure of this paper is as follows. In Section 2 we discuss service oriented performance measures used in vehicle sharing systems and in public transportation in general. In Section 3, a Markovian model of a vehicle sharing system is described. In Section 4, the NR and CPR policies are compared analytically and numerically. A simulation model of a vehicle sharing system is presented in Section 5. Using the model, the superiority of the CPR policy is demonstrated using data from a real world system. In Section 6, concluding remarks and some directions for future research are given.

2. Performance measures

Most of the one-way vehicle sharing systems are service oriented; accordingly, these systems are typically measured by the quality of service given to the users. The most common measure used in practice is based on the percentage of time in which stations are empty or full. Specifically, it is obtained by averaging these percentage values over all stations. We refer to this measure as *station-availability*. Since the vehicle inventory levels in the stations are updated on-line in the information systems, it is quite easy for the operators or other interested parties to monitor this performance measure. This measure has also been at the focus of some previous research studies. However, we claim that this measure is a biased reflection of the quality of service because the state of the stations should be weighted by their demand rates. For example, the adverse effect of an empty station that faces a low rate of demand for rentals is smaller compared with an empty station that faces a higher rate. Moreover, due to changes in demand rates along the day, it may even be desirable for certain stations to be full at certain times, and for other stations, to be empty. For example, during the mornings, many systems experience high demand for journeys from residential areas to business areas. Thus, in these hours it may be beneficial to fill up stations in residential areas and to empty stations in business areas. Other considerations that should be taken into account in a performance measure are the length of the trip, which may affect the inconvenience caused to a user due to an unfulfilled request, the time of the day, the proximity to other stations, available alternative modes of transportation, etc.

In contrast to the station-availability performance measure, it is common in studies of other public transportation systems to represent the service level by the total time users spend in the system. See, for example, road selection (Campbell 1992), buses and trams (Borndörfer et al. 2007), and passenger trains (Schöbel and Scholl 2006). A related measure to the total time is the total *excess time* users spend in the system. As presented in the Introduction, the excess time of a user is the difference between her actual and her ideal (shortest) travel time, see Ceder and Wilson (1986). In this study, we suggest using the total excess times of all users as the performance measure in a vehicle sharing system. The excess time is caused to a user due to an unfulfilled renting request or delays in returning the vehicle. We use this measure to evaluate and compare our suggested reservation policy to the current practice of one-way vehicle sharing systems, in which parking space reservation is not implemented. Note that focusing on the total excess time allows us to remove from the analysis a fixed constant that in any case cannot be reduced, that is, the total ideal time.

Currently, information systems of vehicle sharing systems log the actual journey times and itinerary, i.e., origins, destinations and journey durations. The true preferences of the users are unknown and there are also no records about abandonments. Based on this information, it is impossible to calculate the excess time of the users.

However, the information needed to calculate the excess time can be quite easily obtained. For example, by incorporating reservation policies, the true destinations of the users would be revealed. In addition, operators can encourage users to declare their requested origin and destination and to report if they decide to abandon, in order to improve their own (possibly future) service. Technically, this would be done using smartphone apps or station information kiosks.

We view the excess time as a better performance than the one currently used (station-availability), therefore in this paper we focus on this measure. However, to make the case for parking space reservation stronger, we compare (in Section 5) the CPR policy with the NR policy using both measures. Our simulation study demonstrates the advantage of the CPR with respect to both of these measures.

3. A vehicle sharing system model

In this section, the vehicle sharing system is modeled as a continuous time Markov chain, see for example Ross (2009). This is a simplistic model created in order to derive some general insights into the performance of the system under the NR (*No-Reservation*) and CPR (*Complete-Parking-Reservation*) policies. These insights are verified using actual data and a detailed simulation model in Section 5.

We use the following general notation to describe the vehicle sharing system configuration:

S The number of stations in the system

C_i The number of parking spaces in station i (station capacity)

V The number of vehicles dispersed in the system

$T_{ij}(t)$ The expected travel time between any two stations i, j of users who depart at time period t

$\lambda_{ij}(t)$ The arrival rate of users who wish to travel from station i to station j at time period t

The travel times reflect structural characteristics of the system, such as the distance between the stations, traffic density within the different segments, the topography of the city, and so on. Note that in general, $T_{ij}(t) \neq T_{ji}(t)$. For example, it is clear that riding a bicycle downhill is much faster than riding uphill. The expected travel times represent the duration of journeys that users actually make, which may not necessarily be the shortest possible. For example, $T_{ii}(t) \neq 0$, that is, roundtrip journeys are possible. As for the users, the demand for journeys between each pair of stations is given as a stochastic process. Furthermore, user behavioral rules are set in order to define how users react to shortages in either vehicles or parking spaces. Finally, this model is *service oriented*; the performance of the system is measured by the total excess time spent by users due to shortages of vehicles or parking spaces.

3.1. The NR policy

In this section, we model the system under the NR policy and Markovian assumptions. The arrival of renters to the system follows a Poisson process, and the travel time of journeys is exponentially distributed. For the sake of simplicity, the distributions are set to be homogenous in time. Therefore, we denote $\lambda_{ij}(t) = \lambda_{ij}$ and $T_{ij}(t) = T_{ij}$ for all t. Finally, to be able to assume a steady state, repositioning operations are not included in the model.

The following user behavior is assumed: (1) a user that faces a shortage of vehicles will abandon the system immediately. That is, she will choose an alternative mode of transportation and therefore will spend excess time on her journey. (2) A user that faces a shortage of vacant parking spaces will wait at the destination until a parking space becomes available, as a result of an arrival of a renter. Namely, the user will enter a waiting queue and will spend excess time waiting for her turn to return her vehicle. This assumption is made in order to simplify the Markovian model. In Section 5.1 the user behavior is extended so that roaming between stations is allowed.

The excess time due to abandoning is modeled as proportional to the travel time. Specifically, the excess time of an abandoning user who wishes to travel from station i to j is denoted by $\alpha_{ij}T_{ij}$ where α_{ij} is the penalty ratio.

Given the assumptions above, the state of the system is described by the following vector:

$$\boldsymbol{x} = (x_{01}, x_{11}, \dots, x_{S1}, \dots, x_{0S}, x_{1S}, \dots x_{SS})$$
$$x_{ij} \geq 0 \;\; \forall i = 0, \dots, S \;\; \forall j = 1, \dots, S$$

where x_{0j} denotes the number of vehicles in stations j and x_{ij} denotes the number of vehicles traveling from station i to station j. In particular, if $x_{0j} > C_j$ then all the parking spaces in station j are occupied and there are $x_{0j} - C_j$ users waiting in station j for parking spaces to become available. Furthermore, a possible state satisfies the following:

$$\sum_{j=1}^{S} \sum_{i=0}^{S} x_{ij} = V \tag{1}$$

Note that any given vehicle can be either at one of the S stations or traveling between one of the S^2 pairs of stations. In addition, there can be up to V vehicles in any given station or traveling between the same given pair of stations. We denote the set of all possible states by X. Recall that the number of parking spaces in a station is not binding because the formation of queues is possible. Therefore, the number of states is the number of ways to place V identical vehicles in $S(S+1)$ "bins". Hence, the number of possible states in the system is $\binom{S \cdot (S+1) + V - 1}{S \cdot (S+1) - 1}$ (Since the number of possible ways to distribute K identical items into N bins is $\binom{N + K - 1}{N - 1}$).

A transition between the system states occurs either when a vehicle is rented or when a user arrives at her destination. A *rent* transition between two possible states $\boldsymbol{x}, \boldsymbol{x}'$ is denoted by the following indicator function:

$$\gamma_{ij}(\boldsymbol{x}, \boldsymbol{x}') = \begin{cases} 1, & if\ \boldsymbol{x}' = (x_{01}, \dots, x_{S1}, \dots, x_{0i} - 1, \dots, x_{ij} + 1, \dots, x_{0S}, \dots, x_{SS}) \\ 0, & otherwise \end{cases}$$

where i is the origin station and j is the destination station. A *return* transition between two possible states $\boldsymbol{x}, \boldsymbol{x}'$ is denoted by the following indicator function:

$$\delta_{ij}(\boldsymbol{x}, \boldsymbol{x}') = \begin{cases} 1, & if\ \boldsymbol{x}' = (x_{01}, \dots, x_{S1}, \dots, x_{0j} + 1, \dots, x_{ij} - 1, \dots, x_{0S}, \dots, x_{SS}) \\ 0, & otherwise \end{cases}$$

where i is the station in which the vehicle was rented and j is the destination station.
The transition rates between any two possible states $\boldsymbol{x}, \boldsymbol{x}'$ are given by:

$$\sum_{i,j} \lambda_{ij}\gamma_{ij}(\boldsymbol{x}, \boldsymbol{x}') + \sum_{i,j} \frac{x_{ij}}{T_{ij}} \delta_{ij}(\boldsymbol{x}, \boldsymbol{x}') \tag{2}$$

Note that while the arrival rates of users do not depend on the number of vehicles in the station, the returning rates are linear functions of the number of traveling vehicles.

As discussed in Section 2, we measure the performance of the system by the total excess time users spend in the system. According to our model this is due to waiting at queues for a vacant parking space or due to abandoning the system and use of alternative mode of transportation. In the analysis of the Markovian models, we focus on the total excess time added per time period in steady state, referred to as the *expected excess time rate*.

We denote the total number of users waiting to return a vehicle in state $\boldsymbol{x}$ by $Q(\boldsymbol{x})$, where:

$$Q(\boldsymbol{x}) = \sum_{j:x_{0j}>C_j} (x_{0j} - C_j)$$

The expected number of users waiting to return a vehicle in steady state is given in Equation (3):

$$\sum_{\boldsymbol{x}\in X} Q(\boldsymbol{x})\pi(\boldsymbol{x}) \tag{3}$$

where $\pi(\boldsymbol{x})$ is the limiting probability of state $\boldsymbol{x}$. Note that the expected number of users waiting to return a vehicle is equivalent to the expected excess time rate due to waiting in queues. This is because for every time period a user waits in a queue, one period of excess time is accumulated. The steady state excess time rate due to abandonments in state $\boldsymbol{x}$ is given by:

$$\sum_{i:x_{0i}=0} \sum_{j=1}^{S} \lambda_{ij}\alpha_{ij}T_{ij} \tag{4}$$

In the internal sum of (4), the excess time rate due to abandonments is calculated for each empty station i, and the external sum goes over all empty stations. For the entire system, the expected excess time rate due to abandonments is obtained by summing (4) over all states, multiplied by their respective limiting probabilities, and is given by:

$$\sum_{\boldsymbol{x}\in X} \sum_{i:x_{0i}=0} \sum_{j=1}^{S} \pi(\boldsymbol{x}) \cdot \lambda_{ij}\alpha_{ij}T_{ij} \tag{5}$$

In conclusion, the expected excess time rate is obtained by summing (3) and (5).

It may be argued that when abandoning the system the quality of service is adversely affected not only by the travel time using an alternative mode of transportation but also by the mere fact that the user has to seek such an alternative service. For example, if an alternative mode of transportation is a bus, the user has to wait for it to arrive and to pay for a ticket. The waiting time and the cost of the ticket are typically unrelated to the travel time. Hence, the penalty incurred by the system for each passenger who uses a different mode of transportation may consist of a fixed component, in addition to the variable component which is proportional to the travel time. The fixed component should be expressed in units that are equivalent to travel time. The models presented in this section neglect this fixed component, but we revisit this issue and justify our simplifying assumption in Section 5.3, where we analyze a real world system using discrete event simulation.

3.2. The CPR policy

We now specify how the Markov chain for the NR policy is adapted to the CPR policy. Recall that a user will rent a vehicle only if there is one available in her origin station and there is an available parking space at the destination, namely, empty and non-reserved parking. Therefore, the total number of vehicles that travel to station i or are parked at station i cannot exceed the number of parking spaces in the station. In addition to (1), a state in a system managed by a CPR policy satisfies:

$$\sum_{i=0}^{S} x_{ij} \leq C_j \quad \forall j = 1 \dots S \tag{6}$$

We denote the set of all possible states in the CPR policy by $\tilde{X}$and note that $\tilde{X} \subset X$, that is, the set of possible states in the CPR policy is a subset of the NR set. In particular, $\tilde{X}$ only includes states that satisfy (6). Since the formation of queues is impossible, excess time under this policy is only due to abandoning. The expected excess time rate due to abandoning when in state $\boldsymbol{x}$ is given by:

$$\sum_{(i,j):x_{0i}=0 \ \vee \sum_{k=0}^{S} x_{kj}=C_j} \lambda_{ij}\alpha_{ij}T_{ij}$$

For the entire system, the expected excess time rate due to abandoning is given by:

$$\sum_{\boldsymbol{x}\in\tilde{X}} \sum_{(i,j):x_{0i}=0 \ \vee \sum_{k=0}^{S} x_{kj}=C_j} \tilde{\pi}(\boldsymbol{x})\lambda_{ij}\alpha_{ij}T_{ij} \tag{7}$$

where $\tilde{\pi}(\boldsymbol{x})$ is the limiting probability of state $\boldsymbol{x}$.

4. Comparison between the CPR and NR policies

In this section the vehicle sharing system under the Markovian assumptions and the user behavior assumptions described above is referred to as the M-VSS model (Markovian - Vehicle Sharing System). In Section 4.1 we define the notion of *offered load* in the context of the M-VSS model, and demonstrate that the expected excess time rate is a function of it. In Section 4.2, using the M-VSS model, we compare the CPR and NR policies under a range of offered loads. The comparison is illustrated through a small example in Section 4.3.

4.1. Offered load

The following definitions refer to the ideal situation in which all requested origin-destination journeys are satisfied. Denote T as the expected travel time per arriving user, specifically:

$$T = \frac{\sum_{i=1}^{S}\sum_{j=1}^{S}\lambda_{ij}T_{ij}}{\sum_{i=1}^{S}\sum_{j=1}^{S}\lambda_{ij}}$$

In the same manner, let λ denote the average arrival rate at a station, which can be written as:

$$\lambda = \frac{\sum_{i=1}^{S} \sum_{j=1}^{S} \lambda_{ij}}{S}$$

A user who rents a vehicle at station i and travels to station j will use the vehicle for an expected T_{ij} time units. That is, the expected work added to the system by this user is T_{ij} time units of vehicle usage. We define the *offered load* as the expected work to be added to an average station per time unit if all demands could be satisfied. The offered load is the product of T and λ:

$$\lambda T = \frac{\sum_{i=1}^{S} \sum_{j=1}^{S} \lambda_{ij} T_{ij}}{S}$$

Note that $S\lambda T$ represents the offered load in the entire system. That is, the total travel time added to the entire system per time unit. In other words, it represents the expected number of vehicles that would be in use at any given moment if the system could meet all the demand.

In the following section, we will measure the performance of the system under various loads by varying the values of T and λ. In order to facilitate a fair comparison, the relations between the stations are kept fixed. This is done by noting that all travel times T_{ij} can be written as the proportion τ_{ij} of the expected travel time T, that is $T_{ij} = \tau_{ij} T \ \forall i, j$ and similarly, all arrival rates λ_{ij} can be written as the proportion υ_{ij} of the expected arrival rate, that is $\lambda_{ij} = \upsilon_{ij} \lambda$. Then, when T is changed, all travel times are adjusted proportionally and similarly when λ is changed, the arrival rates at all stations are adjusted proportionally. In other words, while T and λ may be altered, the proportions τ_{ij} $\forall i, j$ and υ_{ij} $\forall i, j$ are kept fixed.

Lemma 1: For the M-VSS model, when υ_{ij} and τ_{ij} are fixed, the limiting probabilities are functions of the offered load λT.

Proof: the coefficients in the steady state equations are generally in the form prescribed in (2), that is, either proportional to λ or to $\frac{1}{T}$. By multiplying all equations by T an equivalent system of equations is received where all the coefficients are either proportional to λT or constant. This set of equations is given in (8).

$$\begin{aligned} &\lambda T \cdot \sum_{\boldsymbol{x}' \in X} \sum_{i,j} \upsilon_{ij} \pi(\boldsymbol{x}) \gamma_{ij}(\boldsymbol{x}, \boldsymbol{x}') + \sum_{\boldsymbol{x}' \in X} \sum_{i,j} \frac{x_{ij}}{\tau_{ij}} \pi(\boldsymbol{x}) \delta_{ij}(\boldsymbol{x}, \boldsymbol{x}') \\ &\qquad = \lambda T \cdot \sum_{\boldsymbol{x}' \in X} \sum_{i,j} \upsilon_{ij} \pi(\boldsymbol{x}') \gamma_{ij}(\boldsymbol{x}', \boldsymbol{x}) + \sum_{\boldsymbol{x}' \in X} \sum_{i,j} \frac{x'_{ij}}{\tau_{ij}} \pi(\boldsymbol{x}') \delta_{ij}(\boldsymbol{x}', \boldsymbol{x}) \qquad \forall \boldsymbol{x} \end{aligned} \tag{8}$$

In addition, the coefficients in (9) are all constants.

$$\sum_{\boldsymbol{x} \in X} \pi(\boldsymbol{x}) = 1 \tag{9}$$

The limiting probabilities are the solution of a system of equations where all coefficients are either proportional to λT or constant. Hence, the limiting probabilities are functions of the two parameters (T and λ) only through their product, regardless of their separate values. ■

Lemma 2: For the M-VSS model, when υ_{ij} and τ_{ij} are fixed, the expected excess time rate can be stated as a function of λT.

Proof: this is straightforward from (3), (5) and Lemma 1.

Consequently, the expected excess time rate is sensitive to changes in the offered load, rather than to changes in the expected travel time per user and the expected arrival rate separately. For instance, if we double the expected arrival rate, $\lambda^* = 2\lambda$, and divide the expected traveling time by two, $T^* = \frac{T}{2}$, the excess time rate will not change.

4.2. Conditions for dominance of the CPR policy

In this section we compare the performance of the M-VSS model under NR and CPR policies over a range of offered loads. In particular, our focus is on the range $\lambda T \in (0, V/S\,)$. The upper bound is received through the following observation: there can be at most V vehicles in use at any given moment. The expected offered load in the entire system is given by $S\lambda T$. Hence $S\lambda T = V$ represents the 100% utilization of the system. This bound is equivalent to the $\rho < 1$ bound in classic queueing systems. In the long run, it is safe to assume that the system will adapt itself to this range. In the case of higher loads, either the system will need to be further developed or the rate of user arrivals will drop due to constant lack of service. We note that in an unlikely case where $V \le \min_i\{C_i\}$, users will always be able to return their vehicles in any station, that is, the system will perform the same under both policies. Since in practice the number of vehicles is much larger than the capacity of any station, we assume that $V > \min_i\{C_i\}$.

We begin with the extreme case of $\lambda T = 0$, namely the case when the expected travel time is negligible.

Lemma 3: For the M-VSS model, when $\lambda T = 0$, the expected excess time rate due to abandoning is zero under both policies.

Proof: To see this we rewrite Equations (5) and (7) as (10) and (11), respectively:

$$\lambda T \sum_{\boldsymbol{x}\in X} \pi(\boldsymbol{x}) \sum_{i:x_{0i}=0}^{S} \sum_{j=1}^{S} \upsilon_{ij}\alpha_{ij}\tau_{ij} \tag{10}$$

$$\lambda T \sum_{\boldsymbol{x}\in \tilde{X}} \tilde{\pi}(\boldsymbol{x}) \sum_{(i,j):x_{0i}=0\ \vee \sum_{k=0}^{S} x_{kj}=C_j} \upsilon_{ij}\alpha_{ij}\tau_{ij} \tag{11}$$

Since the expressions in the internal sums are fixed and the sum of limiting probabilities is bound from above by 1, both sums equal zero when the offered load is zero.

It remains to examine the excess time rate due to queueing in the NR policy.

Lemma 4: For the M-VSS model, in the NR policy, when $\lambda T = 0$, the expected excess time rate due to queueing is a positive constant, independent of λ and T.

Proof: See Appendix A.

In conclusion, when λT=0, the excess time in the CPR policy tends to zero while in the NR policy it tends to a positive value. Therefore, when λT is zero, the CPR policy is preferable. It turns out that there is a discontinuity point at $\lambda = 0$ since clearly there is no excess time when no customers arrive at the system. We are now ready to state an important result.

Theorem 1: For the M-VSS model, there exists some $\beta > 0$ for which the CPR policy delivers smaller excess time (better service) compared to the NR policy under any workload that satisfies $\lambda T < \beta$.

Proof: Both excess time rate expressions are continuous functions of λT (both are sums of continuous functions). By Lemma 3 and Lemma 4, when $\lambda T = 0$ the CPR policy is preferable. If the functions intersect, the first intersection point β defines a range $(0, \beta]$ in which the CPR policy is preferable. Otherwise, the CPR policy is preferable for any λT. ■

4.3. An illustrative example

Theorem 1 presented above is general in the sense that no limitations were imposed on the number of stations in the system or their capacities. To demonstrate this result, we use a small system, configured as follows: the system is comprised of two stations ($S = 2$). The capacity of both stations is $C_1 = C_2 = 2$, the number of vehicles in the system is V=3. The expected travel time between the two stations is identical in both directions, $T_{12} = T_{21} = T$. The arrival rates of renters at the two stations are denoted by λ_{12} and λ_{21}. For the sake of simplicity there is no demand for round-trip journeys, that is $\lambda_{11} = \lambda_{22} = 0$. This allows us to reduce the state representation to $\boldsymbol{x} = (x_{01}\, x_{21}, x_{02}\, x_{12})$. Recall that x_{0i} represents the number of vehicles in station i and x_{ji} represents the number of vehicles traveling from station j to station i. For example, the state $(20{,}01)$ indicates that two vehicles are parked in station 1, no bicycle is travelling to station 1, there are no vehicles parked in station 2 and one vehicle is travelling to station 2. Finally, the excess time due to abandonment is αT ($\alpha_{12} = \alpha_{21} = \alpha$). In Figure 1 we depict the resulting Markov chains for the NR and CPR policies. The entire figure represents the NR policy chain and the sub-graph which consists of solid arcs and nodes represents the CPR policy chain. Notice that even for a small and simple system, the resulting Markov chains are quite intricate.

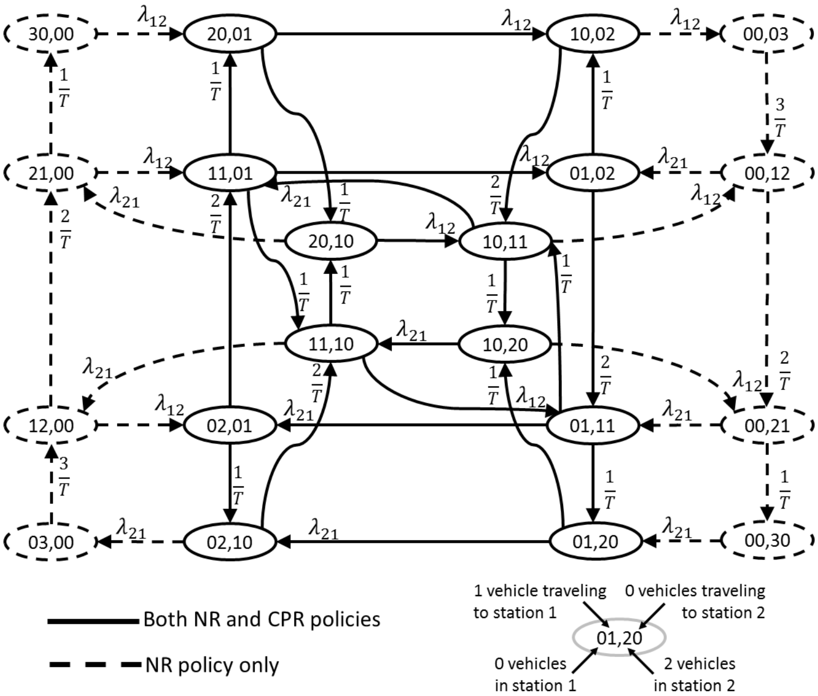


Figure 1: Graph representation of the Markov chains for the NR and CPR polices of the illustrative example

The limiting probabilities of the two Markov chains were calculated by solving the system of equations given by (8) and (9). The expected excess time rate for each policy was calculated according to (3), (5) and (7). We compare the two policies by subtracting the excess time rate of the CPR policy from the excess time rate of the NR policy. The function that describes this difference, as a function of the offered load λT, is termed the *difference function*. A positive difference means that the CPR policy performs better and a negative value means that the NR policy performs better for the corresponding offered load. From Theorem 1 we know that there exists $\beta > 0$ such that for all $\lambda T \in (0, \beta]$ the difference function is positive.

In Figure 2 we present six difference function graphs for various settings. In the top graphs we set demand rates in both stations to be the same, that is $\lambda_{12} = \lambda_{21} = \lambda$. In the bottom graphs the total demand rate is kept constant but demands are unequal with $\lambda_{12} = 0.5\lambda$, $\lambda_{21} = 1.5\lambda$. Namely, the demand rate from station 2 to station 1 is three times the demand rate in the opposite direction. Recall that α is the penalty ratio in the case of abandonment. We set the penalty ratio to three different levels, $\alpha = 0.5, 1, 10$, as presented in the left, middle and right graphs, respectively. To illustrate the range of α values that are likely to be used, we note, for example, that in a bike sharing system, an abandoning user will probably walk to her destination. Assuming that the walking time is twice the riding time, the excess time due to abandoning is equal to the riding time, i.e. $\alpha = 1$. Notice that only in the upper right graph, Figure 2(e), the curve intersects the axis, that is, there exists a range of offered loads in which the NR policy is preferable. This graph describes a case where the demands are balanced and the penalty for abandonment is extremely high.

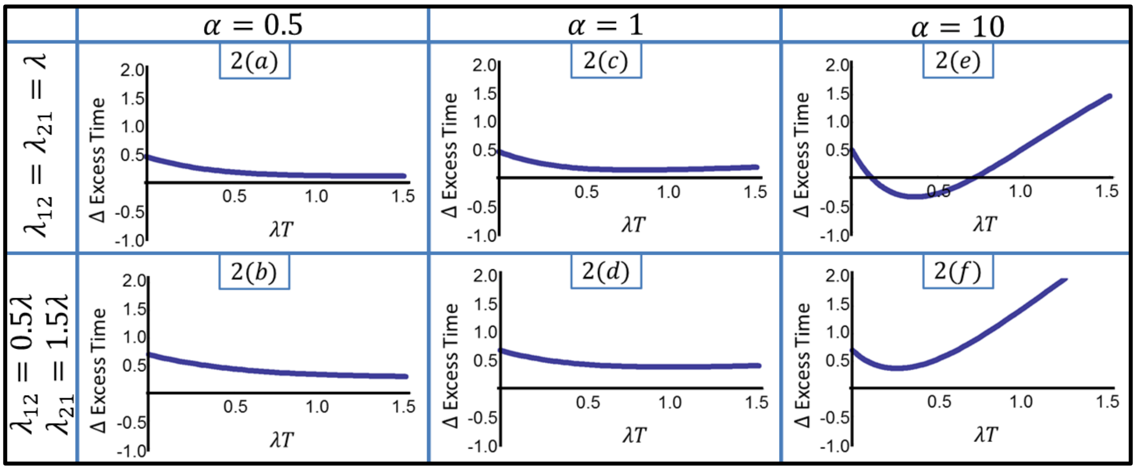


Figure 2: Difference function for symmetric and asymmetric demands and for various abandonment penalty ratios

In order to examine the effect of imbalanced demand on the difference function, we analyze the above system by assigning $\lambda_{12} = \lambda\theta, \lambda_{21} = \lambda(2-\theta)$, $\theta \in [0,2]$, and study the resulting difference function over θ. Indeed, the results show that for any λ, there exists a unique minimum point of the difference function at $\theta = 1$, which corresponds to the case $\lambda_{12} = \lambda_{21}$. Therefore, in this small example, the worst performance of the CPR policy compared to the NR policy is when the system is completely symmetric in terms of travel times and arrival rates. Note that in this case, rebalancing is least needed.

Recall that the bound for the system's offered loads is $S\lambda T \leq V$. In this example, $S = 2,\ V = 3$ and so $\lambda T = 1.5$ represents 100% utilization. To further study the range in which the CPR policy is preferable over NR, we present in (12) and (13) the resulting excess time rate functions when $\lambda_{12} = \lambda_{21} = \lambda$, for the NR and CPR policies, respectively. The resulting expressions are relatively compact compared to the general expressions. Here, it is easy to see that both are functions of the offered load λT. The difference function is given in (14).

NR Policy:
$$\frac{3 + 3\alpha\lambda T + 6\alpha(\lambda T)^2 + 6\alpha(\lambda T)^3 + 4\alpha(\lambda T)^4}{6 + 9\lambda T + 6(\lambda T)^2 + 2(\lambda T)^3} \tag{12}$$

CPR Policy:
$$\frac{2\alpha\lambda T(6 + 10\lambda T + 7(\lambda T)^2 + 9(\lambda T)^3 + 3(\lambda T)^4)}{12 + 32\lambda T + 28(\lambda T)^2 + 14(\lambda T)^3 + 3(\lambda T)^4} \tag{13}$$

$$\frac{36 + 12(8-3\alpha)\lambda T + 12(7-5\alpha)(\lambda T)^2 + 6(7+2\alpha)(\lambda T)^3 + 9(1+8\alpha)(\lambda T)^4 + 67\alpha(\lambda T)^5 + 24\alpha(\lambda T)^6 + 2\alpha(\lambda T)^7}{72 + 300\lambda T + 528(\lambda T)^2 + 552(\lambda T)^3 + 376(\lambda T)^4 + 167(\lambda T)^5 + 46(\lambda T)^6 + 6(\lambda T)^7} \tag{14}$$

One can see that when the penalty parameter is in the range $0 \leq \alpha \leq 1.4$, (14) is positive for any λT, that is, the CPR policy is always preferable. Recall that for bike-sharing systems a reasonable value of α is 1. Also, for $\lambda T \geq 0.82$, (14) is positive for any α, that is, the CPR policy performs better than NR, no matter how high the penalty for abandoning is. For large α values, the excess time due to waiting in a queue is likely to be smaller than the excess time when abandoning due to the inability to make a reservation. Therefore as α increases the range in which the CPR policy is preferable, narrows. Note that when the offered load is extremely high, the difference function converges to a constant, which means that the difference between the policies in terms of excess time per user is negligible. Indeed, if the offered load is extremely high, all vehicles will be constantly traveling, that is, in both policies most users will abandon due to shortage of vehicles, and therefore the system will perform the same under both policies.

Recall that the excess time rate function of the NR policy is composed of two components, the excess time rate due to abandoning and the excess time rate due to waiting. Next, we examine the tradeoff between the two components and compare them to the excess time rate function of the CPR policy, which is due to abandoning only. In Figure 3 we further examine the case where in the illustrative example, the demand rates in both directions are identical ($\lambda_{12} = \lambda_{21} = \lambda$) and the penalty ratio is set to $\alpha = 1$. We present the functions of the two components of the NR policy and their sum in black dotted, dashed and solid lines, respectively, and the excess time rate function of the CPR policy in a solid gray line. As the offered load increases, the excess time due to waiting decreases while the excess time due to abandoning increases. For this instance, the CPR policy is superior (lower excess time) over the entire examined range of offered loads. While for low loads the dominance is mainly due to the waiting times, for higher loads ($\lambda T > 0.8$), the excess time due to abandoning in the NR policy is higher by itself than in the CPR policy. This means that for relatively high loads, fewer users will abandon under the CPR policy.

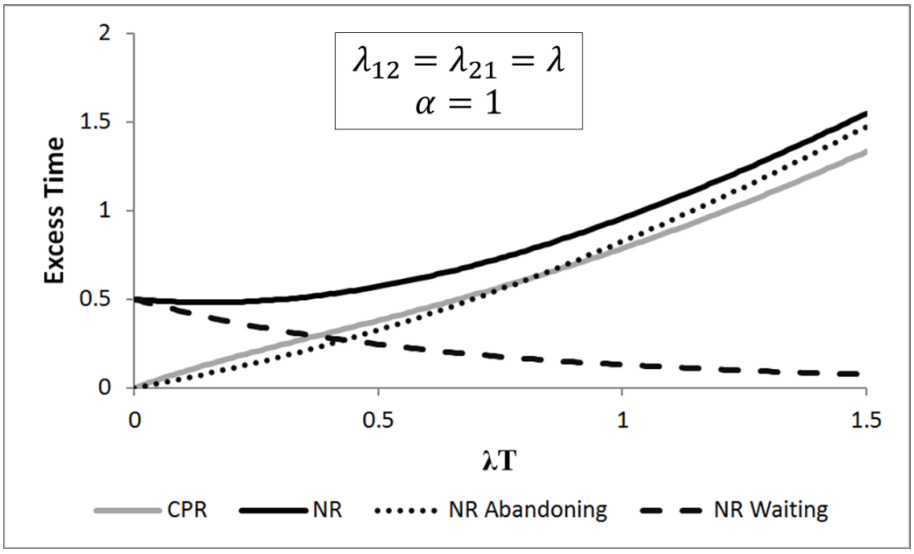


Figure 3: Excess time rate functions: NR policy (abandoning, waiting and total) and CPR policy

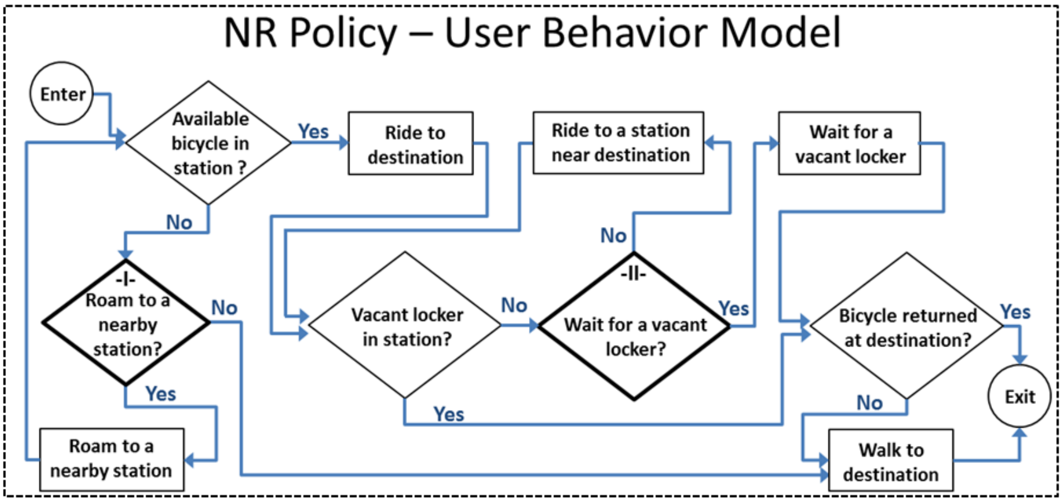
5. Simulation model

To further examine the effectiveness of the CPR policy, real world systems should be examined. Unfortunately, it is intractable to evaluate the excess time rate function for large size vehicle sharing systems. Moreover, in the Markovian models, a simple user behavior was assumed and the demand process was assumed to be homogenous throughout the day. Both assumptions do not fit real systems. In order to better model the complexities of real vehicle sharing systems we use discrete event simulation in which we relax the homogenousity assumption and present extensions made to the user behavior model. In Section 5.2 we describe a real-world bike sharing system that was used as a case study. For a clearer presentation we use throughout this section bike-sharing terminology. Specifically, parking spaces are referred to as lockers and the vehicles are simply bicycles. The alternative mode of transportation selected by the users is assumed to be walking.

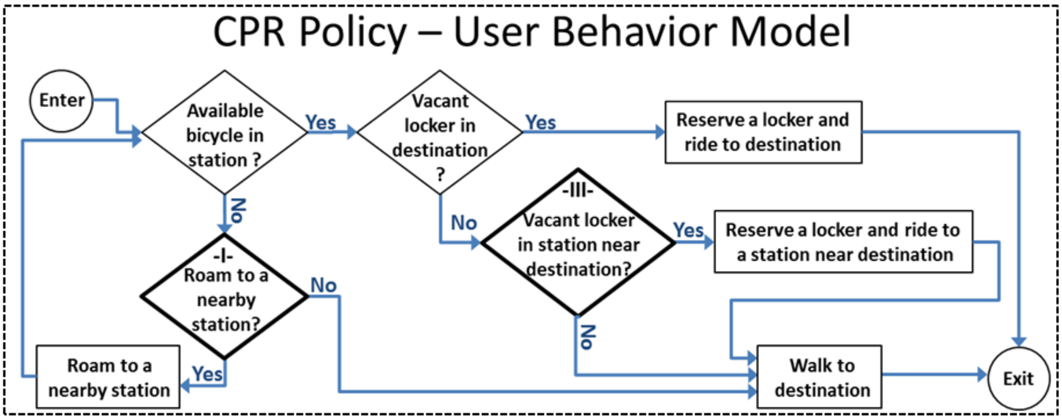
5.1. Enhancing the user behavior model

The movement of users within the bike-sharing system is a complicated process. In particular, users may react in different ways to shortages of bicycles or lockers. Each reaction may project on a different group of users in the system. We assume that the users are strategic and that they make use of information available in the stations' kiosks or accessed by their smartphones through the internet. Specifically, users have full knowledge of the travel times between each pair of stations, arrival rates of renters and real-time inventory levels. Each user is interested in reaching her destination in the shortest time. At decision points, due to an unfulfilled demand, we assume that the users choose the alternative that minimizes the expected remaining time in the system. The users are myopic, in the sense that they do not take into account the implications of possible changes in the inventory levels of intermediate stations during their journey.

The behavior models under the NR and the CPR policies are described in Figure 4(a) and 4(b), respectively. In Table 2 we further elaborate on the user decision processes, presenting the exact calculation made by a user at a decision point. Three types of decision points exist, denoted in Figure 4 and in Table 2 by I, II and III.



4(a)



4(b)

Figure 4: User behavior models in the NR and CPR policies

Under the NR policy, a user that does not find an available bicycle may choose to roam to a nearby station or walk directly to her destination (I). When a bicycle is rented the user rides to her destination. If upon arrival at the destination she finds an available locker, she returns the bicycle there and leaves the system. Otherwise, the user may either enter a waiting queue in that station or ride to a nearby station (II). If the bicycle is not returned at the destination, the user walks from the returning station to the destination.

Under the CPR policy, a user that does not find an available bicycle may choose to roam to a nearby station or walk directly to her destination (I). When a bicycle is available, if the user is unable to reserve a locker at her desired destination, she may choose either to reserve a locker in a station near her destination or to waive the service of the system altogether and walk directly to her destination (III). Again, if the bicycle is not returned at the originally desired destination, the user walks from the actual returning station to the desired destination. The main difference from the NR policy is that once a bicycle is rented, the returning is ensured.

We use the following notation to describe the user decision processes in Table 2:

T_{ij} Expected riding time from station i to station j (as defined in Section 3)

$\lambda_i(t)$ Arrival rate of renters to station i at decision time t $\left(\lambda_i(t) = \sum_{j=1}^{S} \lambda_{ij}(t)\right)$

W_{ij} Expected walking time from station i to station j $\left(W_{ij} = T_{ij} + \alpha T_{ij}\ \forall i \neq j, W_{ii} = 0\ \forall i\right)$

Q_i Queue length in station i at the decision time

B_i Number of bicycles in station i at the decision time

L_i Number of available lockers (not occupied and not reserved) in station i at the decision time

We assume that the riding times ($\boldsymbol{T_{ij}}$) and walking times $(\boldsymbol{W_{ij}})$ do not change along the day. For simplicity, we omit the index $\boldsymbol{t}$ from the state variables $\boldsymbol{Q_i}, \boldsymbol{B_i}, \boldsymbol{L_i}$.

Table 1: Decision processes of users who face shortages of bicycles or lockers

Situation	Question	Decision Process
User with destination j arrives at station i in which there are no available bicycles (NR, CPR)	Roam to a nearby station? (I)	Set $k^* = \arg\min_{k:B_k>0}\left(W_{i,k} + T_{k,j}\right)$ If $W_{i,k^*} + T_{k^*,j} < W_{i,j}$ roam to station k^* (Yes) Else, walk to j (No)
User with destination j arrives with a bicycle to station i, in which there are no vacant lockers (possibly $i = j$) (NR)	Wait for a vacant locker? (II)	Set $k^* = \arg\min_{k:L_k>0}\left(T_{i,k} + W_{k,j}\right)$ If $T_{i,k^*} + W_{k^*,j} \geq (Q_i + 1)/\lambda_i(t) + W_{i,j}$ wait in station i (Yes) Else, ride to station k^* (No)
User finds an available bicycle at station i but cannot make a reservation at her destination j (CPR)	Vacant locker in station near destination? (III)	Set $k^* = \arg\min_{k:L_k>0}\left(T_{i,k} + W_{k,j}\right)$ If $T_{i,k^*} + W_{k^*,j} < W_{i,j}$ reserve a locker and ride to station k^*, from there walk to j (Yes) Else, walk to j (No)

The total time a user spends in the system is taken as the difference between the time the user reaches her destination and her arrival time to the system. The excess time is obtained by subtracting the *ideal time*, the net riding time from the desired origin to the desired destination. We refer to the net riding time as the ideal time because this is the time the user spends in the system if she does not experience shortage of bicycles or lockers, the ideal situation.

5.2. A real world system

As a case study we take the bike sharing system in Tel-Aviv, Tel-O-Fun, with 130 stations scattered in an area of about 50 square kilometers. A total of 2,500 lockers and 900 bicycles are dispersed in the system. Weekday rent transactions were collected over two months. Daily demand patterns did not change significantly throughout this period. The average number of daily trips was about 4,200.

By aggregating the transactions, we estimated the arrival rate of renters during 30 minute periods throughout the day. As may be expected, in most stations the demand process was not homogenous over time. For example, the demand for bicycles in stations located near working areas was low at the beginning of the day and increased significantly towards the end of the working day.

Riding times were estimated using the Google Maps API. For regular trips, it is safe to assume that most users ride directly from the origin to the destination. Indeed, the average time of all transactions was about 12 minutes. This is not the case for round-trips, where the average duration was approximately 30 minutes. Such trips were about 8% of all rent transactions. The penalty for an abandoning user was set to $a = 1$, assuming that walking time is twice the riding time.

We note that in its current state, the information system of Tel-O-Fun cannot document information regarding abandonments. This is mainly due to the fact that when a user arrives at an empty station, she does not attempt to rent a bicycle and therefore is not identified by the system. Moreover, the system cannot tell apart users who returned bicycles at their desired destination and those who had to roam to a nearby station. To deal with this issue, we estimated the proportion of time a station was empty or full and inflated the demand rates accordingly.

5.3. Numerical results

The discrete event simulation, together with the user behavior model logic was coded in MathWorks Matlab™. The average daily bicycle usage (the total rent durations), was about 920 hours. In a system with 900 bicycles there are 21,600 available bicycle hours a day. That is, the average daily utilization was 4.2%. In addition, the average utilization at peak hours was 8.3%.

In order to test the capabilities of the system under various offered loads, we multiplied the measured offered load by the following factors: 0.5,1,2,4,8. This was done by multiplying uniformly all arrival rates by these factors (*load multipliers*). In addition, we used two starting points for the initial inventory level of bicycles at the beginning of the working day: (1) the actual initial station inventories on a randomly chosen day, after the operators executed repositioning activities. (2) the inventory levels prescribed by the method of Raviv and Kolka (2013).

For each load factor, we randomly generated 50 daily demand realizations, including renters' arrival times to each station and their destinations. In order to reduce variation, we used the same realizations for each combination of policy and initial inventory (Common random numbers).

The simulation results are presented in Table 2 and Table 3 for the two starting inventory levels. In the first column we present the load multiplier and in the second we present the policy that was used. In the third and fifth columns we present the average total time spent by users in the system, over 50 replications, and the total ideal time, respectively. The absolute and relative excess times in each configuration are reported in the sixth and seventh columns, respectively. In the fourth and eighth columns we present, respectively, the relative reduction of the total time and of the excess time in the CPR policy as compared to the NR policy. In the ninth and tenth columns we present the percentage of users who received an ideal ride and the percentage of users who did not rent a bicycle at all, respectively.

Table 2: Simulation results for various system loads with initial inventory taken from actual random day

Initial Inventory – Actual day										
Load Multiplier	Policy	Total Time In System (hr./day)	% Total Time In System Reduced	Total Ideal Time (hr./day)	Absolute Excess Time (hr./day)	%Excess	% Excess Reduced	Ideal Rides	Unserved	Fixed Penalty (min.)
0.5	NR	492.40	0.93%	459.29	33.11	7.21%	13.89%	86.23%	2.72%	45.98
	CPR	487.80			28.51	6.21%		86.43%	3.00%	
1	NR	1,009.76	1.33%	919.87	89.90	9.77%	14.99%	82.36%	3.84%	52.45
	CPR	996.28			76.42	8.31%		82.78%	4.21%	
2	NR	2,108.86	2.58%	1,831.42	277.44	15.15%	19.62%	75.45%	6.28%	477.53
	CPR	2,054.43			223.01	12.18%		77.18%	6.36%	
4	NR	4,554.72	3.91%	3,675.82	878.90	23.91%	20.27%	65.11%	11.68%	∞
	CPR	4,376.54			700.71	19.06%		68.83%	10.58%	
8	NR	10,172.35	4.39%	7,335.99	2,836.37	38.66%	15.74%	51.83%	22.72%	∞
	CPR	9,725.77			2,389.79	32.58%		55.93%	19.82%	

Table 3: Simulation results for various system loads with initial inventory calculated according to the method of Raviv and Kolka (2013)

Initial Inventory – Raviv and Kolka (2013)										
Load Multiplier	Policy	Total Time In System (hr./day)	% Total Time In System Reduced	Total Ideal Time (hr./day)	Absolute Excess Time (hr./day)	%Excess	% Excess Reduced	Ideal Rides	Unserved	Fixed Penalty (min.)
0.5	NR	477.20	1.26%	459.29	17.91	3.90%	33.56%	93.00%	0.49%	28.22
	CPR	471.19			11.90	2.59%		92.79%	1.11%	
1	NR	979.91	1.93%	919.87	60.05	6.53%	31.46%	88.69%	1.39%	36.58
	CPR	961.02			41.16	4.47%		88.92%	2.13%	
2	NR	2055.57	3.73%	1,831.42	224.16	12.24%	34.18%	80.51%	3.31%	78.04
	CPR	1978.96			147.54	8.06%		82.42%	4.02%	
4	NR	4383.17	5.05%	3,675.82	707.35	19.24%	31.30%	71.19%	6.98%	559.14
	CPR	4161.77			485.95	13.22%		75.00%	6.92%	
8	NR	9586.93	7.26%	7,335.99	2250.94	30.68%	30.93%	59.03%	15.09%	∞
	CPR	8890.63			1554.65	21.19%		65.89%	11.75%	

It can be observed in both tables that for all the tested values of offered loads, the CPR policy outperformed the NR policy in terms of the total excess time (hence also in terms of total time in the system) and in terms of the percentage of ideal rides. This is accompanied by a slight decrease in the percentage of served users in low load configurations (0.5,1) and a significant increase in the percentage of served users in high load configurations (8). That is, the operator can achieve a significant improvement in the quality of service at a small loss of rent revenue when the offered load is relatively low, and even increase revenue when the offered load is high. Note that unserved users that are not able to make a reservation under the CPR policy, are likely to spend significant excess time looking for a vacant locker had the system allowed them to rent a bicycle. In the long run, better service is essential to retaining users and attracting new ones and hence both the users and the operator are likely to benefit from the CPR policy.

Next, we evaluate how the addition of a fixed penalty component per abandoning user (see discussion at the end of Section 3.1) affects the superiority of the CPR policy. To this end we calculate, for every configuration, the value of the fixed penalty for which the two policies breakeven. These values are presented in the last column of Table 2 and Table 3. The lowest calculated breakeven value is about 40 minutes, and in some configurations the value is infinity, i.e., for any fixed penalty the CPR policy outperforms the NR policy. This happens when the percentage of unserved users is smaller under the CPR policy. Given that an average duration of a rent is about 12 minutes, the superiority of the CPR policy continues to hold for any reasonable fixed penalty.

Another noticeable result is that excess times in Table 3 are lower per offered load and policy as compared to the figures reported in Table 2. That is, better results are obtained by setting the initial inventories according to the method of Raviv and Kolka (2013). Though this result is not a part of the contribution of this study, it addresses the issue of interaction between repositioning and reservation policies. As can be seen in Table 3, even if system operators carry out effective static repositioning, implementation of the CPR policy can further improve performance.

We note that when using the method of Raviv and Kolka (2013), the total initial inventory of bicycles was 1,268 bicycles, about half the total number of lockers in the system. Clearly, during the reviewed period the operators had not dispersed a sufficient number of bicycles in the system. To bridge this gap, we leveled the total inventory by uniformly adding bicycles to the actual initial stations' inventories on the randomly chosen day. We ran the simulation with these initial inventories as well. The resulting system performance measures were similar to those of Table 2. Hence, the difference from Table 3 can be attributed to proper distribution of the bicycles at the beginning of the day.

Next we compare the two policies in terms of the station-availability performance measure. Recall that while under the NR policy a user cannot return a bicycle if the station is full, under the CPR policy a user cannot make a reservation if all the lockers are either occupied or reserved, i.e., the station is blocked. In Table 4 we present the average percentage of time a station is empty or full\blocked. In addition, we present the percentage of users who could not rent a bicycle upon their first attempt and the percentage of users who could not return a bicycle or make a reservation upon their first attempt. In the third to sixth columns we present these figures for the case where the initial inventories were taken from a random day. In the last four columns we give the figures for the case where the initial inventories were set according to Raviv and Kolka (2013).

Table 4: percentages of unfulfilled first attempts to rent or return a bicycle

		Initial Inventory							
		Actual day				Raviv and Kolka (2013)			
Load Multiplier	Policy	Empty	Full\ Blocked	Unfulfilled in rent	Unfulfilled in return\ reserve	Empty	Full\ Blocked	Unfulfilled in rent	Unfulfilled in return \reserve
0.5	NR	6.29%	1.47%	12.17%	2.70%	1.38%	2.16%	2.88%	4.48%
	CPR	6.19%	1.39%	11.80%	2.75%	1.29%	2.08%	2.71%	4.73%
1	NR	9.06%	2.25%	16.59%	3.46%	3.07%	3.32%	6.68%	5.96%
	CPR	8.85%	2.06%	15.87%	3.50%	2.86%	3.04%	5.99%	5.92%
2	NR	12.92%	3.76%	25.57%	5.09%	5.59%	5.92%	14.42%	9.13%
	CPR	12.28%	3.03%	23.55%	4.25%	4.96%	4.86%	12.02%	8.13%
4	NR	19.40%	5.20%	44.55%	5.10%	10.50%	8.77%	28.56%	9.82%
	CPR	17.98%	4.10%	38.76%	3.93%	8.92%	7.23%	22.67%	8.65%
8	NR	27.63%	6.14%	78.05%	3.32%	18.16%	11.93%	55.94%	7.72%
	CPR	25.70%	5.00%	68.59%	2.83%	14.96%	9.76%	40.98%	7.42%

For all simulated configurations the percentage of empty and full\blocked stations are lower under the CPR policy, that is, the CPR policy is preferable also under the station-availability performance measure. As discussed in Section 2, this performance measure does not truly reflect the quality of service given to the users because, among other flaws, it does not take into account the number of arriving users while the stations are empty or full. This is easily noticed in Table 4 when comparing the percentages that appear in the "Empty" columns with the "Unfulfilled in rent" columns as well as when comparing the percentages that appear in the "Full\Blocked" columns with the "Unfulfilled in return\reserve" columns. Moreover, the differences can be quite significant, to either direction.

Focusing now on unfulfilled requests, the percentage of users who cannot receive service due to lack of bicycles (Unfulfilled in rent) is lower under the CPR policy for all configurations. As may be expected, in some configurations the percentage of users who failed to make a reservation was higher than the percentage of users who failed to return a bicycle (unfulfilled in return\reserve). If the two columns are summed, we see that the percentage of unfulfilled requests is higher in the NR policy (except for when the load multiplier is 0.5 and the initial inventory is set according to the method of Raviv and Kolka (2013)). The situations in which users cannot return a bicycle are typically perceived as ones that cause more frustration to the users since they are "trapped" in the system. If indeed a higher priority is assigned to securing the ability of users to return the bicycles, the advantage of the CPR policy increases.

Lastly, we demonstrate that the above results are not sensitive to the penalty ratio. The above experiment was repeated using higher penalty ratios, namely $\alpha = 2,4,8$. Under these settings, the traveling times using alternative modes of transportation (e.g., walking) are three, five and nine times the riding times, respectively. The results of this experiment are reported in Table 5. In the third to sixth columns we present the percentage of excess time spent by users in the system for the case where the initial inventories were taken from a random day. In the last four columns we present these values for the case where the initial inventories were set according to Raviv and Kolka (2013). Note that the total ideal times do not depend on the penalty ratio, and therefore they are identical, for each load multiplier, to those reported in Table 3. Observe that as the penalty ratio grows, the absolute and relative difference between the NR and CPR policy grows, that is, the superiority of the CPR policy increases. As the penalty ratio grows, more users will enter the system, that is, the effective load on the system will grow. This aligns with the results obtained by increasing the load multipliers.

Table 5: percentage of excess time spent in the system for various penalty ratios

		Initial Inventory							
		Actual day				**Raviv and Kolka (2013)**			
Load Multiplier	Policy	$\alpha = 1$	$\alpha = 2$	$\alpha = 4$	$\alpha = 8$	$\alpha = 1$	$\alpha = 2$	$\alpha = 4$	$\alpha = 8$
0.5	NR	7.21%	13.04%	25.06%	49.73%	3.90%	5.99%	10.39%	19.90%
	CPR	6.21%	12.10%	23.90%	47.33%	2.59%	4.86%	9.04%	17.27%
1	NR	9.77%	18.06%	35.42%	71.07%	6.53%	10.81%	20.03%	38.62%
	CPR	8.31%	16.49%	33.23%	66.90%	4.47%	8.74%	17.13%	33.53%
2	NR	15.15%	28.20%	53.61%	105.99%	12.24%	20.98%	38.68%	74.34%
	CPR	12.18%	24.50%	48.19%	96.95%	8.06%	16.14%	32.03%	62.98%
4	NR	23.91%	46.42%	85.40%	166.95%	19.24%	34.98%	64.51%	124.19%
	CPR	19.06%	38.31%	73.44%	146.03%	13.22%	26.70%	52.67%	104.46%
8	NR	38.66%	79.40%	154.35%	291.59%	30.68%	60.49%	112.69%	212.53%
	CPR	32.58%	67.59%	133.31%	252.19%	21.19%	43.25%	84.51%	165.97%

6. Concluding remarks and future research

This is the first study on parking space reservation policies in vehicle sharing systems. We view such policies as a tool to passively redirect demands and balance inventory levels. Specifically, we show that the excess time spent by users in a system managed under the CPR policy is lower compared to the excess time in the base policy, for a large range of offered loads. This is demonstrated by both an analytical analysis of Markovian models and by a simulation of a real world system.

Reservation policies also reduce the uncertainty related to the usage of vehicle sharing systems. The guarantee that a parking space will be available upon the return of the vehicle at the destination can save the time and anxiety associated with the possibility of having to search for a parking space.

Throughout this study, we compared CPR and NR policies. It may be worth examining “smarter” policies. For example: time limited reservation policies, station specific reservation policies or even user- based policies that involve reservations. In the CPR policy, a reserved parking space is unavailable until the vehicle is returned. It may be better not to allow the seizing of resources for too long. In addition, some users may prefer not to declare their destination in advance. If only some of the users reserve in advance a parking space, “revenue management” questions rise. For example, how many parking spaces should be kept for users who happen to arrive at the station without pre-reservation? What kind of incentives should be awarded to users who reserved?

In addition to the advantages discussed in this paper, we note that by placing or trying to place reservations, the users reveal information that is currently not available to the operators of vehicle sharing systems. Such information may be useful both for operational and strategic decisions. For example, information received via reservations can assist the operators in predicting the near future state of the system. This may allow better short term planning of repositioning activities. In addition, information on journeys that cannot be realized is collected. Such information is vital when planning capacity expansion of stations in the system.

In this study we have assumed that users act according to the dictated policy regulations. However, a user may decide to return the vehicle at a station different from the one in which the parking space was reserved. Furthermore, strategic users may declare a different destination than their true desired destination, simply to be able to rent a vehicle. It is interesting to examine the effects of such behavior on the performance of the system and to determine whether the system should allow such vehicle returns. These questions require further research.

To conclude, the findings of this paper suggest that incorporation of parking reservation policies in vehicle sharing systems will improve the quality of service given to the users. Technologically, incorporation of parking reservation policies merely requires minor software and, in some systems, hardware updates.

Acknowledgment

This research was supported by THE ISRAEL SCIENCE FOUNDATION (ISF) grant no 1109/11.

References

Angeloudis, P., Hu, J., Bell, M-G-H, 2012. A strategic repositioning algorithm for bicycle-sharing schemes. In: *Transportation Research Board 91st Annual Meeting Compendium of Papers*.

Benchimol, M., Benchimol, P., Chappert, B., Taille, A-D-L., Laroche, F., Meunier, F., Robinet, L,. 2011. Balancing the stations of a self-service "Bike Hire" system. *RAIRO Operations Research*, 45, 37–61.

Borndörfer, R., Grötschel, M., Pfetsch, ME., 2007. A column generation approach to line planning in public transport. *Transportation Scienece,* 41(1), 123-132.

Campbell, J-F., 1992. Selecting routes to minimize urban travel time. *Transportation Research Part B*, 26(4), 261-274.

Ceder, A., Wilson, N.H.M., 1986. Bust network design. *Transportation Research Part B*, 20(4), 331-344.

Chemla, D., Meunier, F., Wolfer-Calvo, R., 2012. Bike sharing systems: solving the static rebalancing problem. *Discrete Optimization*,10(2), 120-146.

Contardo, C., Morency, C., Rousseau, L-M., 2012. Balancing a dynamic public bike sharing system. https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2012-09.pdf

Correia, G., Antunes, A., 2012. Optimization Approach to Depot Location and Trip Selection in One-Way Carshring Systems. *Transportation Research Part E*, 48, 233–247.

DeMaio, P., 2009. Bicycle-Sharing: History, Impacts, Models of Provision, and Future, *Journal of Public Transportation*, 12(4), 41-56.

DeMaio, P., Meddin, R., 2013. The Bike Sharing World Map. http://bike-sharing.blogspot.com.

Di Febbraro, A., Sacco, N., Mahnam, S., 2012. One-Way Carsharing: Solving the Relocation Problem. *Transportation Research Record* 2319, 113-120.

Efthymiou, D., Antoniou, C., Waddell, P., 2013. Factors affecting the adoption of vehicle sharing systems by young drivers. *Transport Policy*, 29, 64-73.

Erdoğan, G., Laporte, G., Wolfer-Calvo, R., 2012. The One-Commodity Pickup and Delivery Traveling Salesman Problem with Demand Intervals. Working paper

Forma, I., Raviv, T., Tzur, M., 2013. Optimization based Heuristic for Large Scale Static Repositioning Problems. Working paper, Tel-Aviv University.

Fricker, C., Gast, N., 2013. Incentives and regulations in bike sharing systems with stations of finite capacity. http://arxiv.org/pdf/1201.1178v2.pdf

George, D-K., Xia, C-H., 2011. Fleet-sizing and service availability for a vehicle rental system via closed queueing networks. *European Journal of Operational Research*, 211(1), 198-217.

Kanter, E., 2013. Autoblog.green. http://green.autoblog.com/2013/03/13/bollore-bluecar-ev-priced-at-12-000-euros-plus-80-euro-month-ba/

Kek, A., Cheu, R., Meng, Q., Fung, C., 2009. A Decision Support System for Vehicle Relocation Operations in Carsharing Systems. *Transportation Research Part E*, 45, 149–158.

Kumar, A., Teo, KM., Odoni, AR., 2013. A Systems Perspective of Cycling and Bike-sharing Systems in Urban Mobility. Working paper, Singapore-MIT Alliance.

Lin, J-R., Yang, T-H., 2011. Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E*, 47, 284–294.

Lin, J-R., Yang T-H., Chang, Y-C., 2011. A hub location inventory model for bicycle sharing system design: Formulation and solution. *Computers & Industrial Engineering*. Doi:10.1016/j.cie.2011.12.006

Martens, K., 2007. Promoting bike-and-ride: The Dutch experience. *Transportation Research Part A*, 41, 326–338.

Midgley, P., 2011. Bicycle-sharing Schemes: Enhancing Sustainable Mobility in Urban Areas. *Commission on Sustainable Development.* New York: UN Department of Economic and Social Affairs.

Nair, R., Miller-Hooks, E., 2011. Fleet management for vehicle sharing operations. *Transportation Science,* 45(4), 524–540.

Norris, J-R., 1997. Markov Chains. *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, UK.

Pessach, D., Raviv, T., Tzur, M., 2012. Dynamic Repositioning in a Bike Sharing System. Working Paper, Tel-Aviv University.

Raviv, T., Kolka, O., 2013. Optimal Inventory Management of a Bike Sharing Station. *IIE Transactions,* 45(10), 1077-1093.

Raviv, T., Tzur, M., Forma, I., 2013. The static repositioning problem in a bike sharing system: Models and Solution Approaches. *EURO Journal of Transportation and Logistics,* 2(3), 187-229.

Ross, S-M., 2009. Introduction to Probability Models, 10th Edition – Chapter 6. Academic Press.

Schuijbroek, J., Hampshire, R., van Hoeve, W-J., 2013. Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems. *Tepper School of Business.* Paper 1491. http://repository.cmu.edu/tepper/1491

Shaheen, S., Cohen, A-P., 2007. Growth in Worldwide Carsharing: An International Comparison. *Transportation Research Record,* 1992, 81-89.

Shaheen, S., Cohen, A., 2012. Innovative Mobility Carsharing Outlook, Fall 2012. Transportation Sustainability Research Center, University of California, Berkley.

Shaheen, S., Guzman, S., 2011. Worldwide Bikesharing. *Access*, 39, 22-27.

Shaheen, S., Guzman, S., Zhang, H., 2010. Bicycle Sharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record*, 2143, 159-167.

Schöbel, A., Scholl, S., 2006. Line planning with minimal travel time. Kroon LG, Möhring RH, eds. 5th Workshop on Algorithmic Methods and Models for Optim. Railways number 06901 in Dagstuhl Seminar Proc. 2006 (Internationales Begegnungs-und Forschungszentrum für Informatik, Dagstuhl, Germany).

Shu, J., Teo, C.-P., and Shen, Z.J.M. (2005). Stochastic Transportation-Inventory Network Design Problem. Operations Research, 53(1), 48-60.

Shu, J., Chou, M., Liu, Q., Teo, C-P., Wang, I-L., 2013. Models for Effective Deployment and Redistribution of Bicycles within Public Bicycle-Sharing Systems. *Operations Research,* forthcoming, DOI: 10.1287/opre.2013.1215.

Waserhole, A., Jost, V., 2012. Vehicle sharing system pricing regulation: A fluid approximation. URL http://hal.archives-ouvertes.fr/hal-00727041.

Zhao, D., 2010. Carsharing: A Sustainable and Innovative Personal Transport Solution with Great Potential and Huge Opportunities, *Frost & Sullivan Automotive Practice*.

Appendix A - proof of Lemma 4.

Lemma 4: In a NR policy where $\lambda T = 0$, the expected excess time rate due to waiting is a positive constant independent of λ and T.

Proof: In this case the expected travel time is negligible. The system state representation can be degenerated to account only for the number of vehicles in each station, denoted by $\boldsymbol{y} = (y_1, y_2, \dots, y_S)$. We denote the set of all possible states by Y. A rent transition between two possible states $\boldsymbol{y}, \boldsymbol{y}'$ is denoted by the following indicator function:

$$\epsilon_{ij}(\boldsymbol{y}, \boldsymbol{y}') = \begin{cases} 1, & if\ \boldsymbol{y}' = (y_1, \dots, y_i - 1, \dots, y_j + 1, \dots, y_S) \\ 0, & otherwise \end{cases}$$

The transition rates between any two possible states $\boldsymbol{y}$,$\boldsymbol{y}'$ are given by:

$$\sum_{i,j} v_{ij} \lambda\, \epsilon_{ij}(\boldsymbol{y}, \boldsymbol{y}')$$

The resulting set of steady state equations are:

$$\sum_{\boldsymbol{y}' \in Y} \sum_{i,j} v_{ij} \lambda \pi(\boldsymbol{y}) \epsilon_{ij}(\boldsymbol{y}, \boldsymbol{y}') = \sum_{\boldsymbol{y}' \in Y} \sum_{i,j} v_{ij} \lambda \pi(\boldsymbol{y}') \epsilon_{ij}(\boldsymbol{y}', \boldsymbol{y}) \qquad \forall \boldsymbol{y}$$

An equivalent system of equations can be received by multiplying all equations by $\frac{1}{\lambda}$ (eliminating λ from all equations). Therefore, the limiting probabilities are independent of λ and T. Recall that the Markov chain has a finite state space. Assuming that for any station i there exists at least one station j such that $\lambda_{ji} > 0$, the chain is both irreducible and positive recurrent, see Norris (1997). Therefore, all limiting probabilities are positive. As a result, when $\lambda T = 0$ the expected excess time rate due to waiting in a queue, (3), is equal to some positive constant. ■