Regulating vehicle sharing systems through parking reservation policies: Analysis and performance bounds

Mor Kaspi*, Tal Raviv, Michal Tzur, Hila Galili
Department of Industrial Engineering, Tel-Aviv University, Israel

August 2015

Abstract: We study the regulation of one-way station based vehicle sharing systems through parking reservation policies. We measure the performance of these systems in terms of the total excess travel time of all users caused due to shortages of vehicles or parking spaces. We devise mathematical programming based bounds on the total excess travel time of vehicle sharing systems under any passive regulation (i.e., policies that do not include active relocation of vehicles) and in particular under any parking space reservation policy. These bounds are compared to the performance of several partial parking reservation policies, a parking space overbooking policy, as well as the complete parking reservation (CPR) policy and the no-reservation (NR) policy that were introduced in a previous paper. A detailed user behavior model for each policy is presented and a discrete event simulation is used to evaluate the performance of the system under various settings. The analysis of two case studies of real-world systems demonstrates that: (1) a significant improvement of what theoretically can be achieved is obtained by the CPR policy. (2) The performance of the proposed partial reservation policies monotonically improve as more reservations are required. (3) Parking space overbooking is not likely to be beneficial. In conclusion, our results reinforce the effectiveness of the CPR policy and suggest that parking space reservations should be used in practice, even if only a small share of the users are required to place reservations.

Key words: Transportation, vehicle sharing, bike sharing, one-way car sharing, reservations

1. Introduction
In recent years, vehicle sharing systems (VSS) have become an integral part of the transportation services offered by many cities around the world. Such systems consist of a fleet of vehicles spread across the city, which users can rent for a short period of time. This type of service may be considered as an extension of the traditional public transport, which offers more flexibility and enables more multi-modal journeys. With this added flexibility, more citizens are able to shift from private vehicles to public transportation. In turn, this may result with a decrease in traffic congestion, a better utilization of land resources, especially in city centers (since fewer parking spaces are needed) and a reduction in air pollution and greenhouse gasses emissions.

*Corresponding author at: Department of Industrial Engineering, Tel-Aviv University, Israel. Tel.: +972-3-6405988
E-mail addresses: morkaspi@post.tau.ac.il (M. Kaspi), talraviv@eng.tau.ac.il (T. Raviv), tzur@eng.tau.ac.il (M. Tzur)
In this study, we focus on one-way, station based, VSSs such as bike sharing and some car sharing systems. Such systems allow users to rent a vehicle in any of its stations scattered around the city (given that there is an available vehicle in that station), use it for a short period of time and return it back to any station in which a parking space is available. In the case of bike sharing systems, the "parking space" is in fact a docking pole. However, in the rest of the paper we use the term parking space to refer to one unit of storage of a vehicle of any kind. Some car sharing systems are "free floating" (rather than station based). In such systems the vehicles can be rented and returned at any point in the city. These systems are not in the scope of this study. For a detailed description of the structure of VSSs, the renting process, types of users and different operating models, see surveys by: Shaheen and Cohen (2007, 2012), DeMaio (2009), Shaheen et al. (2010), Shaheen and Guzman (2011), Jorge and Correia (2013) and Demaio and Meddin (2014).

The operators of VSSs face a difficult task of meeting the demand for vehicles and for available parking spaces. Indeed, from on-line reports on the number of vehicles in many VSSs one can learn that stations become empty or full frequently (see for example bikes.oobrian.com). This difficulty arises mainly from the characteristics of the demand for journeys during the day. These demand processes are typically stochastic, asymmetric and heterogeneous in time. The system is unable to satisfy the demand either when a user who wishes to rent a vehicle arrives at an empty station or when a user who wishes to return a vehicle arrives at a full station, i.e., a station with no vacant parking spaces. The latter scenario is typically perceived as more severe since a user who is unable to return a vehicle is “trapped” in the system because she cannot complete the renting transaction until she finds an available parking space. Contrary to that situation, a user who cannot rent a vehicle may decide to use an alternative mode of transportation.

Managers of one-way VSS should aim to improve the quality of service provided to its users, subject to the available resources. In this study, we measure the quality of service by the total excess time users spend in the system as a result of shortages in vehicles or parking spaces. The excess time of a user is the difference between the actual time she spends in the system (her exiting minus entering time) and her ideal time, i.e., the riding/driving time between her origin and destination stations. Indeed, we believe that time is a major consideration of commuters in an urban public transit system, and associated with the main cost incurred by them. This is especially true in cities where regular commuters can buy a monthly or annual subscription for the VSS and for the public transit.

The fact that the user cannot use the VSS for her journey, or for part of it, may cause her an additional damage apart from the time she loses. The cost of this damage, in units equivalent to the cost of time, can be added to the measured excess time. For example, if the user needs to take a taxi instead of renting a shared car, the excess cost of the taxi fare (as compared to the cost of renting a shared vehicle) can be
weighted and added to the excess time. While we use the term excess time throughout the paper, it can be replaced by excess cost to represent more general settings. This observation widens the scope of our discussion here and allows it to capture systems with diverse characteristics, e.g., both bike sharing and one way station based car sharing.

Alternative performance measures for the quality of service are: the proportion of empty or full stations, the percentage of users who receive an ideal service and the percentage of users who do not use the system at all, i.e., those who abandon the system. All these measures are positively correlated with excess time (see Kaspi et. al., 2014) but they do not represent directly the inconvenience caused to the user. Excess time is also applicable for situations in which the service cannot be provided ideally at the desired origin and/or destination stations but a substitute service can be provided at nearby stations. For example, if no vehicle is available at the desired origin station the user may rent a vehicle in a nearby station. In such a case the excess time is the net additional time incurred due to using alternative modes of transportation from the desired origin station to the actual renting station.

To reduce the occurrence of shortages in vehicles or parking spaces, the system operators may take either strategic or operational actions. Strategic actions include deploying more stations or enlarging existing stations, see, for example, Lin and Yang (2011), Lin et al. (2011), George and Xia (2011), Correia and Antunes (2012), Shu et al. (2013), Correia et al. (2014) and Boyaci et al. (2015). Operational actions may include dynamically changing the fleet size and actively or passively regulating the system.

By active regulation, we refer to redistributing vehicles between the system’s stations using repositioning trucks (in the case of bike sharing systems) or by designated drivers (in the case of car sharing). Raviv and Kolka (2013), Schuijbroek et al. (2013) and Vogel et al. (2014) devise methods for determining the desired daily initial inventories in the stations, that repositioning should aim to achieve. Kek et al. (2009), Nair and Miller-Hooks (2011), Benchimol et al. (2011), Angeloudis et al. (2012), Chemla et al. (2013a), Raviv et al. (2013), Erdö?an et al. (2014, 2015), Forma et al. (2015), and others, study static repositioning operations. Contardo et al. (2012), Jorge et al. (2014), Kloimllner et al. (2014) and Pessach et al. (2014) study dynamic repositioning operations. However, repositioning of vehicles may be a costly procedure, especially in car sharing systems where each car is repositioned by a designated driver.

By passive regulation, we refer to mechanisms used for redirecting demand in order to improve the performance of the VSS. Such mechanisms do not affect the true demand for journeys but may instead shift users to rent (return) vehicles at stations different from their true origin (destination) station. Fricker and Gast (2014) study a system regulation under which each user declares two optional destination stations and the system directs her to the less congested one. Several studies focus on pricing regulations as means of self-balancing VSSs, see, for example, Chemla et al. (2013b), Pfrommer et al. (2013) and
We note that the study of Waserhole et al. (2013) does not fall within our definition of passive regulations since they assume that the demand is elastic to the price.

In a previous paper, Kaspi et al. (2014) proposed implementing parking space reservations in one-way VSSs in order to improve the quality of service provided by the system. In particular, they studied a complete parking reservation policy (CPR) in which all users are required, upon renting a vehicle, to reserve a parking space in their destination station. If there is a reservable parking space (i.e., not occupied and not reserved), it will be reserved for the user, and will not be available for other users from the moment the rent starts until the user returns the vehicle to the reserved parking space. If upon renting a vehicle there are no reservable parking spaces at the destination, the renting transaction will be denied. The user may then try to make a reservation at another station near her destination or may decide to use an alternative mode of transportation.

Under the CPR policy, a reserved parking space remains empty until the user returns her vehicle. In the meantime, other users cannot use this resource, i.e., it is blocked. The tradeoff in implementing such a policy is that while some users are guaranteed an ideal service (since they will certainly be able to return their vehicle at their desired destination) other users may receive poorer service due to the blocking of parking spaces. In Kaspi et al. (2014), the CPR policy was compared to the base policy entitled no-reservations (NR), using both a Markovian model with simplifying assumptions and an enhanced discrete event simulation model. Both policies are complete in the sense that all the users in the system are required to follow the same regulation. The results of the analysis demonstrated that the CPR policy outperforms the NR policy with respect to several service-oriented performance measures.

In this study, we examine whether and to what extent a further reduction in the total excess time may be achieved by any other passive regulation, and especially by any other parking reservation policy. Towards that we use mathematical programming models to devise lower bounds on the total excess time that users spend in the system under any passive regulation and under any parking reservation policy. We consider the benefit of limiting the requirement to make a reservation to only some of the journeys. We refer to these policies as partial reservation policies, which combine the two extreme (complete) policies in different ways. We evaluate the performance of all policies and compare them to the lower bounds.

We note that while mathematical programming may not be perceived as the natural choice to analyze such a system, its advantage is in its ability to process an extremely large number of possible occurrences. On the other hand, in order to model the VSS using stochastic tools, for example closed queueing systems, some simplifying assumptions are required to make the model tractable. Fricker and Gast (2014) and Kaspi et al. (2014) make simplifying assumptions regarding the behavior of users when they face vehicle/parking space shortages. George and Xia (2011) assume that the capacities of the stations are unlimited. Indeed, using these assumptions tractable models are generated. However, they do not reflect
the true dynamics of the VSS since the interactions between neighboring stations due to shortages are neglected.

The focus of this study is on improving VSSs from the user perspective. Indeed, the operators have their own perspective that reflects their revenue and costs, which plays a crucial role when making strategic decisions about the design of the system, about active regulation, pricing etc. As for the short run, when a reservation is denied, the operator indeed may lose revenue, thus potentially creating a conflict between the user's and the operator's goals. However, under a good reservation policy, this would provide better service to the users and thus increase revenue in the long run. Therefore, when optimizing the parking reservation policy, it is reasonable to consider only the quality of service as an objective function.

The contributions of this paper are: First, using mathematical programming models, we provide for the first time lower bounds on the performance of a VSS, measured by the total excess time, under any passive regulation and under any parking reservation policy. Second, we introduce the concept of partial reservation policies. We examine three different partial policies, each is based on a simple sound principle, which is easy to control by the system’s managers and communicate to the users. We define the user behavior under these policies and examine their performance using discrete event simulation of real world systems. Third, we examine the potential benefit of parking space overbooking.

The rest of the paper is organized as follows: in Section 2, a generic description of the VSS is presented and mathematical models are formulated to bound its performance under passive regulations and in particular under parking reservation policies. In Section 3 a behavior model of users in VSSs is presented and the proposed partial parking reservation policies are described. A utopian overbooking policy is presented at the end of this section. A description of two real world VSSs and numerical results for their performance are presented and discussed in Section 4. Concluding remarks are provided in Section 5.

2. Lower bounds on the total excess time in a VSS

As mentioned in the introduction, the operational actions that a VSS operator can take in order to provide a high quality service can be of two forms, namely, active regulations and passive regulations. In this study, we focus on passive regulations, i.e., mechanisms used for redirecting demand.

VSSs are decentralized systems, that is, each user makes decisions regarding her planned itinerary so as to minimize her own expected excess time. Such decisions depend on the availability of vehicles or parking spaces at the system’s stations at the renting time as well as on the user’s expectations regarding future availability. In addition, the user’s decisions are subject to the passive regulation prescribed by the system. Under passive regulations, the system may influence the users’ decisions by limiting their choices
or by incentivizing them to prefer certain itineraries. However, the system does not assign itineraries to the users. For example, under the CPR policy, if a user cannot make a parking reservation at a certain station, she is not allowed to travel with a shared vehicle to that station, but the choice of her actual alternative itinerary is hers. From the operator’s point of view the question is: how to design a passive regulation so that the outcome of all users’ decisions minimizes the expected total excess time?

A passive regulation can be formally defined as a mapping of the state of the system and the demand for journeys to a set of itineraries allowed for each journey. The set of possible passive regulations is enormous. However, a major share of these regulations may be hard to implement or hard to communicate to the users. In this study we introduce and analyze regulations in the form of parking reservation policies, which are based on simple principles and are easy to communicate to the users. In order to assess the potential improvement that may be achieved by passive regulations in terms of the expected total excess time, we formulate mathematical programs that provide lower bounds. First, we devise a lower bound on the expected total excess time under any passive regulation. Second, since this study focuses on parking reservation policies, we devise a tighter bound specifically for any parking reservation policy.

The rest of this section is organized as follows: Section 2.1 includes a description of a VSS, and assumptions concerning the demand. Section 2.2 contains a mixed integer program (MILP) whose optimal value is a lower bound on the excess time that may be achieved by any passive regulation. Section 2.3 modifies the MILP formulation to account for passive regulations that involve only parking reservations, thus obtaining a tighter bound. A formal proof for the validity of this lower bound is presented.

2.1. Description of the VSS

In this section we discuss the information needed in order to model a VSS. Such information is used both in the mathematical models that are presented in this section, and in the user behavior model presented in Section 3. The information needed in order to describe the system is as follows:

- The number of stations in the system
- The number of parking spaces in each station (referred to as the capacity of the station)
- The initial inventory level (number of vehicles) at each station
- The expected travel time between any pair of stations using the shared vehicles.
- The expected travel time between any pair of stations, using an alternative mode of transportation.

Note that information about the locations of the stations is not needed. In order to describe the relations between the stations, it is enough to specify the traveling time between each pair of stations. The distance between the stations, the topography of the city, congested roads and other considerations are all
taken into account in the traveling times. In some situations, due to shortages in vehicles or parking spaces, users may roam to nearby stations (using an alternative mode of transportation) or may decide to abandon the system altogether and make their entire journey using an alternative mode of transportation. Therefore, the traveling times between any pair of stations using an alternative mode of transportation are also needed. The travelling time can represent any additional cost or inconvenience incurred by the user in her journey, in addition to the actual value of the time spent in the system. In bike sharing systems the alternative mode for most of the potential journeys is walking. This is also the case in car sharing systems for roaming between neighboring stations.

Each demand for a journey is defined by the desired origin and destination stations and by the desired starting time. An underlying assumption of the models introduced in this paper is that all the journeys in the system start and end at stations of the VSS. In reality, users’ journeys start and end at general locations (GPS points) in the city, however, such fine spatial granularity is not required for strategic decisions on reservation policies. Moreover, information regarding exact origins/destinations is currently not available for the operators of VSS. Finally, we assume that an alternative mode of transportation is always available to the users while the shared vehicle (resp., parking space) may or may not be available at the origin (resp., destination).

2.2. A lower bound on the total excess time under any passive system regulation

Our goal in this section is to establish a lower bound on the total excess time that results from the users’ decisions under any passive regulation. Given the system's characteristics and a demand realization for journeys during a predetermined planning horizon (typically a day), we formulate an optimization problem that centrally selects the itineraries of the VSS's users so as to minimize the total excess time.

The solution value of this optimization problem is a lower bound on the total excess time that may be achieved by any passive regulation, due to the following two assumptions on which the optimization problem is based:

1. All demands for journeys are known in advance.
2. A central planner determines the itinerary of each user, in a way that benefits the entire system. The justification of this assumption is that any solution selected by the central planner may be selected by the users under some passive regulation.

In practice, each user determines her own itinerary, according to her individual objective and given the information she has. Thus, the excess time of the optimal assignment obtained by a central planner with full information is a lower bound on the excess time that results from any passive regulation policy, i.e., a policy that affects the itinerary selection of the users in some restricted way. We note that due to the
system's limited resources, this bound is typically strictly positive, thus it is better than the trivial bound of zero excess time (no shortages of any type).

In practice the demand for journeys is a stochastic process. Therefore, the average solution value of the optimization problem for numerous demand realizations, drawn from a given stochastic process, provides an estimator for a lower bound on the expected total excess time under any passive regulation.

A demand realization is described by a set of journeys where each journey is characterized by an “origin-destination-time” tuple. Each journey can be materialized by one of several possible itineraries. We assume that a possible itinerary is in one of the following forms:

a) Use a shared vehicle all the way from the origin to the desired destination.
b) Use a shared vehicle from the origin station to another station with an available parking space and then use an alternative mode to get to the desired destination.
c) Use an alternative mode of transportation to get to a station with an available vehicle and then use a shared vehicle from this station to the desired destination.
d) Use an alternative mode of transportation to get to a station with available vehicle and then use a shared vehicle from this station to another station with an available parking space. Then, from this station use an alternative mode to get to the desired destination.
e) Use an alternative mode all the way from the origin to the desired destination.

We refer to the station where the vehicle is actually rented (resp., returned) as the renting (resp., returning) station. Upon attempting to return a shared vehicle, the user may be required to wait at the returning station until a parking space becomes available, and then proceed with her itinerary (leave the system or continue with an alternative mode of transport). We assume that the user will not wait for a vehicle to become available in a renting station, since information regarding the number of vehicles in each station is available to her in real time. Instead, she would roam to a nearby station or use an alternative mode of transportation for the entire journey.

In Figure 1 we present an example with several possible itineraries that materialize the journey of a user who wishes to travel from station A to station B. The travel time of each segment of the journey is depicted on the corresponding arc and the itinerary's excess time (denoted by X) is presented below each graph. For example, in Figure 1(b) the excess time is 5 since the travel time is 12, as compared to a travel time of 7 in the ideal journey, see Figure 1(a). In Figures 1(a) – 1(e), we present examples for each of the five forms of itineraries presented above, respectively. Note that since the excess time associated with using an alternative mode of transportation for the entire journey, as in 1(e), is 14, itineraries with larger excess time, such as 1(f), will never be selected by a user and thus can be excluded from consideration of the central planner.
Each possible itinerary can be defined by its renting station, returning station and renting time. The returning time is determined by the renting time and the traveling time between the two stations. Waiting times at the returning stations are not considered as part of the itinerary times, they are calculated separately. In addition, a journey can be materialized by an itinerary that includes only an alternative mode of transportation. Clearly, such an itinerary is not associated with renting and returning stations.

![Figure 1: Example, possible itineraries that materialize a journey](image)

We define a set of possible events, where each event is a time-station tuple that refers to a renting or returning time and location of a possible itinerary. We assume without loss of generality that at most one event can occur at each station at a given time. At the time of each event the state of the corresponding station is defined by the number of vehicles parking at the station and the number of users that are (possibly) waiting to return their vehicle at the station.

The assignment of itineraries to users, carried out by the central planner, is constrained by several considerations that are related to the availability of vehicles and parking spaces at the stations. In Figure 2, we use a network flow graph to present the movement of vehicles in the system over time. We use the possible itineraries of the journey depicted in Figure 1 as an example, assuming that the journey starts at time 20. Each possible itinerary that includes a movement of a vehicle [itineraries (a)-(d) of our example]
is depicted by a black solid arc from a node that represents the renting time and location to a node that represents the returning time and location. The costs of these arcs are the excess times associated with their itinerary and their capacities are 1. The use of an alternative mode of transportation is not reflected directly by arcs in the network. However, the times of the nodes and the costs of the arcs are affected by the use of the alternative modes of transportation. For example, in itinerary (c), depicted in both Figure 1 and Figure 2, the use of a shared vehicle begins at station D at time 26 although the itinerary starts at station A at time 20. The cost of this arc is 5, which represents the excess time of the itinerary.

To represent an entire demand realization we construct a network as in Figure 2, with a set of nodes and arcs for all possible itineraries of all demanded journeys. The nodes in this network correspond to events. Each pair of consecutive nodes on the time axis, that are associated with the same station, is connected by two “horizontal” parallel arcs. The solid gray arc represents the number of vehicles parked in the station between the two events and the dashed arc represents the number of vehicles (and drivers) waiting in the station for a vacant parking space, during this time interval. Since the pair of nodes is consecutive, the number of vehicles parking and waiting in the station does not change during this time interval. The costs of the parking arcs are zero and their capacity equals the capacity of the station. The per unit cost of the waiting arcs equal to the time difference between their end nodes and their capacity is not limited. For example, if the flow on the waiting arc that connects nodes (B,27) and (B,32) is 3 then the excess time due to waiting in station B between time periods 27 and 32 would be 15.

![Figure 2: Network flow graph representing the flow of vehicles in the system](image)

The network also includes one source node for each station, with a supply that represents the initial inventory in the station and one sink node. The net demand of the rest of the nodes is zero. A feasible assignment of itineraries to journeys is obtained as a feasible integer flow on this network with additional side constraints. These constraints limit the total flow on all the itinerary arcs associated with each
journey to at most 1. A solution where the total flow on the arcs associated with a certain journey is zero represents selection of an itinerary that uses an alternative mode only, e.g. Figure 1(e). The excess time incurred in such a solution is the sum of the flow costs plus the costs due to the journeys that use alternative modes only. Thus, our lower bound is obtained by minimizing this excess time. We solve this optimization problem using an MILP that is formulated below. Next, the notation required to formulate this model is presented.

Indices:
- $s$: Station
- $t$: Time
- $j$: Journey
- $i, k$: Itinerary

Parameters:
- $S$: Set of stations
- $J$: Set of journeys
- $C_s$: Capacity of station $s$
- $L_s^0$: Initial vehicle inventory in station $s$
- $E$: Set of possible events $\{(s, t)\}$
- $I_j$: Set of possible itineraries of journey $j$, we also use $I \equiv \cup_{j \in J} I_j$
- $X_i$: The excess time of itinerary $i$ (not including waiting time)
- $D_{s,t}$: The time difference between event $(s, t)$ and the next event at station $s$
- $B(s, t)$: Set of itineraries in which a vehicle is rented at station $s$ at time $t$
- $F(s, t)$: Set of itineraries in which a vehicle is returned at station $s$ at time $t$
- $(s, t)'$: The event that precedes event $(s, t)$ at station $s$

In addition, we define two artificial events $(s, 0)$ and $(s, H)$ for each station $s$, that denote the beginning and the end of the planning horizon, respectively. Note that $X_i$ represents the excess time associated with selecting itinerary $i$. This excess time includes the additional time incurred by using alternative modes of transportation to materialize the entire journey or part of it. It does not include additional excess time that the user may experience due to waiting for a vacant parking space at the returning station. This waiting time is reflected by the parameter $D_{s,t}$. Without loss of generality the set $I_j$ consists only of journeys with excess times that are not greater the excess time of using alternative mode for the entire journey.

Decision variables:
- $r_i$: 1 if itinerary $i$ is selected, 0 otherwise
- $p_{s,t}$: Number of vehicles parking at station $s$ right after event $(s, t)$
- $w_{s,t}$: Number of users waiting to return a vehicle at station $s$ right after event $(s, t)$

With respect to the network flow model, the $r_i$ variables represent the flows on the itinerary arcs. The $p_{s,t}$ variables represent the flows on the parking arcs and the $w_{s,t}$ variables represent the flows on the waiting
The objective function (1) sums over the excess time of the selected itineraries and the waiting times of all the users who wait to return their vehicle at their returning station. These are the two components of the total excess time of all the users in the system. Constraints (2) assure that for each journey exactly one itinerary is selected. Constraints (3) are vehicle inventory balance equations. Constraints (4) set the initial vehicle inventory in each station. Constraints (5) limit the number of parked vehicles in a station to the capacity of the station. Constraints (6) and (7) state that no user is waiting to return a vehicle at the beginning or at the end of the planning horizon. Constraints (8) stipulate that the itinerary decision variables are binary. Constraints (9) and (10) are non-negativity constraints on the number of parked vehicles and waiting users after each event.

In this model the central planner may assign a user with any of its given potential itineraries. In some cases the users may be referred to relatively distant rent or return stations, merely in order to balance the vehicle inventories, for the system’s benefit, and not necessarily because the system cannot satisfy their demand by better itineraries. In the next section we extend the model to limit such occurrences.

Theoretically, a user may begin her ride and return the vehicle at any station. Therefore, the number of potential itineraries of a journey is the square of the number of stations. However, most of these potential itineraries would take longer than simply using the alternative mode of transportation for the
entire journey (i.e., abandoning the system). Under many regulations it is safe to assume that users will not accept such itineraries. In the numerical experiment reported in Section 4, we let the central planner consider only potential itineraries that are not longer than using the alternative mode of transportation for the entire journey. Moreover, in order to reduce the computational effort required for solving the PR-LB model (1)-(10), we relaxed the integrality constraints (8) and replaced them with non-negativity constraints. This clearly preserves the result as a lower bound. In our numerical experiment we observed that the effect of this relaxation on the obtained lower bound is negligible, see the discussion in Section 4.

When restricting ourselves to itineraries that are not longer than using the alternative mode of transportation, an alternative lower bound on the total excess time could be obtained by including the possible waiting times at the destination within the itineraries and removing the waiting variables \((w_{s,t})\) from the model. In such a formulation we could allow only sequences of travel and waiting times that together are shorter than using the alternative mode of transportation for the entire journey. This further limits the decision space of the central planner and thus may result in a tighter lower bound. However, such an approach would result in a significantly larger number of possible itineraries (and thus decision variables). In the instances that we have solved, we observed that the total waiting time was negligible, as compared to the total excess time. Therefore, we believe, that the potential improvement of the lower bound is not significant.

Though this study focuses on parking reservation policies, the above model serves as a lower bound on the excess time under any passive regulation. In particular, since the input for this model consists of all the demand for journeys it can also serve as a lower bound for vehicle reservation policies, trip reservation policies, the best of two regulation proposed by Fricker and Gast (2014), and the pricing regulations proposed by Chemla et al (2013) Pfrommer et al. (2013).

2.3. A lower bound on the total excess time under any parking reservation policy

In this section we focus on a subset of all possible passive regulations, namely, parking reservation policies. A parking reservation is a process in which, when attempting to rent a vehicle, the user declares on her destination and the trip is either allowed or denied by the system. If allowed, a parking space is reserved for the user at the desired destination. If denied, the user may try to place reservations to other destinations until one is allowed. A parking reservation policy is a set of rules which determine: in which subset of trips the user is required to place reservations, whether a reservation request is allowed or denied and when allowed, whether the parking space is reserved for the user temporarily or permanently (until her arrival to the destination). The operator is allowed to overbook parking spaces that are currently not available. However, for a parking reservation policy to be enforceable and sustainable over time, the
operator must not deny parking reservation requests unjustifiably. Next, we formally define such a set of parking reservation polices, to be studied in this paper.

**Definition 1 (A parking reservation policy).** A passive regulation in which the operator can deny renting a vehicle only if there are no reservable parking spaces at the destination at the renting time.

Recall that under a parking reservation policy, reservation is not always required, but when it is required, the condition of definition 1 must hold. For example, the CPR and NR policies are both legitimate parking reservation policies. In the CPR policy, reservation of a reservable space is always required. The NR policy satisfies the requirement of definition 1 trivially, since under this policy no reservation is required and thus reservations are never denied.

Under any parking reservation policy, the set of possible itineraries that can materialize journey \( j, I_j \), can be partitioned into three subsets given the state of the system when the journey begins: (I) Itineraries that cannot be denied by a parking reservation policy. This set includes any itinerary with an available vehicle at its renting station at its renting time and a reservable parking space at its returning station at the renting time. In addition, the itinerary that consists of the alternative mode only is always included in this set, since it is also an itinerary that cannot be denied. (II) Itineraries that can be either denied or allowed by a parking reservation policy. This set includes any itinerary with available vehicle in its renting station at its renting time but no reservable parking space at its returning station at the renting time. Assignment of itineraries from this set may be materialized in overbooking policies or in partial reservation policies where some users start their journey without reservations at all. (III) Itineraries that cannot be allowed by a parking reservation policy. This set includes any itinerary with no available vehicles at its renting station at its renting time. The parking reservation policy dictates which of the itineraries in (II) are available to the user. The user, from her side, selects the itinerary from (I) or from the allowed itineraries in (II) that minimizes her excess time.

Note that under a general passive regulation, the system may offer to the user any subset of itineraries from the union of (I) and (II), as long as this subset contains the itinerary that uses the alternative mode only. However, under a parking reservation policy the offered subset must include all the itineraries in (I) and possibly some of the itineraries in (II). Thus, in these policies the system has less control over decisions of the users.

The PR-LB model (1)-(10) is modified such that the central planner may assign to each journey either the shortest itinerary in (I) or a shorter itinerary from (II). Recall that in the original model, any itinerary from the union of (I) and (II) could be assigned. The partition of potential itineraries to the sets (I), (II), and (III) cannot be pre-defined as an input to the model. This is because the selection of the itineraries that are included in these subsets depends on the system’s state at the decision time and on all the decisions that were made for journeys that begin prior to that journey. Instead, we modify the PR-LB
model (1)-(10) by adding sets of decision variables and constraints so as to properly assign itineraries that users who minimize their excess time would actually choose under some parking reservation policy. We refer to this extended model as the Parking Reservation Policy Lower Bound model (abbreviated PRP-LB). We use the same notation as in the PR-LB model (1)-(10) and add the following parameters and decision variables:

Parameters:
- \( O(i) \) A \((s, t)\) tuple that represents the renting station and renting time of itinerary \( i \)
- \( D(i) \) A \((s, t)\) tuple that represents the returning station and returning time of itinerary \( i \)
- \( J(i) \) The journey that can be materialized by itinerary \( i \)
- \( T(s, t) \) Time of node \((s, t)\)
- \( S(s, t) \) Station of node \((s, t)\)
- \( \mathcal{R}_i \) A set of itineraries for which a parking space may be reserved at the returning station of itinerary \( i \) at the renting time of itinerary \( i \). That is, an itinerary \( k \) is in the set if:
  - It is of a different journey, \( J(k) \neq J(i) \)
  - It has the same returning station as itinerary \( i \), \( S(D(k)) = S(D(i)) \).
  - The renting time of itinerary \( k \) is earlier than the renting time of itinerary \( i \), \( T(O(k)) < T(O(i)) \)
  - The returning time of itinerary \( k \) is later than the renting time of itinerary \( i \), \( T(D(k)) > T(O(i)) \)
- \( M \) A very large number (for example, twice the capacity of the largest station)

Auxiliary decision variables:
- \( e_{s,t} \) 0 if a vehicle is available at station \( s \) at time \( t \), otherwise it can either be 0 or 1.
- \( f_i \) 0 if at renting time \( T(O(i)) \) there are some reservable parking spaces at station \( S(D(i)) \), otherwise it can either be 0 or 1.

The PRP-LB model can be stated now as (1)-(10) with the following additional constraints:

\[
\begin{align*}
    e_{O(i)} + f_i &\geq r_k \quad \forall i, k \in I \setminus X_i < X_k \quad \forall j \in J & \quad (11) \\
    M \cdot (1 - e_{s,t}) &\geq p_{s,t} + w_{s,t} \quad \forall (s, t) \in E & \quad (12) \\
    c_{s(D(i))} \cdot f_i &\leq p_{D(i)} + w_{D(i)} + \sum_{k \in \mathcal{R}_i} r_k \quad \forall i \in I & \quad (13) \\
    e_{s,t} &\in \{0, 1\} \quad \forall (s, t) \in E & \quad (14) \\
    f_i &\in \{0, 1\} \quad \forall i \in I & \quad (15)
\end{align*}
\]

Constraints (11) stipulate that each journey is materialized by the shortest possible itinerary, i.e., the one with the shortest excess time that is allowed by some parking reservation policy. For any itinerary \( k \), if an itinerary \( i \) of the same journey with shorter excess time that belongs to (I) exists then itinerary \( k \) cannot be selected. Recall that if itinerary \( i \) is in (I) then a vehicle is available at its renting station \( (e_{O(i)} = 0) \) and a parking space is available at its returning station \( (f_i = 0) \). In this case the left hand side of
(11) is zero and thus \( r_k \) is forced to be zero. Note that if itinerary \( i \) is in (II) the right hand side of (11) is greater than zero. In this case, the model may or may not assign itinerary \( k \) to the journey. By constraints (12), a station can be considered "empty" in a given time, only if there are no vehicles parked or waiting in it at that time. Constraints (13) assure that the \( f_i \) variables are set to zero if there are reservable parking spaces at the returning station of itinerary \( i \) at the renting time. The decision variable \( w_{s,t} \) is added to the right hand side of constraints (12) and (13) in order the make sure that the central planner will not “leave” vehicles waiting outside a non-full station in order to gain more flexibility in selecting possible itineraries. Constraints (14) and (15) stipulate that the variables \( e_{s,t} \) and \( f_i \) are binary.

The value of the solution of the PRP-LB model (1)-(15) provides a tighter bound on the total excess time as compared to the PR-LB model (1)-(10). This is because it is based on a super-set of its constraints and since parking reservation policies are a subset of any passive regulation. As in the case of PR-LB, this model was solved while relaxing the binary variables \( r_i \). The binary variables \( e_{s,t} \) and \( f_i \) were not relaxed since without imposing their integrality the resulting relaxation is very weak. This is due to the effect of the big-M terms in constraints (12) and (13). Indeed, this model is more difficult to solve (see Section 4).

In the PRP-LB model, unlike the PR-LB, if vehicles are available at the station at the arrival time of a renter, the system must offer one to the user. Therefore, this model cannot provide a lower bound for the performance of vehicle or trip reservation policies. Next, we formally prove the validity of the optimal solution value of PRP-LB as a lower bound on the total excess time under any parking reservation policy.

Proposition 1: For any demand realization, the total excess time associated with the optimal assignment of itineraries to journeys under PRP-LB is not greater than the excess time under any parking reservation policy.

Proof: Consider the assignment of itineraries to journeys obtained by some parking reservation policy (that satisfies the condition of Definition 1). We refer to this assignment as \( A^* \). We claim that such an assignment can be mapped to a feasible solution of PRP-LB and therefore the optimal solution of PRP-LB is a lower bound on the excess time that results from any parking reservation policy. First note that since \( A^* \) is a feasible assignment of itineraries to journeys, it must satisfy constraints (2)-(10) when setting the \( r_i \) variables to represent the actual itineraries that were selected by the users under policy \( A^* \) and setting the values of the variables \( p_{s,t} \) and \( w_{s,t} \) to represent the number of vehicles that are parking and waiting at the stations after each event \((s,t)\), respectively. Next, we will show that the values of the binary variables \( e_{s,t} \) and \( f_i \) can be set such that the rest of the constraints of PRP-LB can be satisfied. First we set the value of \( e_{s,t} \) as follow

\[
e_{st} = \begin{cases} 
0 & p_{s,t} + w_{s,t} > 0 \\
1 & otherwise
\end{cases}.
\]
Such an assignment would immediately satisfy constraints (12) for each event \((s,t)\). Similarly we set

\[
 f_i = \begin{cases} 
 0 & \text{if } p_D(i) + w_D(i) + \sum_{k \in R_i} r_k < c_{S(D(i))} \\
 1 & \text{otherwise}
\end{cases},
\]

which immediately satisfy constraints (13) for each itinerary \(i\). Now, it is left to show that with this assignments constraint (11) is satisfied for each pair of itineraries of the same journey \((i,k)\) such that \(k\) is selected under policy \(A^*\) and \(X_i < X_k\), that is, itinerary \(i\) has a shorter excess time than itinerary \(k\). Recall that if \(k\) was selected, then \(r_k = 1\). Assume by contradiction, that constraint (11) is violated, which implies \(e_{O(i)} = 0\) and \(f_i = 0\). This means that for itinerary \(i\) a vehicle was available and a reservable parking space was available at the renting time. By Definition 1, such an itinerary could not be denied by a parking reservation policy. Finally, since it is shorter than itinerary \(k\), it must have been selected by the user, which is a contradiction. 

By proposition 1, the assignment of itineraries that can result from any parking reservation policy under any demand realization is a feasible solution of PRP-LP. Thus, the excess time that can be achieved by any parking reservation policy is bounded from below by the optimal solution of the model.

3. Parking reservation policies

The lower bounds developed in the previous section may be used to evaluate the effectiveness of any regulation or parking reservation policy. In this section, we introduce several parking reservation policies. The performance of a VSS under these policies or under any other regulation can be evaluated only with respect to the response of the users to the rules prescribed by the regulation. We base our analysis, with respect to users’ response, on an axiomatic approach and model the users as rational independent agents whose goal is to minimize their own excess time. However, achieving this goal may be too complex for many users due to the stochastic nature of the VSS. Therefore, we postulate a user behavior model that heuristically approximates this minimization problem and, in fact, provides an optimal solution in most of the cases.

In Section 3.1 we present this user behavior model. The model describes the decisions taken by the users at different decision points. These decisions are affected by the state of the system and the settings of the regulation. In Section 3.2 we present three partial reservation policies, discuss the motivation for using them and explain how they are reflected in the user behavior model. In Section 3.3 we present a utopian parking overbooking policy, whose performance is used to gauge the potential benefit of parking overbooking policies.
3.1. User behavior model
The movement of users in the system depends both on its regulation and on the state of the system (availability of vehicles and parking spaces). A user who enters the system acts as follows: If there are no available vehicles at her origin station, she may either decide to go to a nearby station, using an alternative mode of transportation, in search for an available vehicle, or she may decide to abandon the system. An abandoning user is assumed to travel to her destination using an alternative mode of transportation. Note that in modern VSS, the user can make this decision based on real time information about the availability of vehicles in the stations of the system. Once a user finds an available vehicle, there are two options: (1) A parking reservation is not required and (2) a parking reservation is required. In option (1), the user rents a vehicle and travels to her destination. When the user reaches her destination (with a vehicle), if she finds an available parking space she returns the vehicle and exits the system. If there is no available parking space at the destination station, the user may decide to wait at the station until a parking space becomes available (i.e. she enters a waiting queue). Alternatively, the user may decide to roam to a nearby station in search for an available parking space. Again, this decision is based on real time information on the availability of parking spaces in the stations. In option (2), the user attempts to make a parking reservation in her destination station. If the reservation is approved, the user makes a rent-and-reserve transaction and travels to her destination station. If the parking reservation is guaranteed, the user can immediately exit the system upon reaching her destination. If the reservation is not guaranteed, the user travels to the returning station and proceeds as in option (1). If the parking reservation is not approved, the user can either try to make a reservation at another station near her destination or she may decide to abandon the system. Finally, if, for some of the above reasons, the vehicle is returned to a station different than the user’s destination, the user uses an alternative mode of transportation to reach her destination station and then exits the system.

This behavior model is described in Figure 3. At decision points, we assume that users have full knowledge of the system state, including inventory levels at each station and the arrival rate of renters to each station (for example, the operator, or a third party, can provide this information via a smartphone application). Users are assumed to be strategic so that at decision points they will choose the alternative that minimizes their expected remaining traveling time. An alternative user behavior model could be based on the maximum utility theory, introducing randomness in the itinerary selection decisions and reflecting factors that are not included in the current model. However, we use a deterministic itinerary selection model that is based solely on excess time, because it is based on data that is readily available to the operators. We believe that such a model is sufficiently accurate to provide insights on the effect of various parking reservation polices.
We further elaborate on the user decision processes, denoted in Figure 3 by I, II and III:

I. A renter who arrives at a station with no available vehicles would consider a nearby station such that the total time spent using an alternative mode of transportation to get to that station and the traveling time from that station to the destination, is the shortest among all stations with available vehicles. The user would choose an alternative mode of transportation for the entire journey, if it is shorter than the above alternative.

II. A renter who arrives with a shared vehicle to a station with no available parking spaces would consider a nearby station such that the total time spent traveling with the shared vehicle to that
station and using an alternative mode of transportation from there to the destination is the shortest among all stations with available parking spaces. The user would choose to wait in the station until a parking space becomes available, if the expected time until this happens is shorter than the above alternative.

III. A renter who cannot make a parking reservation at the destination station would consider making reservation at a nearby station such that the total time spent traveling with the shared vehicle to the chosen returning station and using an alternative mode of transportation from there to the destination is the shortest among all stations for which it is possible to make parking reservation. The user would choose using an alternative mode of transportation for the entire journey if it was shorter than the above alternative.

In the user behavior model, there are three junctions that represent the settings of the policy:

- Parking reservation required?
- Reservation approved?
- Vacant parking space guaranteed?

To highlight these junctions, we plot them in Figure 3 as trapezoids. The NR and CPR policies, are complete in the sense that under each of these policies the answer to each of the above three questions is identical for all the users of the system. For example, under the CPR policy, all users are required to make a parking reservation, the reservation is approved if at the renting time there is an available parking space at the returning station and a vacant parking space is guaranteed to all users who are able to make a parking reservation.

3.2. Partial parking reservations policies
In this section we present three types of partial parking reservation policies. Each type is based on a simple, yet reasonable principle. The common motivation for these policies is to enforce parking reservations only when they are likely to make a positive effect on the performance of the system. In the descriptions below, a trip is defined as a direct ride between a pair of origin-destination stations.

3.2.1. Trip based partial reservation policy
Under this policy, parking reservations are required only for trips with expected traveling times shorter than a given threshold. At rent, a user will declare her destination and if the expected traveling time is shorter than the given threshold she will be required to reserve a parking space in her destination. As in the CPR policy, if at the renting time there is no vacant parking space at the destination, the transaction is denied and the user may try to make a reservation at a different station. A user with expected riding time longer than the threshold time, will not be required to make a parking space reservation. If such a user
finds an available vehicle at her origin, she will be able to rent it and ride to her destination, as in the NR policy. The rationale behind this policy can be stated as follows: since a parking space is a valuable resource in a VSS and a reservation practically blocks it for the duration of the trip, the parking space should be reserved only for short trips. Moreover, users with short traveling times may be more sensitive to excess time due to shortage of parking space at the destination.

Note that if the threshold time is set to zero, this policy coincides with the NR policy. On the other hand, if the threshold is set to a large enough value this policy coincides with the CPR policy. Different partial policies of this type can be obtained by setting the value of the threshold parameter between the two extremes.

3.2.2. Station based partial reservation policy

Under this policy a parking reservation is required only if the difference between the expected returning and renting rates at the destination station during a certain time interval is higher than a pre-specified value, referred to as the difference threshold. Otherwise, no reservation is required. Expected renting and returning rates can be estimated using past transactions. The difference is calculated for each station during predefined time intervals of each day. If the calculated difference is lower than the difference threshold the user will behave as under the NR policy.

The rationale behind this policy can be stated as follows: the probability of parking space shortages in a station grows as the imbalance (difference) between the demand rates for parking spaces and for vehicles grows. Such imbalances may be consistent, for example, in bike sharing stations at relatively low altitude locations, where bicycles are more likely to be returned than rented. Or, the imbalance may change during the day, for example, in stations located at working areas where returning in the morning is much more prevalent than renting. When the demand rate for parking spaces (returning) is higher than the demand rate for vehicles (renting), there is a greater chance that users will find the station full. By enforcing parking reservations at such stations, the system prevents users from traveling to stations with no available parking spaces. This will shift some of the users to nearby, less congested stations. Such a shift would have probably happened anyway, since users who find a full station typically roam to a nearby station in order to return their vehicle. With parking reservations, the change in the retuning station is determined in advance, which is likely to reduce the users’ excess time. On the other hand, it seems less effective to enforce parking reservations in stations that are likely to be empty anyway.

Note that the higher the difference threshold is, the fewer the cases in which reservations are required. For extremely high threshold values the policy coincides with the NR policy, while for extremely low (negative) values it coincides with the CPR policy.
3.2.3. Time limited partial reservation policy

Under this policy all users are required to make a parking reservation, as in the CPR policy, but this reservation is valid for a limited time. After the reservation expires, the reserved parking space becomes available to other users, and a vacant parking space is no longer guaranteed to the user. If the reservation expires and no parking space is available by the time the user arrives at the destination station, she will have to either wait by the station or roam to a nearby station (as in the NR policy).

The rationale behind this policy can be stated as follows: by making a reservation, a user with a long traveling time that reaches her destination only after her reservation expires, still affects the system because as long as her reservation is valid, she may block other users from making a reservation. That is, the reservation may divert subsequent demand, which may increase the probability of the user to find a vacant parking space, even if her reservation expired.

Note that if the time limit is set to a large enough value, this policy coincides with the CPR policy. However, if the time limit is set to zero, the resulting policy is still different from the NR policy. This is because users still must make a reservation, and they cannot begin traveling to a station that is full at the renting time. In Section 4 we compare the performance of this specific setting (in which the time limit is zero) to the performance of the NR policy and discuss the differences and their implications.

3.3. Utopian parking space overbooking policy

In many service systems in which reservations are carried out, it is a common practice to allow overbooking. It means accepting reservations for resources that, based on previous reservations, will not be available at the required time. Overbooking may be an effective policy in the presence of stochasticity in the arrival or service process that expects no-shows of customers who made reservations. In a VSS that practices parking reservations, no-shows are not an issue because the reservations are made at the renting time and the users must return the vehicle at the stated destination. Nevertheless, it might be beneficial to allow in some cases for users to travel to a station even if it has no available parking spaces, or in other words, to allow overbooking. This is because a parking space may become available by the time the user reaches her destination station. A good overbooking policy is based on reliable forecast that is capable of predicting such occurrences.

In order to evaluate the potential benefit of overbooking policies, we envision a system that has full information regarding the demand for vehicles at the station for which it considers allowing overbooking. The overbooking decisions will be based on this information, thus it is referred as a utopian overbooking policy. Note however, that this policy optimizes the service provided to each user individually rather than taking the system point of view as in the lower bounds presented in Section 2.
Under this policy, upon renting the user is required to declare her destination, and then the system decides whether a reservation can be made or not. The system’s decision is made based on knowledge of the current state of the destination station, including users who are on their way to that station with a vehicle, and of all future arrivals of renters to that station (including their exact arrival times). We refer to the system decision process as look-ahead, since the system’s decision is based on anticipating whether there will be an available parking space in the destination station upon arriving to it. The look-ahead algorithm is presented in Table 1. We use the following notation to describe it:

- \( E \) A list of future events in the returning station, including return events of reservations that were already approved and all future rent events. The list is sorted in ascending order of time, where each event is of the form \((\text{time}, \text{type})\)
- \( x.t \) The time of event \( x \)
- \( x.\text{type} \) The type of event \( x \)
- \( rt \) Return time of the user who is attempting to make a reservation.
- \( lt \) The latest returning time of a reservation that was previously made in the return station
- \( O \) Occupancy at the return station (parked and waiting vehicles)
- \( C \) Capacity of the return station

Table 1: Look-ahead algorithm

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>While (x.t &lt; rt) &lt;br&gt; If (x.\text{type} = \text{rent}) and (O &gt; 0), set (O = O - 1).&lt;br&gt; If (x.\text{type} = \text{return}), set (O = O + 1).&lt;br&gt; Set (x) to the following event in the list</td>
</tr>
<tr>
<td>2.</td>
<td>Set (O = O + 1) (for current user)&lt;br&gt; If (O &gt; C), go to 5.</td>
</tr>
<tr>
<td>3.</td>
<td>If (lt &gt; rt)&lt;br&gt; While (x.t \leq lt)&lt;br&gt; If (x.\text{type} = \text{rent}) and (O &gt; 0), set (O = O - 1).&lt;br&gt; If (x.\text{type} = \text{return}), set (O = O + 1).&lt;br&gt; If (O &gt; C), go to 5.&lt;br&gt; Set (x) to the following event in the list</td>
</tr>
<tr>
<td>4.</td>
<td>Return “Reservation Allowed”</td>
</tr>
<tr>
<td>5.</td>
<td>Return “Reservation Denied”</td>
</tr>
</tbody>
</table>

Interestingly, in some cases under this utopian overbooking policy, users may arrive at their returning station and find that there is no vacant parking space to return the vehicle to. This may occur because in the look-ahead algorithm, it is assumed that all future demand for journeys outgoing from the destination station will reduce the occupancy of the station. But, some renters may decide to abandon the system due to their inability to make a reservation at their destination and therefore the occupancy of the station may be higher than anticipated. Here, the system is not penalized for shortages of parking spaces. Instead, the users are assumed to leave the system at their destination as if they were allowed to park their vehicles near the station. In other words, the station capacity is increased artificially until a parking space becomes
available. Therefore, all users who made reservations are guaranteed that they will be able to return their vehicles upon arrival at their destination.

We note that in a real stochastic setting, overbooking is likely to lead to more shortage events than in this utopian policy since the demand forecast is less accurate. Moreover, in reality, when shortage events occur, users are not allowed to leave their vehicles near the stations. Instead, they will have to waste more time in search for a vacant parking space or to wait for a parking space to become available. Therefore, under an actual overbooking policy, the total excess time is likely to be higher than in our utopian overbooking policy.

In Table 2 we summarize the answers to each of the three questions that appear in the user behavior model, which characterize the settings of the parking reservation policies described above.

<table>
<thead>
<tr>
<th>Parking reservation policy</th>
<th>Parking reservation requirement</th>
<th>Conditions to approve a parking reservation</th>
<th>Vacant parking space guaranteed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>For none of the users</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>CPR</td>
<td>For all users</td>
<td>A vacant parking space at the destination at the renting time</td>
<td>Yes</td>
</tr>
<tr>
<td>Partial: trip based</td>
<td>For users with trip time shorter than a given threshold</td>
<td>A vacant parking space at the destination at the renting time</td>
<td>Yes</td>
</tr>
<tr>
<td>Partial: station based</td>
<td>For users with a destination station in which the difference between the returning rate and the renting rate is higher than a given threshold</td>
<td>A vacant parking space at the destination at the renting time</td>
<td>Yes</td>
</tr>
<tr>
<td>Partial: time limited</td>
<td>For all users</td>
<td>A vacant parking space at the destination at the renting time</td>
<td>Only for users with trip time shorter than the time limit</td>
</tr>
<tr>
<td>Utopian overbooking</td>
<td>For all users</td>
<td>The system anticipates that there will be a vacant parking space at the destination at the returning time</td>
<td>Yes, in this hypothetical utopian setting the user is allowed to return a vehicle even when no parking space is available</td>
</tr>
</tbody>
</table>

4. Numerical study

In this section, we evaluate the proposed partial reservation policies with various threshold parameters and demand characteristics, using discrete event simulation of VSSs. The simulation is based on the user behavior model presented in Section 3.1. The results are compared to the lower bounds that were devised in Section 2. The numerical study is based on data from two real world bike sharing systems, Capital Bikeshare and Tel-O-Fun. In Section 4.1 we describe the two bike sharing systems and the trip data that was used to generate the input to our models. In Section 4.2 we present the results of the numerical experiments and discuss their implications.
4.1. Case studies

The Capital Bikeshare system was launched on September 2010. The system operates in Washington D.C., Arlington County and Alexandria, Virginia and Montgomery County, Maryland. The operating Company, Alta Bicycle Share, provides full trip history data, which can be downloaded from the following link: http://capitalbikeshare.com/trip-history-data. In this study we use trip data from the second quarter of 2013. In that period there were 232 operative stations with 3860 parking spaces and about 1750 bicycles in the system. The average number of daily trips on weekdays was about 7800.

In Figure 4 we present a map with the stations that were operative in that period. On the map we mark three clusters of stations: Arlington, Alexandria and Crystal City. As can be observed, in these clusters the stations are scattered densely while they are relatively distant from other stations in the system. Indeed, most of the trips that originated or ended in these clusters were within the cluster. In Alexandria, about 90% (resp., 88%) of the journeys that originated (resp., ended) in the cluster ended (resp., originated) in the cluster. In Crystal City, the figures are 77% and 74% and in Arlington, 70% and 76%, respectively. In the following section, we present results for the entire system and for each of the three clusters separately. While generating the data for each of the clusters, we neglected trips from/to other stations in the system. Though the resulting data does not fully reflect the occurrences in these stations, it allows us to analyze small systems in varying sizes which are “close to real”.

![Figure 4: Map of Capital Bikeshare stations (2nd quarter 2013)](image)

The second system that was studied is the Tel-O-Fun bike sharing system in Tel-Aviv. The system was launched on April 2011 and the trip data was collected during a period of two months in the beginning of 2012. At that time, the system consisted of 130 stations scattered in an area of about 50
square kilometers, a total of 2500 parking spaces and about 900 bicycles. During this period, the average number of daily trips (on weekdays) was about 4200.

The input for the simulation for both systems was generated as follows. We assume that the alternative mode of transportation is walking, which we believe is typically the case for bike sharing systems. The riding times and walking times were estimated using the Google Maps API. The capacities of the stations were retrieved from the systems’ websites. The arrival rates of renters during 30 minutes periods throughout the day were estimated by aggregating the weekday trips. Assuming Poisson demand processes, for each system 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 all examined policies (Common random numbers). In addition, for each demand realization, we generated the input for the PR-LB and PRP-LB models, namely the set of potential itineraries per realized journey.

Two methods for setting the initial inventory level of vehicles at the stations were used: (1) the actual initial station inventories on a randomly chosen day, after the operators executed repositioning activities; (2) the initial inventory levels prescribed by the method of Raviv and Kolka (2013). The purpose of using two different initial inventory levels is merely to check the sensitivity of our results and insights to these parameters. Clearly, we could use other methods known in the literature to determine the initial inventory, as discussed in the introduction.

4.2. Results
The discrete event simulation, together with the user behavior model and the preprocessing of the input for the mathematical models, were coded in MathWorks Matlab™. The PR-LB and PRP-LB models were solved using IBM ILOG CPLEX Optimization Studio 12.5.1. The codes and data are available upon request from the authors.

We begin by discussing the results of the lower bounds and the utopian overbooking policy. The results and discussion regarding the partial reservation policies are presented at the end of this section. In Table 3 we present results for the Capital Bikeshare and Tel-O-Fun systems. In the first and second columns, the name of the system and the number of stations in the system are given. In the third column we present the method according to which the initial inventory levels were set. In the fourth to seventh columns we present the average total excess time, over 50 realizations, for the NR policy, the CPR policy, the utopian overbooking policy and the PR-LB model. In the last column we present the average total ideal times, over 50 realizations. Recall that the ideal time is the total traveling time if all the journeys could be served ideally by a shared vehicle from the desired origin to the desired destination. The problem instances for the PRP-LB model could not be solved using the available computational resources and thus
this lower bound is not presented here. We revisit this model when analyzing the smaller sub-systems below.

### Table 3: Results for the two real-world systems

<table>
<thead>
<tr>
<th>System</th>
<th>Stations</th>
<th>Initial Inventory</th>
<th>NR</th>
<th>CPR</th>
<th>Over-booking</th>
<th>PR-LB</th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital Bikeshare</td>
<td>232</td>
<td>Actual Day</td>
<td>346.9</td>
<td>282.4</td>
<td>271.5</td>
<td>114.7</td>
<td>1,347.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>183.9</td>
<td>141.1</td>
<td>132.1</td>
<td>58.3</td>
<td></td>
</tr>
<tr>
<td>Tel-O-Fun</td>
<td>130</td>
<td>Actual Day</td>
<td>89.9</td>
<td>76.4</td>
<td>75.8</td>
<td>23.7</td>
<td>919.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>59.5</td>
<td>41.2</td>
<td>38.9</td>
<td>15.4</td>
<td></td>
</tr>
</tbody>
</table>

We observe from Table 3 that the lower bound on the total excess time provided by the PR-LB model is significantly tighter than the trivial lower bound obtained by assuming that all the journeys are materialized by their ideal itineraries, i.e., no excess time, as in Kaspi et al. (2014). For example, in Capital Bikeshare, about 40% of the gap between the CPR policy and the trivial lower bound (zero excess time) is explained by the PR-LB model. That is, at least 40% of the excess time under the CPR policy, cannot be reduced by any passive system regulation. Furthermore, recall that in the PR-LB model we assume that all demands for journeys are known in advance and that a central planner determines the itinerary of each user. As this setting is unrealistic, we can expect that the excess time under any real policy would be much higher. In other words, a major part of the remaining gap can be explained by these assumptions. Recall that each of the figures in Table 3 is an estimation of the excess time under a certain reservation policy based on an average of 50 demand realizations. The differences between the values in each row of the table were tested by a one-sided sign test and were found to be significant at $p-value < 0.000012$.

The results presented for the PR-LB model are based on the LP relaxation of the model. In addition, we solved the original MILP model for the smaller instances that are based on the Tel-O-fun data. In 97 out of these 100 instances, the value of the LP relaxation solution was identical to the one obtained by the MILP model, where the latter were obtained at a substantially longer processing times. In the remaining three instances, the lower bound obtained by the MILP model was slightly higher but the difference was negligible, less than 0.002%.

Using the initial inventories as prescribed by the method of Raviv and Kolka (2013), a significant reduction in excess time is obtained, as can be observed in Table 3 for all policies. Indeed, proper planning of static repositioning results with a major improvement in the service level. Nevertheless, the results for the CPR policy and the PR-LB model suggest that an additional substantial reduction in the
total excess time can be achieved by integrating repositioning activities with an efficient passive regulation.

As can be observed in Table 3, the utopian overbooking policy produced only slightly better results as compared to the CPR policy. This is quite surprising, when recalling the assumptions on which the utopian overbooking policy is based on. That is, even with full knowledge of the demand realizations and the use of overbooking, a significant improvement cannot be obtained. This implies that, realistic overbooking policies are not likely to be significantly (or at all) beneficial in terms of reducing the excess time in VSS. This unexpected finding can be explained by the fact that in VSS, a positive side-effect of parking space reservations is the diversion of the demand toward less congested stations. This in turn may have a positive effect on future users of the system who are less likely to face shortages of vehicles and parking spaces. Allowing overbooking reduces this positive side-effect. Given the fact that a good overbooking policy is much more complicated to implement than the CPR policy, and that it also introduces additional uncertainty and thus reduces the trust of the users in the system, we believe that this type of policy should not be practiced in VSSs.

In Table 4, we present some statistics on the dimension of the PR-LB instances that we solved and the solution times. We present the number of stations in each system, the average number of users (over the 50 demand realizations), the average number of itineraries per user, the number of variables in the linear programming model and the average solution time of both the LP relaxation and the MILP model, where the itinerary variables are defined as binary ones. Note that the MILP model could be solved in a reasonable time only for the smaller instances of the Tel-O-Fun network. The solution time of the PR-LB model is not of a particular interest in our study since such a model is not supposed to be solved very often. It is observable that the solution times are reasonable for most of the strategic and operational scenarios. That is, a similar formulation can be used for other purposes, where time consideration is more important.

<table>
<thead>
<tr>
<th>System</th>
<th>Stations</th>
<th>Number of users</th>
<th>Average number of itineraries per user</th>
<th>Initial Inventory</th>
<th>PR-LB</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Number of variables</td>
<td>Solution time LP (sec)</td>
<td>Solution time MILP (sec)</td>
<td></td>
</tr>
<tr>
<td>Capital Bikeshare</td>
<td>232</td>
<td>7,826.4</td>
<td>204.5</td>
<td>Actual Day Raviv &amp; Kolka</td>
<td>4,993,194</td>
<td>2,211.87</td>
<td>N/A</td>
</tr>
<tr>
<td>Tel-O-Fun</td>
<td>130</td>
<td>4,154.9</td>
<td>62.7</td>
<td>Actual Day Raviv &amp; Kolka</td>
<td>765,050</td>
<td>49.20</td>
<td>232.97</td>
</tr>
</tbody>
</table>

Solving the PRP-LB model that was presented in Section 2.3 is impractical for large real-world systems, due to the large number of binary variables. To obtain the insights provided by the solution of the PRP-LB model, we generated three small sized systems based on three clusters of stations in the
Capital Bikeshare system, namely, Alexandria with 8 stations, Crystal City with 15 stations and Arlington with 30 stations. In Table 5, we present the results for these systems. The table is supplemented with an additional column (the seventh), presenting the lower bound on the expected total excess time produced by the PRP-LB model. We observe from the table, that for the three small sized systems, the value obtained from the PR-LB model explained about 56-66% of the gap from the trivial (zero) lower bound. However, a larger portion of this gap, namely 67%-81% was explained by the PRP-LB value. This result further strengthens our belief that no parking reservation policy is likely to result in a dramatic (if at all) improvement over the CPR policy. We also note that for these systems, the excess time for the utopian overbooking policy is sometimes slightly higher than that of the CPR policy. Recall that each of the figures in Table 5 is an estimation of the excess time under a certain reservation policy based on an average of 50 demand realizations. The differences between the values in each row of the table were tested by a one-sided sign test and were found to be significant at $p - value < 10^{-7}$.

Table 5: Results for the three clusters in Capital Bikeshare

<table>
<thead>
<tr>
<th>System</th>
<th>Stations</th>
<th>Initial Inventory</th>
<th>Total Excess Time (hours/day)</th>
<th>Total Travel Time (hours/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td>CPR</td>
</tr>
<tr>
<td>Arlington</td>
<td>30</td>
<td>Actual Day</td>
<td>2.907</td>
<td>2.262</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>1.600</td>
<td>1.129</td>
</tr>
<tr>
<td>Crystal City</td>
<td>15</td>
<td>Actual Day</td>
<td>1.314</td>
<td>1.120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>0.656</td>
<td>0.564</td>
</tr>
<tr>
<td>Alexandria</td>
<td>8</td>
<td>Actual Day</td>
<td>0.589</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>0.225</td>
<td>0.184</td>
</tr>
</tbody>
</table>

In Table 6, we present statistics of the instances for the three clusters of Capital Bikeshare and of mathematical model that were used to create the lower bounds. The table is in the same format as Table 4. Interestingly, it is observed that the initial inventory has a significant effect on the solution time of PRP-LB. The optimized inventory levels obtained by the method of Raviv and Kolka (2013) results with models that can be solved much more quickly, although the dimension of the mathematical models are identical.

Next, we consider the partial reservation policies that were presented in Section 3.2., and examine whether they can improve the performance obtained by the CPR policy. In Figure 5, we present the simulation results for these policies. The figure contains six graphs, one for each combination of the two studied real-world systems and the three partial policies. In each graph, two curves are displayed, representing the percentage of excess time obtained using the two methods for setting the initial

Table 6: Statistics for the three clusters in Capital Bikeshare

<table>
<thead>
<tr>
<th>System</th>
<th>Stations</th>
<th>Number of users</th>
<th>Average number of itineraries per user</th>
<th>Initial Inventory</th>
<th>PR-LB</th>
<th>PRP-LB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Number of continuous variables</td>
<td>Solution time (sec)</td>
<td>Number of Auxiliary Binary variables</td>
</tr>
<tr>
<td>Arlington</td>
<td>30</td>
<td>255.6</td>
<td>42.2</td>
<td>Actual Day</td>
<td>36,223</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>0.72</td>
<td>374.55</td>
</tr>
<tr>
<td>Crystal City</td>
<td>15</td>
<td>128.5</td>
<td>17.0</td>
<td>Actual Day</td>
<td>8,052</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Alexandria</td>
<td>8</td>
<td>68.6</td>
<td>5.9</td>
<td>Actual Day</td>
<td>1,601</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Raviv &amp; Kolka</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: Partial reservation policies - percentage excess time under various settings
For each partial policy, we plot the percentage of excess time (relative to the ideal time) under various settings. For the trip based partial policy, we tested 31 time thresholds in an interval of three minutes. For the time limited partial policy we tested 31 time limits, again with three minutes intervals. For the station based partial policy, we tested 11 difference thresholds, of 0%, 10%,…, 100%; the thresholds were calculated for one hour time intervals during the day. In order to use the same scale on the horizontal axis for the two systems, we present the percentage of stations in which a parking space reservation is required, instead of presenting the difference thresholds.

Recall that extreme settings of such partial policies result with the complete policies (except for the lower extreme of the time limited partial policy). The figure demonstrates that as the time threshold increases, the same trend appears in all six graphs, i.e. when more reservations are required, the excess time decreases. The best performance is achieved when parking reservations are required from all the users, i.e., under the CPR policy.

These results demonstrate that using a simple rule to define a partial parking reservation policy is not likely to produce better results than the CPR policy. Another outcome of this analysis is that the more users are required to reserve parking spaces the better the performance of the system is. However, in situations where it is possible to require reservations only from some users, it is better to apply a partial reservation policy rather than not require reservations at all.

Recall that in the time limited partial reservation policy (Section 3.2.3), all users must make a reservation, but the reservation expires after a given time. If the time limit is set to zero, users would only be able to travel to stations that are not full at the renting time, but a parking space will not be guaranteed in the destination in any case. In Figure 3, the graphs of the time limited policy begin at lower points, compared with the other two partial policies. That is, as compared to the NR policy, it is observable that a significant improvement may be obtained simply by redirecting users to returning stations that are not full at the renting time. In fact, most of the improvement accomplished by the CPR policy may be attributed to the redirection of users to stations with vacant parking spaces.

Next, we examine whether the above insights are relevant to systems with other characteristics. In particular, we consider systems with the same geography and similar demand patterns, but with different congestion, i.e., offered load, and with different station capacities. For each of the systems (Tel-O-Fun and Capital Bikeshare) we generated new instances by multiplying the demand rates in all stations by several load multipliers, where 1 represents the original systems. Fifty demand realizations were generated based on each of these load multipliers.

In Figure 6, we present the performance of the trip based partial reservation policy with various time thresholds and load multipliers. It is observable that in both systems and under various congestion levels
the excess time is reduced as more reservations are required. This implies that the effect observed under the original demand load is, qualitatively, not affected by the congestion level. However, as the congestion increases the benefit obtained from the reservation is increasing. This can be attributed to the fact that in a more congested systems, shortage events are more likely to occur. For the sake of brevity, we present the results only for the trip base partial reservation policy. Very similar trends were observed under the station based and the time limited partial reservation policies.

Further on, we examined the effect of the capacities of the stations on the performance of the system under the same 50 demand realizations. To this end, we have conducted the following test: the capacities of all the stations in the system were decreased/increased by 25% (and rounded to the closest integer). In Figure 7 we present the results for the trip based partial reservation policy. Similar to previous results, the excess time reduces as the time threshold increases. That is, the same trends are observed regardless of the capacity of the stations. For a given demand rate, as the capacity of the stations are increased, the number of parking space shortages are reduced. It is observable that the excess time under the NR policy, the various settings of the partial reservation policies and the CPR almost converge to the same value as the station capacities are increased. As may be expected, the benefit of implementing parking reservations increases when the parking spaces are scarcer. Again, similar trends are observed under the station based and the time limited partial policies.

Figure 6: Trip based reservation policy - percentage excess time under various load factors with initial inventory obtained from an actual day
5. Concluding remarks

The main message of this study is reinforcing the effectiveness of parking reservations in VSS as a method to improve the service provided to its users. It is shown that the simplest possible parking reservation policy (namely, CPR) appears to be the most effective one in terms of reducing the total excess time. This was observed through empirical tests under numerous settings that are based on the geography and demand trends of two real-world systems, diverse offered loads, station capacities and initial inventories. Our case studies, presented in Section 4, are based on data received from bike sharing systems. However, we believe that the concept of parking space reservation, and other passive regulations, is even more relevant to car sharing systems where the cost of active regulation (i.e., relocation of vehicles) is prohibitive.

Using a lower bound calculated by the PR-LB model, we have demonstrated that, in our case studies, a significant share of the excess time that could be theoretically saved by any passive system regulation, is already saved by the CPR policy. Our extended PRP-LB model shows that other parking reservation policies are not likely to be able to save substantially (if at all) more excess time.

We also studied several partial reservation policies and demonstrated that while these policies are slightly inferior to the CPR, they may also be a good alternative to the basic NR policy in cases where CPR cannot be implemented for some reasons. Finally, we precluded reservation policies that are based on overbooking as a parking reservation approach that is likely to outperform the CPR policy. This was achieved by showing that even under a utopian scenario in which the system looks ahead into future demand, such policies cannot significantly reduce the excess time obtained by CPR.
The PR-LB based lower bound, introduced in Section 2, can be used to evaluate the effectiveness of various other VSS related policies. This model reflects the fact that each journey may be assigned to one of several itineraries. This adds a lot of flexibility to VSSs and affects its dynamics in a way that should not be ignored by a strategic planner. Though we have focused on reducing the excess time of users, our model can be extended to accommodate other objectives of the users. That is, each potential itinerary can be assigned with a measure that reflects a combination of several objectives. We also suggest using our model in the future to incorporate considerations of the operator. For example, if a car sharing operator faces profit losses due to some possible itinerary choices of the users, these values can be weighted and added to the excess time. It would be interesting to examine the effect of parking reservation policies on the obtained profit under various pricing schemes.

Acknowledgments
- This research was supported by the Israel Science Foundation (ISF) grant no. 1109/11.
- The authors wish to acknowledge the valuable comments of two anonymous referees on a previous version of this article.
References


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


