# Optimal Deployment of Stations for a Car Sharing System with Stochastic Demands: a Queueing Theoretical Perspective

Elisabetta Biondi, Chiara Boldrini, and Raffaele Bruno
IIT-CNR
Via G. Moruzzi 1, 56124, Pisa, ITALY
{e.biondi,c.boldrini,r.bruno}@iit.cnr.it

*Abstract*— **Car sharing holds a promise of reducing traffic congestion and pollution in cities as well as of boosting the use of public transport when used as a last-mile solution in a multi-modal transportation scenario. Despite this huge potential, several problems related to the deployment and operations of car sharing systems have yet to be fully addressed. In this work, we focus on station-based car sharing and we define an optimization problem for the deployment of its stations. The goal of this problem is to find the minimum cost deployment (in terms of number of stations and their capacity) that can guarantee a pre-defined level of service to the customers (in terms of probability of finding an available car/parking space). This problem combines insights from queueing theory (used to model the stochastic demand for cars/parking spaces at the stations) with a variant of the classical set covering problem. For its evaluation, we use a trace of more than 100,000 pickup and drop-off events at a free-floating car sharing service in The Netherlands, which are used to model the input demand of the car sharing system. Our results show that the proposed solution is able to strike the right balance between cost minimisation and quality of service, outperforming three alternative schemes used as benchmarks.**

## I. Introduction

Car sharing systems are innovative mobility services that are becoming increasingly popular in urban and sub-urban areas and have the potential to solve real-world problems of urban transports [1]. The principle of a car sharing system is that customers can rent for limited period of times a car from a fleet of shared vehicles operated by a company or a public organisation. Nowadays, the most widespread type of car sharing service is one-way free-floating car sharing [2]. In this case, the return of the rented vehicle is possible at any parking spot within the operational area of the car sharing service [3]. The main advantage of this approach is the great flexibility for car sharing members, who can pick up the nearest car and return it *anywhere* within a given area. Examples of free-floating car sharing services are Car2go, DriveNow, and Enjoy.

While the vast majority of car sharing fleets around the world feature gasoline-powered cars, there is a general shift, both in car sharing and within the automotive sector in general, towards electric cars. However, the combination of a free-floating car sharing service with electric vehicles is not straightforward and it typically relies on many and well-located charging stations. Unfortunately, it is rarely the case that such an efficient and powerful infrastructure is available, hence car sharing companies may run into trouble. For example, in March 2016 Car2go decided to replace its electric fleet in San Diego, US, with gas-powered cars due to the lack of charging stations in the city [4].

An alternative form of one-way car sharing is station-based car-sharing (e.g., Autolib in France) in which dedicated stations are deployed by the car sharing operator and users are required to pick up and drop off shared vehicles only at one of the available stations. Clearly this limits the freedom of movement for the users. However, these stations can be easily equipped with the necessary infrastructure for recharging electric vehicles [5], which makes station-based car sharing systems a good candidate for electric car sharing services [6]. Another advantage of station-based car sharing is that it provides higher reliability and predictability of car locations and parking, the latter being a particularly attractive option for city centers, where finding a parking space can be a nightmare for drivers of private cars.

While the customers' desire for flexibility can be addressed by smartly placing stations within the operational area, the huge drawback of station-based car sharing is that it requires a significant capital investment to build the necessary station infrastructure, making it less financially attractive than a free-floating service. Thus, when deploying a station-based car sharing system it is crucial to strike the right balance between the costs for the operator and the quality of service provided to the customers. In fact, deploying fewer stations allows the operator to save money, but if the service does not provide a good experience to the customer, the money saved on the infrastructure is rapidly lost on missed rentals and general customer dissatisfaction. Vice versa, an effective deployment may initially cost more but it may also increase ridership and customer satisfaction in the end.

Existing planning frameworks typically rely on time-space models, which are models that assume a deterministic knowledge of the demand of vehicles at each time interval of the control period [7], [8], [9]. However, in real-world scenarios the demand process of customers is stochastic and it exhibits seasonal patterns [10] as well as weather-dependent ones. Furthermore, in car sharing there is a complex interdependence between travel demands and availability of vehicles

*This work was partially funded by the ESPRIT project. This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 653395.

and parking at each station. In other words, the way users make rental requests affects in a complex manner the future availabilities of cars and parking spaces. Hence, modelling approaches that are able to take into account a stochastic demand ([11], [12], [13]) look more suited to address this complexity. However, to best of our knowledge, the literature currently lacks a model that can consider simultaneously decisions related to the location and size of stations while taking into account that customers' demands, hence station dynamics, are stochastic and not deterministic.

The contribution of this paper is twofold. First, we develop an *optimization model* for the problem of determining the number and location of the stations of a one-way car sharing systems in the presence of stochastic demand and congestion. We use a set-covering model to formulate the problem of locating stations and sizing their parking capacity with the aim of minimising the deployment costs while providing a configurable service levels. Following the approach adopted in [14], we model each candidate site as a *queueing-based* system using standard assumptions for the arrival and service processes. Then, we develop a computationally efficient, near-optimal solution based on a greedy search method. As a second contribution, we apply our planning methodology to a real case study using a dataset containing the pickup and drop-off events of a car sharing operator in The Netherlands with a fleet of almost 400 shared vehicles. Our results show that the proposed optimization model is able to strike the right balance between minimising the infrastructure costs of the car sharing operator and minimally affecting the service provided to the customers (in terms of availability of shared cars and parking spaces at stations).

The remainder of the paper is structured as follows. Section II provides an overview of previous related work. Section III discusses the preliminaries to our optimization problem, i.e., how the demand within the operational area can be represented and how this demand relates to the capacity of stations. In Section IV we present the optimization problem for planning the car sharing infrastructure deployment in terms of location of stations and capacity of stations. Then, in Section V, we evaluate the proposed optimization strategy using as input demand the pickup and drop-off events at a real-life car sharing system. Finally, in Section VI we draw our conclusions and present directions for future research.

## II. Related Work

Facility location is an optimization problem extensively studied in the field of logistics and transportation planning. Facility location models are often cast as set covering problems, and the interested reader is referred to [15] for a comprehensive survey on the topic. Primarily interesting for our work are the solutions proposed for the optimal deployment of stations for bike-sharing and car-sharing systems. Several works in the literature ([16], [8], [9]) consider a demand process that is static and deterministic. This limits both the realism (the actual demand is definitely not deterministic) and the power of the model. In fact, the practical advantage of a modelling tool lies in its predictive power, i.e., in its capacity to capture the salient features of a complex dynamic system and, possibly, to represent it in a compact and synthetic way. Works that move towards this direction are [11], [12], [13]. The mathematical tools that are used to achieve this goal are queueing theory and fluid models. In [11], a closed-queueing network is used to model a general rental system, and the closed-queueing framework is exploited for optimising the number of shared vehicles in order to maximise their availability. The work in [12] focuses on the rebalancing problem in a sharing system assuming that vehicles are autonomous (i.e., that they are driverless, hence they can autonomously reach under-served locations if needed). The sharing system is modelled using a fluidic approach, where moving vehicles and arriving/departing customers are seen as *flows* within the network. The goal is to find how large should be the flow of vehicles that autonomously move to stations that are under-served in order to rebalance the system. The same problem of rebalancing vehicles is investigated in [13], this time using queueing theory. The authors argue that the latter model provides the foundation for the former.

Our contribution falls within the same category of the works discussed above, in that we exploit queueing theory to model the dynamics of the car sharing system. However, we depart from the state of the art in several ways. Differently from the related literature, we combine the knowledge on the system dynamics acquired through queueing theory with a facility location problem that aims at selecting the location of the stations as well as their optimal capacity. The optimal capacity value is derived by treating the station as a queue, then estimating the availability of vehicles at the stations as if they were jobs waiting to be served at a queue. Differently from [11], [12], [13], we thus address the problem of stations with finite capacity $K$, which is what is observed in real car sharing systems. Finally, again differently from the related literature, we evaluate the proposed algorithm using a trace of pickup and drop-off events at a real car sharing system.

## III. Preliminaries

In this study, we assume that the study area is partitioned into a set $\mathcal{N}$ of non-overlapping square cells. Each cell may contain a set of arrival and departure events of shared cars, or it can be empty. Each of these cells is a potential candidate location for the stations of the car sharing system. Without loss of generality, we assume that stations are placed at the center of cells.

We assume that users are willing to travel at most a certain distance in order to reach the nearest station[1]. We control this willingness using parameter $R$. When $R = 0$ users are very conservative and want to find a stations exactly inside the cell (let us assume its location to be $(x, y)$ in the grid) where they want to pick/drop a shared car. With $R = 1$, users are willing to explore at least the cells neighbouring $(x\pm1, y\pm1)$ the one where they want to pick/drop a shared car. With $R = 2$, they can additionally explore the neighbours of the

[1]For example, in the London bike sharing system a rule of thumb of $300m$ is used when choosing the distance between stations (http://oobrien.com/category/bike-share/).

neighbours $(x \pm 2, y \pm 2)$ of their interested cell, and so on and so forth. Since users will not explore beyond $R$, their requests for cars/parking spaces will not be satisfied unless they find a station within $R$. Thus, the *coverage radius* of each station can be at most equal to $R$, since each station can serve requests that happen at most $R$ cells away.

The key idea of the proposed optimization model is that, given the sets of arrival and departure events, the evolution of the cars dropped off and picked up at the car sharing stations can be modelled and, thus, predicted using a *queue-based model* [17]. In fact, from the sets of arrival and departure events at each cell, a stochastic travel demand at the stations can be computed. More precisely, assuming a coverage radius $R$, let $\lambda_i$ be the average rate of trips ending at station $i$ (i.e., the number of shared vehicles dropped off per unit time), and $\mu_i$ the average rate of trips that depart from station $i$ (i.e., the number of vehicle pickups per unit time). For the sake of model tractability, similarly to the related literature [12], [11], [13], we assume that the arrival process is Poisson, and that the time between two consecutive pickups of parked cars at a station is Exponential. This is clearly an approximation, since pickups and drop-offs exhibit well-known daily and weekly trends in reality [18]. However, in Section V we show that the model based on these assumptions still maintains an impressive predictive power.

Under the above assumptions, if the stations had infinite parking spaces, they could be modelled as M/M/1 queues [17]. However, real stations have a finite capacity, corresponding to the number of parking spaces that they can fit in. In order to take into account this aspect, we model stations as M/M/1/K queues, i.e., as queues with Poisson arrivals, Exponential service times, and capacity (parking spaces in our case) equal to $K$. This formulation is very convenient, since it allows us to describe with closed-form expressions the dynamics of cars parked at stations. In fact, the probability that, at steady-state, there are exactly $0, 1, \dots, K$ available shared cars at a station is equal to the steady-state probability of the M/M/1/K queue describing the station. For the sake of clarity, we remind that the steady-state probability $\pi_n$ that there are $n$ cars parked in the system (or jobs, using the queueing analogy) can be written as follows:

$$\pi_n = \frac{(1-\rho)\rho^n}{1-\rho^{K+1}}, \tag{1}$$

with $n \in \{0, 1, \dots, K\}$ and $\rho = \frac{\lambda}{\mu} \neq 1$ (with $\rho = 1$, $\pi_n = \frac{1}{K+1}, \forall n$ [17]).

From the car sharing operator's point of view there are two key aspects to consider when evaluating the service provided to the customer: the probability that customers find an available car where they request one, and the probability that customers find an available parking space at the station where their trip ends. These two probabilities correspond, using the queue analogy, to the probability that the queue is not empty $(1-\pi_0)$ and to the probability that the queue is not full $(1-\pi_K)$, respectively. Thus, when planning the capacity of a station, the goal of the car sharing operator is to choose the minimum value of $K$ such that these two probabilities are above some target performance levels, which we denote as $p_{car}$ and $p_{parking}$, respectively. Exploiting Equation 1 above, the following important result can be derived.

*Lemma 1 (Optimal capacity):* Assuming a coverage radius $R$, in order for the availability of cars and parking spaces to be above the desired thresholds $p_{car}, p_{parking} \in (0, 1)$, the capacity $K_i$ of a tagged station $i$ should be equal to the minimum solution to the following system of inequalities:

$$\begin{cases} K_i & > & \log_{\rho_i}\left(\frac{\rho_i - p_{car}}{1-p_{car}}\right) - 1 \\ K_i & > & \log_{\rho_i}\left(\frac{1-p_{parking}}{1-p_{parking}\cdot\rho_i}\right), \end{cases} \tag{2}$$

where $\rho_i$ denotes the ratio $\frac{\lambda_i}{\mu_i}$ of the arrival and departure rates at station $i$ when the coverage is $R$.

*Proof:* The two inequalities are obtained by substituting to $1-\pi_0$ and $1-\pi_K$ the expressions for $\pi_0$ and $\pi_K$ derived from Equation 1, then inverting to get $K_i$. ■

Please note that the optimal capacity value $K_i$ is only dependent on the stochastic demand at the station ($\rho_i = \frac{\lambda_i}{\mu_i}$, which in turn depends on the coverage $R$) and on the target availability of cars/parking spaces $p_{car}$ and $p_{parking}$. Lemma 1 also tells us that the desired target performance cannot be always achieved. In fact, since the argument of the logarithm must be positive, we have that (dropping subscript $i$ for the sake of clarity) $p_{car} < \rho$ and, at the same time, $p_{parking}\rho < 1$. It follows that, when $\rho < 1$, every $p_{parking}$ is attainable for parking space availability but only $p_{car} < \rho$ is possible for the availability of cars. Vice versa, when $\rho > 1$, any $p_{car}$ is possible but only $p_{parking} < \frac{1}{\rho}$ is attainable for parking spaces. Intuitively, this is due to the fact that, when $\rho < 1$, shared cars are picked up quickly at stations once available (in fact, $\rho < 1$ implies $\mu > \lambda$, i.e., a rate of departure greater than the rate of arrival). When cars are picked up quickly, there is not much the operator can do about the availability of cars when planning the station infrastructure, because simply increasing $K$ has no effect on the number of available cars. In this case, it is strategic that the car sharing operator performs vehicle redistribution (like in [12]), which would alter the value of $\lambda$, hence rebalancing the ratio between pickups and drop-offs. The opposite holds true when $\rho > 1$: the station tends to fill up quickly, hence customers will most certainly find an available car but not necessarily a parking space.

## IV. Optimising the station infrastructure

After having laid out the building blocks of our model in the previous section, we now focus on a car sharing operator that has to decide how to deploy its station infrastructure. Specifically, the operator has to decide i) where to place the stations ii) which capacity to assign to each station. This problem is an instance of a set cover problem [19]: we have to identify the infrastructure that requires the minimal capital investment and achieves the desired service level in terms of coverage and availability of cars/parking spaces.

In order to approach this set cover problem, we first have to identify candidate stations. In principle, stations could be placed at any cell of the grid. However, as we discussed in Section III, a station can only serve those pickup/drop-offs happening within $R$, i.e., within the maximum distance that customers are willing to tolerate. Hence, candidate stations should only be located in cells for which there are pickup/drop-offs events within $R$. In the following we denote the set of candidate stations as $\mathcal{S}$.

The control variables of our optimization problem are the capacities $K_i$, with $i \in \mathcal{S}$. Variable $K_i$ is equal to the capacity of candidate station $i$ if station $i$ is actually selected, zero otherwise. When a station is selected for deployment its capacity $K_i$ should satisfy the constraint in Lemma 1. In real life, the car sharing operator could be forced to keep the capacity of the stations below a certain value (e.g., because the municipalities may not be willing to allocate too many parking spaces to a single car sharing service). We can incorporate this constraint into our model by forcing the capacity of the station to be smaller than a certain value $K_{max}$, which could be different for different stations (depending on where they are positioned, for example). Here, without loss of generality, we can assume $K_{max}$ is equal for all stations.

Deploying stations is a big cost for the operator. A fraction of this cost is independent of the size of the station. We denote this cost as $C_s$ and we refer to it as *fixed cost*. It corresponds to, e.g., the sum of costs for building the payment booth at the car sharing station, maintenance, insurance, etc. A fraction of the cost is instead dependent on the size of the station. For example, we can imagine that larger stations will cost more to the car sharing operator in terms of cost for renting the parking space or for deploying a charging spot at the parking bay. We denote these latter costs with $C_k$ and we call them *variable costs*.

Relying on the above discussion, the optimization problem can now be cast into an integer programming problem (IP) as in Problem 1. The objective function (Equation 3) states that we want to minimise both fixed costs and the variable costs (those proportional to the capacity of the stations). We denote with $\mathbb{1}_{K_i>0}$ the indicator function, which is equal to 1 when condition $K_i > 0$ holds, equal to 0 otherwise. Then, the first constraint (Equation 4) forces that each cell in which drop-off/pickup events happen is served by at least one station within the coverage radius $R$. We have denoted with $\mathcal{R}_j$ the set of potential stations that are within $R$ from cell $j$. We also denote with $\mathcal{N}'$ the subset of $\mathcal{N}$ containing only those cells in which pickup/drop-off events take place. The second and third constraints (Equations 5-6) are related to the target availability that the car sharing operator wants to achieve. The only difference with what we have discussed in Section III is that, in order to maintain the problem always feasible, we redefine $p_{car}^{(i)} = \min\{p_{park}, \rho_i - \epsilon\}$ and $p_{parking}^{(i)} = \{p_{parking}, \frac{1}{\rho} - \epsilon\}$, with $\epsilon$ arbitrarily smaller. This is necessary to keep the argument of the logarithms in Equation 2 positive and it reflects what we discussed in Section III: assuming, e.g., $\rho < 1$, the only $p_{car}$ really attainable by station $i$ corresponds to the smallest value between the desired $p_{car}$ and the $\rho_i$ of station $i$. The fourth

**Problem 1** Infrastructure optimization

$$\underset{K_i}{\text{minimize}} \sum_{i \in \mathcal{S}} C_s \mathbb{1}_{K_i>0} + C_k K_i \tag{3}$$

$$\text{subject to} \sum_{i \in \mathcal{R}_j} \mathbb{1}_{K_i>0} \geq 1, \forall j \in \mathcal{N}' \tag{4}$$

$$K_i > \log_{\rho_i}\left(\frac{\rho_i - p_{car}^{(i)}}{1 - p_{car}^{(i)}}\right) - 1 \tag{5}$$

$$K_i > \log_{\rho_i}\left(\frac{1 - p_{parking}^{(i)}}{1 - p_{parking}^{(i)} \cdot \rho_i}\right) \tag{6}$$

$$K_i \leq K_{max} \tag{7}$$

$$K_i \in \mathbb{N} \tag{8}$$

constraint (Equation 7) forces the capacity to be smaller than or equal to the maximum allowed capacity. Finally, the fifth constraint (Equation 8) states that $K_i$ should be a natural number.

### A. Greedy approximation method

Set covers problems like the one in Problem 1 are known to be NP-hard. Luckily, greedy algorithms typically work well in these cases and find a set cover not too far from the optimal cover [19]. Hence, we define a greedy strategy for solving Problem 1. Its pseudocode is provided in Algorithm 1. This greedy strategy works by picking, at each step, the candidate station that has the lowest cost per event covered (line 10). The capacity of this station is equal to the value of $K_i$ that satisfies Lemma 1 (line 7). If this value of $K_i$ is greater than $K_{max}$, we reduce the coverage radius of the station and we repeat the search (lines 11-16). Otherwise, the candidate station is selected for deployment and we store the mapping between the selected station and its optimal capacity in $\mathcal{K}_{opt}$ (line 18). Then, we remove the station from the candidate stations set (line 19) and we update the set of events to be covered (line 20). The latter update alters the demand at the remaining candidate stations, hence we also need to update the $\rho_j^{(R_j)}$ and $K_j$ at those candidate stations that are affected by the change in the set of events to be covered (lines 21-24). When all events have been covered, the solution to the optimization problem ($\mathcal{K}_{opt}$) is returned (line 27).

## V. Evaluation

In this section we evaluate the proposed optimization strategy against three benchmark algorithms. The demand for pickup and drop-off of vehicles is obtained from the historical data of a real-life free-floating car sharing system operating in The Netherlands.

*The dataset.* It contains $100,650$ events corresponding to the pickup/drop-off times and GPS locations for 349 vehicles of a real-life free-floating car sharing fleet. The dataset covers the period from May 17, 2015 to July 1, 2015 and it has a granularity of 1 minute. We have divided the service area of the car sharing operator in $100m \times 100m$ cells, obtaining a grid with 162 rows, 128 columns, and 20736 cells. In order to

**Algorithm 1** Greedy algorithm

$\triangleright$ $\mathcal{E}$ set of all pickup/drop-off events
$\triangleright$ $\mathcal{N}$ set of all cells in the scenario

$R_i = R$ $\triangleright$ $\forall i \in \mathcal{N}$
$\mathcal{C}_i^{(R_i)} = \{e \in \mathcal{E} : dist(e, i) < R_i\}$ $\triangleright$ $\forall i \in \mathcal{N}$
$\mathcal{S} = \{i \in \mathcal{N} : \mathcal{C}_i^{(R_i)} \cap \mathcal{E} \neq \emptyset\}$
$\mathcal{K}_{opt} = \emptyset$
**for all** stations $i \in \mathcal{S}$ **do**
  compute $\rho_i^{(R_i)}$ $\triangleright$ based on events in $\mathcal{C}_i^{(R_i)} \cap \mathcal{E}$
  find $K_i$ for $\rho_i^{(R_i)}$ according to Lemma 1
**end for**
**while** $\mathcal{E} \neq \emptyset$ **do**
  find station $i \in \mathcal{S}$ such that $\frac{C_s + C_k K_i}{\mathcal{C}_i^{(R_i)} \cap \mathcal{E}}$ is maximised
  **if** $K_i > K_{max}$ **then**
    $R_i = R_i - 1$
    update $\mathcal{C}_i^{(R_i)}$
    **if** $\mathcal{C}_i^{(R_i)} = \emptyset$ **then** $\triangleright$ candidate station removed
      $\mathcal{S} = \mathcal{S} - \{i\}$
    **end if**
  **else** $\triangleright$ candidate station selected
    $\mathcal{K}_{opt} = \mathcal{K}_{opt} \cup \{i \mapsto K_i\}$
    $\mathcal{S} = \mathcal{S} - \{i\}$
    $\mathcal{E} = \mathcal{E} - \mathcal{C}_i^{(R_i)}$
    **for all** stations $j \in \mathcal{S} : \mathcal{C}_j^{(R_j)} \cap \mathcal{E}$ has changed **do**
      update $\rho_j^{(R_j)}$
      update $K_j$ for $\rho_j^{(R_j)}$ according to Lemma 1
    **end for**
  **end if**
**end while**
**return** $\mathcal{K}_{opt}$

run our algorithm, we need to compute the demand $(\lambda, \mu, \rho)$ at each candidate station, based on the events that happen within its coverage radius $R$. To this aim, we follow the simple method discussed in [20]: for each potential station, we count the number $n_a$ of arrivals, the number $n_d$ of departures, and we measure the busy time $T_{busy}$ (i.e., the time the station is not empty) during the observation period $T$. Then, $\lambda$ can be obtained as $\frac{n_a}{T}$, $\mu$ as $\frac{n_d}{T_{busy}}$, and $\rho$ simply as $\frac{\lambda}{\mu}$.

*The benchmarks.* We compare our planning algorithm (which is hereafter denoted as OPT) to three benchmarks. First, we consider a simple random strategy that places stations uniformly at random and places as many of them as to guarantee that any event has a station within distance $R$. We denote this strategy as RAND1. The capacity assigned to stations is fixed and equal to 6, i.e, equal to the average capacity observed in the real station-based car sharing system studied in [18]. The second random scheme we consider (denoted as RAND2) places as many stations as OPT but their location is selected uniformly at random among all cells where pickup/drop-off events take place. RAND2 has an economic advantage with respect to RAND1 in that it uses an optimal number of stations (the number computed by OPT). We will see later on that this is not enough for meeting the required quality of service: where stations are located and which capacity is assigned to them is as important as their number. The third benchmark (denoted as KMEANS) is a smarter algorithm that groups events based on $k$-means clustering and then assigns a station to each cluster center. The number of clusters, i.e., the $k$ of the $k$-means, is set equal to the number of stations placed by OPT. KMEANS is oblivious to the station coverage but groups events such that they can be served effectively with the desired number of stations $k$. Please note that $k$-means has been also used in [13] for station planning in their evaluation.

For KMEANS and RAND2 there is no obvious way to select the capacity of stations. In order to make them compete on equal grounds with OPT, we use a simple strategy whereby we force them to have the same overall system capacity of OPT ($\sum_i K_i^{OPT}$). This is obtained by solving the following system of equations:

$$\begin{cases} n_{floor}\lfloor \bar{K}^{OPT} \rfloor + n_{ceil}\lceil \bar{K}^{OPT} \rceil = \sum_i K_i^{OPT} \\ n_{floor} + n_{ceil} = |\mathcal{S}^{OPT}| \end{cases}, \quad (9)$$

where $\bar{K}^{OPT}$ denotes the average capacity assigned by OPT, and $n_{floor}$ and $n_{ceil}$ the number of the largest previous and the smallest following integer of $\bar{K}^{OPT}$ that should be assigned as capacity to stations.

*Model parameters.* For our OPT algorithm we need to select the target availability probability for both shared cars and parking spaces ($p_{car}$ and $p_{parking}$). In this evaluation we set them to 0.8. We also have to set the maximum distance that customers are willing to explore for finding a car/parking space, which determines the coverage radius of stations. We test the range $R \in \{1, 2, 3, 4, 5\}$, which corresponds to a maximum distances between a cell and its reference station that goes from $280m$ ($R = 1$) to $850m$ ($R = 5$). Please note that we do not test case $R = 0$ since this would imply that each cell should be equipped with a station: in fact, users would not be willing to look for one outside the cell where they want to drop/pick up the vehicle. In order to better evaluate the ability of the proposed scheme to fine tune station capacity, we will ignore the constraint $K_i < K_{max}$ in this evaluation. Finally, as for the cost values $C_s$ and $C_k$, they will be discussed in Section V-A.1.

### A. Results

We consider separately two aspects: costs and performance in terms of service provided to the customers.

*1) Costs:* We first discuss the performance of OPT in terms of costs for the car sharing operator to deploy the infrastructure. In order to provide a meaningful comparison, we contrast the costs for OPT against the costs for RAND1, since RAND1 is the only benchmark for which the number of stations deployed and the capacity of stations are completely independent of OPT. We test three configurations for OPT. In the first configuration, we set $C_s = 0$€ and $C_k = 1K$€. This corresponds to a case where there is no fixed cost for building a station but every parking space costs $1K$€. With this configuration, OPT will aim at minimising the overall capacity in the system (i.e., the sum of all capacities assigned

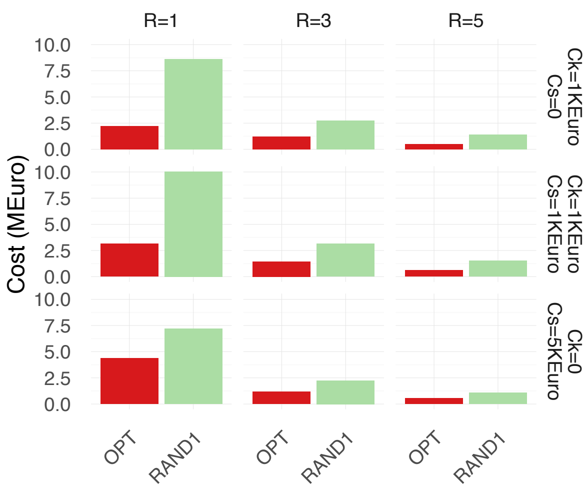


Fig. 1. Costs for deploying the infrastructure.

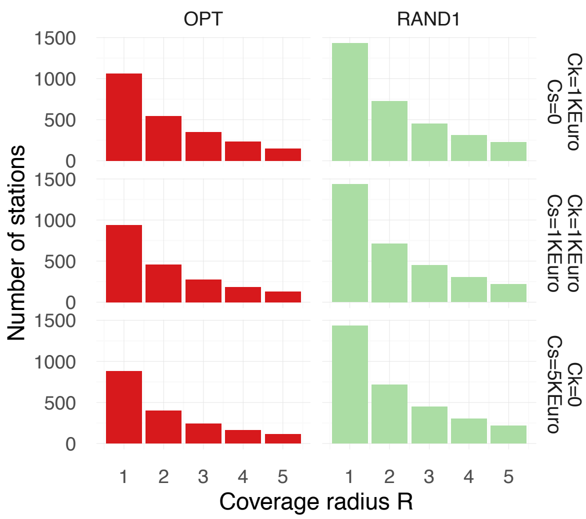


Fig. 2. Number of stations deployed.

to the deployed stations) but not the number of stations. Vice versa, with configuration $C_s = 5K€$ and $C_k = 0€$, OPT will try to minime the number of stations, ignoring the capacity that it is assigned to them. Finally, in the third configuration, where $C_s = 1K€$ and $C_k = 1K€$, OPT tries to strike a balance between the number of stations deployed and their size.

We observe in Figure 1 that OPT always outperforms RAND1, regardless of the specific cost configuration. This is due to a combination of factors. First, since it prioritises the coverage of areas where more events are taking place, OPT is able to always use fewer stations than RAND1 to cover the same set of events (Figure 2). Second, OPT is able to fine tune the capacity of the station to the demand at the station, hence it can use higher or lower values of $K_i$ depending on the situation. On the contrary, RAND1, being oblivious to the demand as well as to the costs, always uses the same capacity value, as shown in Figure 3

It is interesting to note the key role played by the coverage radius $R$ in infrastructure planning. As Figure 2 shows, the number of stations that we need to deploy (on the y-axis) decreases rapidly (regardless of the planning strategy)

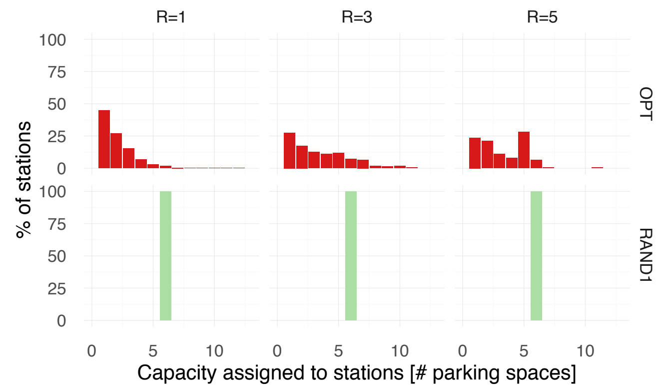


Fig. 3. Histogram of capacity distribution

as we increase $R$, which corresponds to the willingness of customers to explore more their surroundings. Unfortunately, $R$ is not something that the car sharing operator can control directly. In a real life scenario, the most realistic values of coverage are $1, 2, 3$, which correspond to distances from roughly 300m to 600m.

2) *Quality of service provided to customers:* The goal of our optimised planning is to provide probabilistic guarantees to the car sharing members in terms of availability of cars and parking spaces. With the next set of plots we will investigate whether this objective is achieved. In order to provide a fair comparison, we use as benchmarks KMEANS and RAND2, which, by definition, use the same number of stations and the same total capacity of OPT for a given scenario. Thus, the different performance will be due only to a different distribution of stations and capacities. Due to lack of space, we only show the results for $C_s = 0€$ and $C_k = 1K€$, but the trend remains similar with the other cost configurations.

In Figure 4 we plot the probability that a parking space is found. The black horizontal line corresponds to the target probability chosen for the availability of parking spaces, equal to $p_{parking} = 0.8$. Each point in the plot corresponds to one station, and it is colored in red if the availability at the station is below the desired threshold, in green otherwise. The vast majority of points with OPT fall above the $0.8$ threshold, with an average of $96\%$ of points in the realistic range $R \in [1, 3]$. In this same range, KMEANS features an average of $92\%$ of stations correctly predicted, while RAND2 shows the worst performance with an average of $85\%$. It is also interesting to notice that with OPT, for which capacity values $K_i$ depend on the $\rho$ at the stations, missed predictions always affect extreme values of $\rho$, either close to zero or close to one. This is probably where the real system departs more significantly from the queueing theoretical model.

Next, we look at car availability. As shown in Figure 5, all $\rho$ values at stations are in the range $(0, 1)$. Recalling the discussion in Section III, this means that only $p_{car} < \rho$ is attainable when optimising $K_i$ for car availability. Hence, we set $p_{car} = \rho - \epsilon$, with $\epsilon = 0.01$. Correct predictions will thus lie close to the bisector. As Figure 5 shows, all algorithms perform quite well. However, OPT clearly outperforms the others for small coverage values.

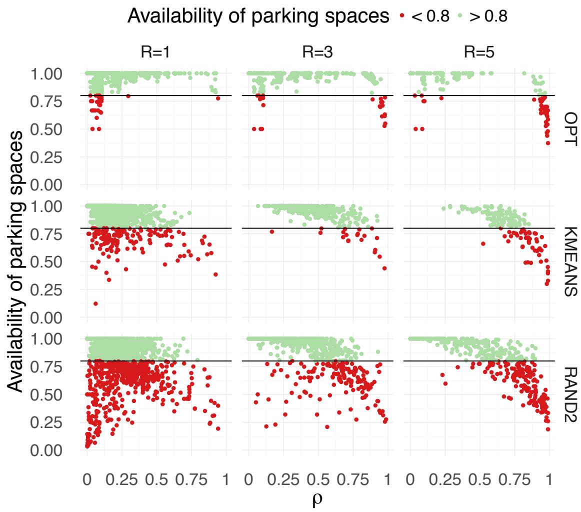


Fig. 4. Observed probability that drop-offs are missed, as a function of $\rho$.

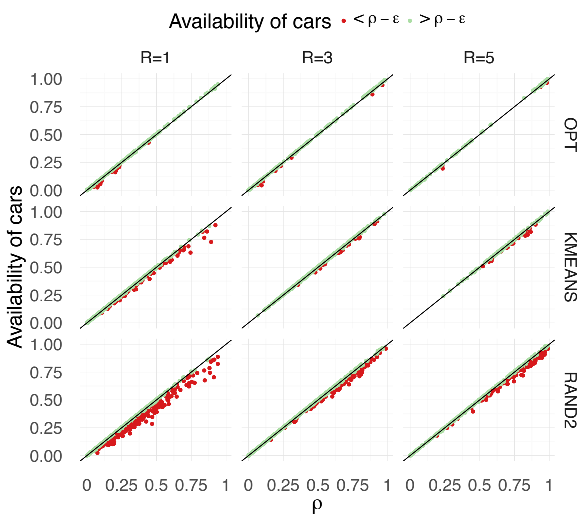


Fig. 5. Probability that cars are found, as a function of $\rho$.

## VI. Conclusions

In this work we have considered the problem of optimally deploying the station infrastructure of a car sharing system. From the car sharing operator's standpoint, an optimal deployment should minimise the cost for the stations (in terms of their number and capacity) while at the same time meeting certain pre-defined levels in the quality of service provided to the customers. We have defined this quality of service in terms of the probability that a customer will find a car/parking space upon reaching a station. Then, we have formulated an optimization problem that blends queueing theory with a variant of a set cover problem. Specifically, the optimal capacity of stations is obtained describing the station as an M/M/1/K queue, while the optimal coverage is achieved through the set covering problem. In order to solve this NP-hard problem, we have then defined a greedy approximation algorithm, that, at each step, selects the candidate station with the smallest cost per covered event. Finally, we have tested the proposed optimal algorithm using a real-life trace of pickup and drop-off events at a free-floating car sharing system. The results shown demonstrate that the proposed solution significantly reduces the costs for station deployment, while at the same time not impacting the quality of service provided to the users, when compared against three benchmark approaches.

We believe that this work has several avenues of future research. First, we want to investigate the validity of alternative queuing models besides M/M/1/K to include time-variant arrival processes (e.g., for modelling peak and off-peak arrivals), of batch arrivals (e.g., for modelling relocation schemes). Furthermore, we plan to extend the problem formulation to jointly consider the optimal planning of the parking and charging infrastructure for an electric car sharing operator.

## References

[1] R. Hampshire and C. Gaites, "Peer-to-peer carsharing: Market analysis and potential growth," *Transp. Res. Rec.*, no. 2217, pp. 119–126, 2011.

[2] B. Schwieger, P. Victorero-Solares, and D. Brook, "Global Carsharing Operators Report 2015," Team Red, Tech. Rep., 2015.

[3] M. Barth and S. A. Shaheen, "Shared-Use Vehicle Systems: Framework for Classifying Carsharing, Station Cars, and Combined Approaches," *Transport. Res. Rec.*, vol. 1791, pp. 105–112, 2002.

[4] Car2go switching electric cars to gas. [Online]. Available: http://goo.gl/6cnv6O [Last checked: Sept. 6, 2016]

[5] E. Biondi, C. Boldrini, and R. Bruno, "Optimal charging of electric vehicle fleets for a car sharing system with power sharing," in *IEEE Energycon*, 2016, pp. 1–6.

[6] P. Fairley, "Car sharing could be the EV's killer app," *IEEE Spectrum*, vol. 50, no. 9, pp. 14–15, September 2013.

[7] A. G. Kek, R. L. Cheu, Q. Meng, and C. H. Fung, "A decision support system for vehicle relocation operations in carsharing systems," *Transportation Research Part E*, vol. 45, no. 1, pp. 149–158, 2009.

[8] G. H. de Almeida Correia and A. P. Antunes, "Optimization approach to depot location and trip selection in one-way carsharing systems," *Transportation Research Part E*, vol. 48, no. 1, pp. 233–247, 2012.

[9] B. Boyaci, K. G. Zografos, and N. Geroliminis, "An optimization framework for the development of efficient one-way car-sharing systems," *European Journal of Operational Research*, vol. 240, no. 3, pp. 718–733, 2015.

[10] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, 2013.

[11] D. K. George and C. H. Xia, "Fleet-sizing and service availability for a vehicle rental system via closed queueing networks," *European Journal of Operational Research*, vol. 211, pp. 198–207, 2011.

[12] M. Pavone, S. Smith, E. Frazzoli, and D. Rus, "Load Balancing for Mobility-on-Demand Systems," *The International Journal of Robotics Research*, vol. 31, no. 7, pp. 839–854, 2012.

[13] R. Zhang and M. Pavone, "Control of robotic mobility-on-demand systems: A queueing-theoretical perspective," *The International Journal of Robotics Research*, vol. 35, pp. 186–203, 2016.

[14] O. Baron, O. Berman, S. Kim, and D. Krass, "Ensuring feasibility in location problems with stochastic demands and congestion," *IIE Transactions*, vol. 41, pp. 467–481, 2009.

[15] R. Z. Farahani, N. Asgari, N. Heidari, M. Hosseininia, and M. Goh, "Covering problems in facility location: A review," *Computers & Industrial Engineering*, pp. 368–407, 2012.

[16] J.-R. Lin and T.-H. Yang, "Strategic design of public bicycle sharing systems with service level constraints," *Transportation Research Part E*, vol. 47, no. 2, pp. 284–294, 2011.

[17] M. Harchol-Balter, *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. CUP, 2013.

[18] C. Boldrini, R. Bruno, and M. Conti, "Characterising Demand and Usage Patterns in a Large Station-based Car Sharing System," in *Proc. of IEEE INFOCOM Workshop on Smart Cities and Urban Computing*, 2016, pp. 1–6.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press Cambridge, 2001, vol. 6.

[20] A. B. Clarke *et al.*, "Maximum likelihood estimates in a simple queue," *The Annals of Math. Stat.*, vol. 28, no. 4, pp. 1036–1040, 1957.