Transportation Research Part B 150 (2021) 370–385

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Performance of reservation-based carpooling services under detour and waiting time restrictions

Yanfeng Ouyang[a,*], Haolin Yang[a], Carlos F. Daganzo[b]

[a] University of Illinois, Urbana-Champaign, IL 61820, United States
[b] University of California, Berkeley, CA 94707, United States

ARTICLE INFO

Keywords:
Carpooling
Reservation-based
Detour
Wait time
Guarantee
Vehicle distance traveled
Passenger distance traveled

ABSTRACT

This paper examines many-to-many carpooling services with advance reservations, and constraints on waits and detours. An analytic model yields approximate formulas for the percent of requests matched, the expected vehicle-distance driven, and the passenger-distance traveled in some idealized scenarios. Simulations of these scenarios validate the formulas. In the most favorable cases carpooling reduces the vehicle-kilometers driven by all users by a few percent. The paper also shows how the formulas can be used by service providers to optimize offerings, and by city governments to design regulatory policies that will perform as intended. A simple example illustrates the ideas.

1. Introduction

Many different kinds of ridesharing services have been rolled out by transportation network companies (TNCs), such as Uber and Lyft in the U.S., Yandex in Europe, and Didi-Chuxing in China. These services have become an important transportation mode for many cities; see Chan and Shaheen (2012) and Furuhata et al. (2013) for recent surveys. Carpooling stands out amongst these services because it does not require dedicated chauffeurs. This simplifies its management and implementation.

With carpooling each vehicle belongs to a traveler who may accept an extra rider, and as a result has to execute some detours before completing his/her own trip. The vehicle is used only once. Clearly, there are two groups of users: "drivers" who provide a car and drive it (for the complete trip of two travelers if successfully matched); and "riders" who merely ride as passengers for the duration of their own trip. As discussed in Daganzo and Ouyang (2019), a carpooling operation can be spontaneous (also called "dynamic" in the literature) or based on reservations (also known as "static").

In spontaneous systems a user declares whether he/she is interested in getting or providing a ride moments before the intended trip. Unfortunately, if a spontaneous rider is not matched, the rider is forced to find a travel alternative with little lead time. Since this quite undesirable, especially if the rider does not have a car, spontaneous systems tend to be found only in special situations where travel alternatives exist and/or the probability of a match is extremely high; e.g., during the morning commute near transit stations serving concentrated destinations. See Graziotin (2013) for more discussion.

In reservation-based systems, users declare their intentions well ahead of time; e.g., a day or hours before travel, so in case of failure users can make alternative travel arrangements. For this reason, these systems may be of use in more general, many-to-many scenarios with dispersed origins and destinations. This is the central question in this paper. Many current carpooling services in the U.S., such as

* Corresponding author.
E-mail address: yfouyang@illinois.edu (Y. Ouyang).

https://doi.org/10.1016/j.trb.2021.06.007
Received 3 November 2020; Received in revised form 20 May 2021; Accepted 17 June 2021
Available online 12 July 2021
0191-2615/

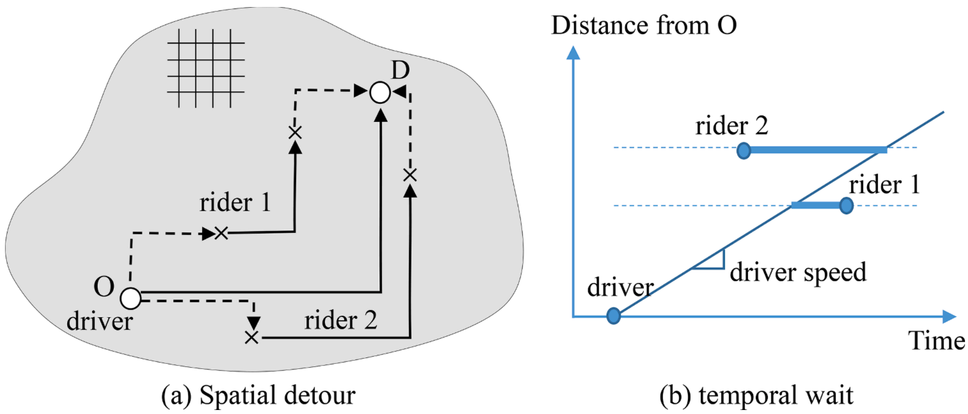


Fig. 1. User inconvenience in carpooling.

those offered by Scoot and Waze Pool, are reservation-based.

Like other ridesharing services, carpooling imposes spatial and temporal penalties on its users. The spatial penalty arises from detours. Fig. 1(a) shows how. It depicts a service region with an underlying square grid of streets. It also displays the origin O and destination D of a driver, whose intended path is marked by a solid arrow. Two ride requests are also shown by means of crosses and arrows. Although our driver can serve rider 1 without any increase in travel distance—by altering its original path as shown by the dashed arrows—the driver would have to detour and suffer a penalty to serve rider 2.

The temporal penalty arises from waiting. If the intended departure time of the driver plus the trip time to a rider's origin is significantly different from the intended departure time of the rider, then the rider has to adjust its departure time and experience some wait. As illustrated by Fig. 1(b), this wait can be positive or negative. The circles depict the users' intended departure times, as well as their relative locations with respect to "O". The slanted line is the position of the driver at different times. The length of the thick horizontal segments are the waits—negative for rider 1 and positive for rider 2. Clearly, although the detour penalty is solely borne by the driver, the penalty from waiting can be borne by either user.

Because long detours and waits may suppress demand, TNCs may want to offer services that guarantee low detours and waits. This should be done judiciously, however, because very strict guarantees can reduce the matching rate and depress participation. In view of this, this paper will build on the existing literature by systematically investigating how guarantee levels affect the matching rate and service performance.

For the most part, the literature on service guarantees describes simulations of specific scenarios, with more emphasis on temporal flexibility than spatial guarantees. For example, Hosni et al. (2014) formulated the matching and routing problem as a mixed-integer program in which the users' temporal flexibility was taken as an input. But the model did not enforce limits on detours. As a result spatial detours could be as large as 75% of the direct trip distance. In the same spirit, Ghilas et al. (2016) and Stiglic et al. (2018) used simulations to evaluate performance, using the passengers' latest acceptable arrival time as the reference point to express temporal flexibility. Spatial flexibility seems to have been addressed only in Stiglic et al. (2015, 2016), which simulated some corridors and networks to show how the "matching rate" is influenced not just by temporal flexibility but also by detour limits and the introduction of meeting points.

These simulation efforts are accurate, realistic and very informative. However, by evaluating only a few scenarios and producing numerical results only for some metrics, the results cannot be easily generalized or used as inputs for mathematical programs, e.g., to optimize the services offered in arbitrary settings or to make multi-modal plans.

On the analytic front, the only analytical work that applies to carpooling appears to be Section 7.3 in Daganzo and Ouyang (2019). This reference, however, only examines an unrealistic special case where distance-adding detours are not allowed at all, so the matching rate is very low. As a result, the work is incomplete.

In the broader context of ridesharing, more efforts have been made very recently. Ke et al. (2020) derive pairing rate and average detour time for ridesharing systems based on a simplifying assumption that the random detour length between consecutive requests follows a (truncated) exponential distribution; spatial distribution of origins/destinations of these trips are not considered in the development of these formulas. Zhang and Nie (2019) derive formulas for expected wait time for both solo riders and pooling riders under different pricing and regulation strategies, and later incorporate the results into an inter-platform competition model (Zhang and Nie, 2020). Some more germane works (e.g., Yang et al., 2020; Daganzo et al., 2020) present analytical formulas for systems with both waiting and detour guarantees. But these works only consider chauffeured systems. The presence of chauffeur as well as spontaneous (on-call) service leads to significantly different likelihood of spatiotemporal feasible matches. In addition to that, the derivation in Daganzo et al. (2020) ignores, for the sake of simplicity, spatial variability of drivers and riders as well as the possibility of competition among drivers and riders when many-to-many matches are feasible under high demand. In light of these, this paper found a simple way to remove those simplifications and derive much more accurate formulas that are still in a closed form.

To fill this gap, this paper will present closed-form formulas for reservation-based carpooling systems with broad, generic operating strategies. Analysis is conducted for idealized settings; e.g., involving a square city with dense city streets and homogeneous demand.

The derived formulas express how temporal and spatial guarantees affect three important metrics for many-to-many trip services: the matching rate; the expected passenger distance traveled; and the expected car distance driven. The derivation of these formulas is quite involved, so we adopt a "building block" approach, starting with a basic scenario and then building on it. The basic "low-demand" scenario has two features: (i) the demand is so low that each user can encounter at most one candidate for a feasible match, and (ii) every user declares a unique role as either a driver or rider. We then consider "high-demand" and relax (i). In this scenario, multiple candidates for feasible matches are available for a driver or a rider. The results are presented as second-order corrections to those of the basic scenario. Finally, we relax (ii) and present additional corrections for the case where all users are flexible and the matching algorithm determines their roles.

The derivations in this paper are not intended to reproduce what TNCs are currently doing, but to explore possibilities. The specific goals are: (i) analyzing a broad space of carpooling strategies that could be considered by TNCs (e.g., pertaining to detour limits, waiting limits and the declaration of user roles); (ii) providing closed-form formulas to evaluate the performance of such strategies; and (iii) help illustrate how TNCs and cities can use these formulas to optimize decision-making. For example, TNCs could use the formulas to customize their service offerings and prices for different cities; and cities could use them to anticipate how TNCs may respond to incentives and regulations, and avoid unintended consequences. Although (iii) could be a research topic in itself, the basic idea will be illustrated with a simple 'toy" example.

The remainder of this paper is organized as follows. Section 2 develops approximate formulas for the basic scenario with low demand and fixed user roles; Section 3 extends the formulas to high demand; and Section 4 to the flexible-role scenario. Section 5 then validates the formulas, with micro-simulations of the proposed strategies in the considered scenarios. Finally, Section 6 presents the toy example, showing how the formulas can be used for policy analysis; and Section 7 closes with some comments.

2. Low-demand

We consider a square service region of size R [km^2], covered by a dense grid of streets parallel to the square's sides. Vehicles travel at speed v [km/hr]. The stopping time to collect and deliver drivers is negligible. User trips are generated steadily over time and space according to a homogenous Poisson process, at rate λ [rides/hr-km^2]. The users' origins and destinations are uniformly and independently distributed over the square. Every user is tagged as a "rider" with probability f and as a "driver" with probability $(1-f)$. Thus, the demand density of riders is λf and that of drivers $\lambda(1-f)$.

The scenario under consideration also has some operational features. First, every successful user match involves exactly one driver and one rider, and the driver carries the rider. Second, unmatched users drive alone. In addition, user trips must be matched not only spatially such that the driver is never detoured by more than an upper limit d [km], but also temporally within a tolerance window of $(-\tau/2, \tau/2)$ [hr] around the rider's desired departure time. Because each driver (or rider) can possibly find multiple riders (or drivers) that can form a feasible match, a bipartite matching algorithm should be used to select the best set of feasible matches.

This section considers very low demand λ, so low in fact that the probability that any rider can be feasibly matched with more than one driver is negligible. The simplification allows us to dispose with the bipartite matching algorithm and tackle the problem analytically by considering users individually.

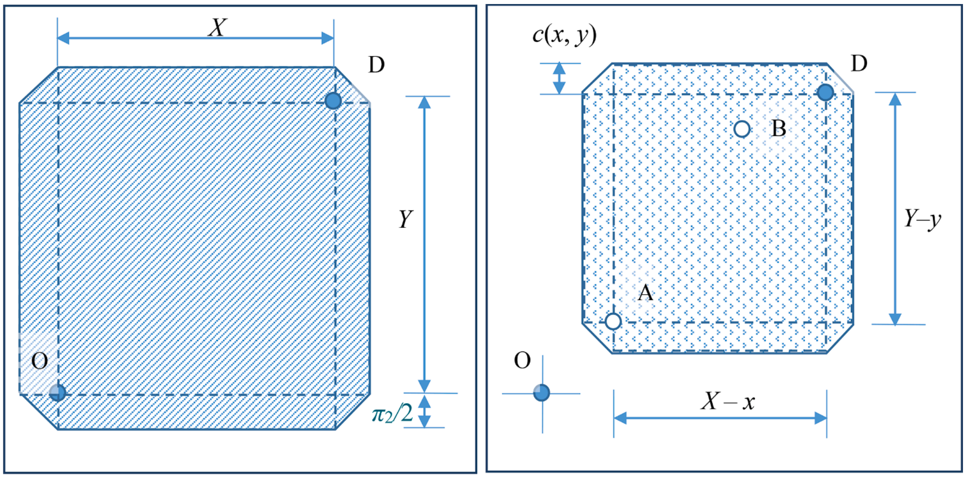
To do this, we shall consider a single randomly selected driver and all the riders that can be feasibly matched to him/her. (Note that these riders are exclusive to our driver because of the low-demand assumption, and that they can be considered one at a time due to our independence assumptions.) The formula is developed in two stages, using these ideas. First, Section 2.1 examines a driver/rider pair where the driver is given and the rider is random, ignoring for the moment their temporal match. It develops an expression for the probability that the match is "spatially feasible", in the sense that driver detours do not exceed the limit. The subsection also develops formulae for the expected vehicle distance saved and the user distance added by a successful match. Then, Section 2.2 adds the temporal dimension to the mix. It derives formulae for the probability that a spatiotemporally feasible match exists and for the fraction of users in the population that can be matched. Finally, Section 2.3 provides additional formulas for the global savings in vehicle-distance driven and the global increase in passenger distance traveled.

2.1. Spatial feasibility of a random rider for a given driver

Before getting into details, note that the system of interest can be specified by six parameters $\{\lambda, f, R, v, \tau, d\}$, which collectively involve two dimensions (time and distance). With this in mind, dimensional analysis (Johnson, 1944; Daganzo and Ouyang, 2019) allows us to reformulate the problem in terms of only four independent dimensionless inputs: $\{f, \pi_0 \equiv \lambda R^{3/2}/v, \pi_1 = \tau v/R^{1/2}, \pi_2 \equiv d/R^{1/2}\}$.[1] These reformulated inputs can be interpreted as follows: Parameter π_0 is a dimensionless measure of demand, expressing the number of user requests in the whole region during the time for the vehicle to traverse the region; parameter π_1 is the ratio of the tolerance time to the time it takes to cross one side of the region; and π_2 is the ratio of the detour limit to the side of the region. Both thresholds are assumed to be far smaller than the size of the region; i.e., $\pi_1 << 1$ and $\pi_2 << 1$.

Consider now a driver/rider pair and the spatial feasibility of their match. We look first at the probability of a successful match conditional on the driver, and then at the changes in expected distance arising from a successful match.

[1] This way, as long as the values of these four inputs are fixed, any performance of the system is also fixed regardless of the individual values of the original six parameters. This reduction of "degree of freedom" helps reveal useful interrelationship among parameters; e.g., in case of need, the number of simulations that must be run can be reduced by orders of magnitudes.



(a) Origin catchment (b) Destination catchment

Fig. 2. Feasible catchment areas for a rider's origin and destination.

2.1.1. Probability of a spatially successful match

Note, the only relevant parameters in this problem are d and R. Therefore, by choosing the distance unit so that $R = 1$ we can model our region as a unit square with detour limit: $d = \pi_2$. No generality is lost.

To start, refer to Fig. 2, which shows our square region. The driver's given origin and destination are points O and D. It is assumed without loss of generality that the picture has been rotated so D is up and to the right from O, as shown.

We then sample a random rider with origin A and destination B anywhere in the square. For this rider to be spatially feasible it must be possible to add its origin to the driver's route from O to D without increasing length by more than π_2. This is a necessary but insufficient condition. If it is met we say that the rider is "origin feasible". Consideration shows that the condition is met if point "A" is in the shaded catchment area of Fig. 2(a). Note this area includes an inner rectangle with opposite corners at O and D, where the added distance is zero, plus a periphery with a width of $\pi_2/2$ and beveled corners. Note how points on the outer edge of the periphery increase the driver's route length exactly by π_2.

For the complete rider trip to be spatially feasible, it should also be "destination feasible"; i.e., it should be possible to add point B to route OAD with a detour whose length does not exceed the balance available after the pickup at point A. If this happens then the length of route OABD will not exceed the length of route OD by more than π_2, as required. This is illustrated in Fig. 2(b), where O, D are solid circles and A, B are hollow. Consideration shows that for the given positions of points O, A and D, the catchment area for feasible destinations B is the dotted polygon shown. As before, it consists of an inner rectangle and a periphery. However, the rectangle's corners are now at A and D; and the periphery's width is now one half of the detour distance that can still be used after the pickup at A.

To analyze this problem we shall use a rectangular coordinate system with origin at O and axes parallel to the square's sides. The coordinates of points D and A are respectively labeled (X, Y) and (x, y); see figure. Recall that the former are given and the latter are random.

Consideration of the problem's geometry reveals that the detour distance required for the pickup at A is twice the L_1 distance from A to the perimeter of the inner rectangle formed by O and D. Keeping this in mind, the detour length for A can be expressed as: $2\{\max(-x, 0) + \max(x-X, 0) + \max(-y, 0) + \max(y-Y, 0)\}$. Clearly, then:

$c(x, y) = \pi_2/2 - [\max(-x, 0) + \max(x-X, 0) + \max(-y, 0) + \max(y-Y,0)] \le \pi_2/2.$

Note that $c(x, y) = \pi_2/2$ if A is in the inner rectangle formed by O and D (i.e. if $0 \le x \le X$ and $0 \le y \le Y$), and that $c(x, y) \ge 0$ if and only if the origin is feasible.

Note that regardless of where A is located relative to O and D, the destination catchment is always composed of an inner rectangle pinned at A and D, four peripheral rectangles of width $c(x, y)$, and four triangles of size $½c(x, y)^2 = O(\pi_2^2)$. Accordingly, the destination catchment size is:

$$|X-x||Y-y| + 2c(x,y)(|X-x| + |Y-y|) + O(\pi_2^2) \text{ where } c(x,y) \ge 0. \tag{1}$$

Since the region's size is 1, expression (1) is also the probability that our rider's destination is feasible, conditional on both the driver's data (X, Y) and the rider's origin (x, y). Thus, integrating (1) with respect to (x, y) over the origin catchment area, we obtain the probability that our rider is both origin- and destination-feasible, conditional on (X, Y). This is what we set out to find, which we denote $N(X, Y)$.

To simplify the final formula for $N(X, Y)$ we neglect the last high-order term in (1) and approximate the integration domain by the slightly larger rectangle $\left[-\frac{\pi_2}{2}, X + \frac{\pi_2}{2}\right] \times \left[-\frac{\pi_2}{2}, Y + \frac{\pi_2}{2}\right]$. This is reasonable because $\pi_2 << 1$. Then, a few algebraic manipulations yield:

$$N(X,Y)=\frac{X^2Y^2}{4}+\frac{3\pi_2}{4}XY(X+Y)+\frac{\pi_2^2}{8}\left(3X^2+4XY+3Y^2\right)+\frac{\pi_2^3}{12}(X+Y). \tag{2}$$

Note, the derivation of (2) assumes that the origin and destination catchments are not truncated by our region's boundary. Although such truncation can happen, the event should be rare because $\pi_2 << 1$. Therefore we expect (2) to be reasonably accurate despite all our simplifications. Eq. (2) is the probability formula we were seeking. It will be used later.

2.1.2. Changes in the expected distances traveled after a successful match

The above simplifications can also be used to derive the changes in the expected distances traveled by both, vehicles and users when our random rider is successfully matched with our given driver. We consider vehicle distance first.

The reduction in vehicle distance due to a successful match is a random variable which we denote L. We look for its expectation conditional on (X, Y). Note, this expectation is a function of (X, Y). Therefore, it is denoted $L(X, Y)$.

The random variable L is the difference between the vehicle distances traveled with and without carpooling; i.e., the difference between the combined length of the two individual paths OD and AB, and the length of path OABD. Thus, using an overbar for the length of a path, we can write: $L=[\overline{OD}+\overline{AB}]-[\overline{OA}+\overline{AB}+\overline{BD}]=\overline{OD}-\overline{OA}-\overline{BD}$. Now note that $\overline{OD}=X+Y$, $\overline{OA}=|x|+|y|$, and if we use (x_B,y_B) for the coordinates of "B", then $\overline{BD}=|X-x_B|+|Y-y_B|$. Thus, $L=X+Y-|x|-|y|-|X-x_B|-|Y-y_B|$. The conditional expectation $L(X, Y)$, is the expectation of this formula with respect to (x, y) and (x_B,y_B).

To derive an expression, we recognize that both of these tuples are uniformly and independently distributed over their respective catchments, and for simplicity we approximate both catchments by slightly larger rectangles—as we did in the previous subsection for the origin catchment. Unfortunately, because the catchment of "B" depends on "A", our two 2-tuples are not independent. This complication can be resolved by taking the expectation in two steps: first with respect to (x_B,y_B) conditional on (x, y), and then with respect to (x, y). If this is done, the manipulations are straight-forward, albeit somewhat lengthy. For these reasons, they are omitted here but briefly summarized in Appendix A. The final result is:

$$L(X,Y)=\frac{X+Y}{3}+\frac{\pi_2^2}{144N(X,Y)}\left(21XY(X+Y)+\left(11X^2+16XY+11Y^2\right)\pi_2+3(X+Y)\pi_2^2\right). \tag{3}$$

Consider now the users. In this case the added distance is L' and its conditional expectation, $L'(X, Y)$. Note that $L'=[\overline{OABD}+\overline{AB}]-[\overline{OD}+\overline{AB}]=\overline{OABD}-\overline{OD}$. Algebraically, this translates to: $L'=|x|+|y|+|x_B-x|+|y_B-y|+|X-x_B|+|Y-y_B|-X-Y$.

Although the formula for L' is slightly different from the formula for L, the same derivation logic applies. And again, the reader can refer to Appendix A for the steps that give the following final result:

$$L'(X,Y)=\frac{\pi_2^2}{48N(X,Y)}\left(18XY(X+Y)+\pi_2\left(12X^2+16XY+12Y^2\right)+3\pi_2^2(X+Y)\right). \tag{4}$$

2.2. Successful matches: probability and the expected total for a random driver

The above results are now used to derive formulae for both, the probability with which a random driver can be spatiotemporally matched and for the total number of users in a given population that can be successfully matched. Since we are assuming that each rider can be matched with only one driver (as per the low demand approximation) we will imagine that each driver scans the complete population of riders for possible matches and chooses the best. This is possible because conflicts with other drivers cannot arise and because the system is reservation-based so all the information is available when the choices are made.

Let F be the number of spatiotemporally feasible matches available to a random driver, so that the sought probability is $\Pr\{F>0\}$. To express this probability we will first look for the distribution of F conditional on the driver (X, Y), which we denote $[F|(X, Y)]$ as is customary, and then remove the conditioning constraint.

As a preliminary step in evaluating this conditional distribution, we first look for the expected number of temporally feasible riders available to our driver. Note that for a match to be temporally feasible the driver must be able to reach the rider's origin "A" in a time window of duration τ that straddles the rider's desired departure time. Now, since "A" can be anywhere in the region, and the rider generation rate is time-independent and homogeneous, it follows that the expected number of temporally valid riders available to our driver is the product of the rider generation rate, λf, the time window duration, τ, and the region's size, R; i.e., it is $\lambda fR\tau=f\pi_0\pi_1$. This is true regardless of the driver's origin and destination.

Moreover, since the riders' spatial coordinates are independent of time it follows that each temporally feasible rider has the same probability of being spatially feasible. This probability was found to be $N(X, Y)$ in Section 2.1. And since the number of temporally feasible riders is a Poisson random variable, it follows that the number of those who are spatiotemporally feasible, $[F|(X,Y)]$, is itself a Poisson random variable with mean $f\pi_0\pi_1N(X, Y)$. Hence, the probability that driver (X, Y) has a successful match is $1-e^{-f\pi_0\pi_1N(X,Y)}$.

Note now that F is a probability mixture of these Poisson variables. Therefore, the unconditional probability $\Pr\{F>0\}$, which we abbreviate as p_1, is the average of $1-e^{-f\pi_0\pi_1N(X,Y)}$ across (X, Y). This average can be easily found because both X and Y are mutually independent distances between two uniformly distributed random points in a segment of length 1, and as such they each have a triangular probability density in [0, 1] (with the mode at 0). Thus, taking expectation over $[0, 1]^2$, the formula for p_1 is:

$$p_1 = \int_0^1 \int_0^1 \left(1 - e^{-f\pi_0\pi_1 N(X,Y)}\right) \times 2(1-X)dX \times 2(1-Y)dY. \tag{5}$$

The value of p_1 for any set of π's can be obtained numerically.

It is also possible to derive closed-form analytic approximations of (5) in several ways. The one we favor assumes that the mean, $f\pi_0\pi_1 N(X, Y)$, of our conditional Poisson random variable can be approximated by a gamma distribution. This is reasonable because the value of expectation (5) should mainly depend on the mean and variance of our conditional random variable.[2] The approximation is useful because, as is well known, gamma-mixtures of Poisson variables are negative binomial random variables. Therefore, F follows this negative binomial distribution. Furthermore, if we use $(n, \omega) \in \mathrm{R}+$ for the mean and variance of the gamma variable, then the parameters of the corresponding negative binomial variable F are $n^2/\omega \in \mathrm{R}+$ and $\omega/(n+\omega) \in (0, 1)$. Thus, $\Pr\{F = 0\} = [n/(n+\omega)]^{n^2/\omega}$, so we finally have:

$$p_1 \equiv \Pr\{F > 0\} = 1 - \Pr\{F = 0\} \approx 1 - [n/(n+\omega)]^{n^2/\omega}. \tag{6}$$

The mean n and variance ω of $f\pi_0\pi_1 N(X, Y)$ are easily obtained because $N(X, Y)$ is a third order polynomial in X and Y, and the joint distribution of (X, Y) is simple. The reader can verify that:

$$n = \frac{f\pi_0\pi_1}{144}\,(1 + 12\pi_2), \tag{7a}$$

$$\omega = (f\pi_0\pi_1)^2 \left(\frac{119}{518400} + \frac{83\pi_2}{21600}\right) \tag{7b}$$

The combination of (6), (7a) and (7b) is the proposed approximation to (5).

The fraction r of users that are successfully matched in a large population can now also be written. Note, for each user there are $(1-f)$ drivers; and therefore $(1-f)p_1$ drivers who are successfully matched. And, since every successful match involves exactly one rider and one driver, there must be $2(1-f)p_1$ matched users for each user; i.e.:

$$r \approx 2(1-f)p_1. \tag{8}$$

2.3. Distance predictions

Let us now turn our attention to expected vehicle distance saved and user distance added per individual user trip, denoted δ and δ' respectively. Think of these quantities as the ratios of the expected total vehicle (or user) distance saved (or added) to the number of users.

Since the expected vehicle distance saved by a successfully matched driver is $L(X, Y)$, and the probability of a successful match is $1 - e^{-f\pi_0\pi_1 N(X,Y)}$, it follows that the expected savings per driver, which we denote $l(X, Y)$, are:

$$l(X, Y) = \left(1 - e^{-f\pi_0\pi_1 N(X,Y)}\right)L(X, Y).$$

To obtain the unconditional expectation, which we denote l, plug (2) and (3) into the above and take the expectation with respect to X, Y. The formula is:

$$l = \int_0^1 \int_0^1 \left(1 - e^{-f\pi_0\pi_1 N(X,Y)}\right)L(X, Y) \times 2(1-X)dX \times 2(1-Y)dY. \tag{9}$$

Similarly, we can get the expected increase in passenger distance l' by replacing $L(X, Y)$ by $L'(X, Y)$ in the derivation of (9). The result is:

$$l' = \int_0^1 \int_0^1 \left(1 - e^{-f\pi_0\pi_1 N(X,Y)}\right)L'(X, Y) \times 2(1-X)dX \times 2(1-Y)dY. \tag{10}$$

Eqs. (9) and (10) can be computed numerically. They give the expected changes in distances (in distance units of $R^{1/2}$) per driver.

Finally, to express these changes per user (a more intuitive measure) we multiply (9) and (10) by the fraction of drivers, which is $(1 - f)$. The result is:

$$\delta = (1-f)l, \text{ and } \delta' = (1-f)l'. \tag{11}$$

---

[2] To see this, recall from Taylor series expansion that $1 - e^{-f\pi_0\pi_1 N(X,Y)} \approx f\pi_0\pi_1 N(X,Y) - ½\,[f\pi_0\pi_1 N(X,Y)]^2$ for small values of $f\pi_0\pi_1 N(X,Y)$. Hence, the first two moments of $f\pi_0\pi_1 N(X,Y)$ would dictate the value of (5).

3. High demand

In this section, we consider normal levels of demand so a rider can encounter multiple drivers as candidates for a feasible match. In this case, driver-based formulas (8) and (11) may become inaccurate because a driver's candidate riders might not be available for the driver. To account for this effect, we first develop an approximate expression for the probability distribution of the rider's choices in the bipartite matching, and then use this expression to obtain an improved formula for p_1.

We first note that the (probabilistic) choices of a rider are different from those of a driver, simply due to the geometric asymmetry between their trips (i.e., since a driver's trip must "contain" that of the rider). To obtain the distribution, therefore, we must repeat the analysis of Section 2.1 from the perspective of a rider. To stay true to this perspective, we use the fraction of riders g rather than the fraction of drivers f to characterize our user population. The steps are summarized in Appendix B. It shows that the number of spatiotemporally feasible drivers corresponding to an arbitrary rider has the following mean m and variance ψ :

$$m \approx \frac{g\pi_0\pi_1}{144}\ (1+12\pi_2). \tag{12a}$$

$$\psi \approx (g\pi_0\pi_1)^2\left(\frac{7}{186624}+\frac{7\pi_2}{4320}\right), \tag{12b}$$

where $g = 1 - f$.[3]

For the same reasons that led to (6) and (7) in Section 2.2, the distribution of the number of feasible drivers for a random rider, which we denote K, is now negative binomial with parameters m^2/ψ and $\psi g/(m+\psi g)$.

To obtain our correction term we will imagine that each of the F riders matched with a random driver (as per the analysis of Section 2.1) is now asked for his/her availability to join the carpool. In Section 2.1 the rider was assumed to be always available because under low demand each rider encounters at most one driver that can be feasibly matched, so a carpool match will be formed if $F > 0$. Now, however, the answer can be "no."

To evaluate this effect we now derive the probability, p_2, with which a selected rider replies "yes". Because the rider is "selected" we know that $K > 0$. Moreover, we also know that if the rider has $K = k$ potential driver matches, and if for simplicity we ignore the rider's possible preference for certain drivers (e.g., those imposing less delay), then the probability of saying yes to one of them is $1/k$. It then follows that the probability of acceptance is:

$$p_2 = \sum_{k=1}^{\infty}\frac{1}{k}\Pr\{K=k\} \Big/ \Pr\{K>0\}. \tag{13}$$

Let us now use "j" as the dummy argument for the mass function of F, which we momentarily abbreviate by $\Pr(j)$. Then, the probability of receiving at least one acceptance conditional on $\{F=j\}$ is $[1-(1-p_2)^j]$. It then follows from the total probability theorem that:

$$\begin{aligned} p_1 &= \sum_{j=0}^{\infty}\Pr(j)\left[1-(1-p_2)^j\right] = 1-\sum_{j=0}^{\infty}\Pr(j)(1-p_2)^j \\ &= 1-\sum_{j=0}^{\infty}\begin{pmatrix} j+\frac{n^2}{\omega}-1 \\ j \end{pmatrix}\left(\frac{\omega}{n+\omega}\right)^j\left(\frac{n}{n+\omega}\right)^{n^2/\omega}(1-p_2)^j \\ &= 1-\left(\frac{n}{n+\omega p_2}\right)^{n^2/\omega}. \end{aligned} \tag{14}$$

This is the improved formula.

A similar correction can be applied to the distance formulas. In this case, the probability that an arbitrary driver (characterized by X and Y) is matched with a rider is now approximated by $N(X, Y)p_2$.[4] Using this simplification, we can simply replace the $(1 - e^{-f\pi_0\pi_1 N(X,Y)})$term in (9) and (10) by $(1 - e^{-f\pi_0\pi_1 N(X,Y)p_2})$, and still use (9)–(11).

4. Flexible user roles

In the scenarios of Sections 2 and 3, each user has a predetermined role as either a driver or rider – i.e., that the fraction of drivers (1 – f) in the user pool is known. This fraction was the basis for (8), which returns the match success rate r from the probability of matching a driver p_1. There may also be situations where some users can perform either role. It should be clear that if these flexible

[3] Note how (7a), (7b) and (12a), (12b) are structurally similar but not identical. Although the formula for the expected number of matches is the same in both cases, the variance is now greater. This is due to the asymmetry between the riders and drivers.

[4] Although this approximation wrongly treats p_2 as if it was independent of (X, Y), its result should be much better than arbitrarily setting $p_2 = 1$ by ignoring the correction.

users allow the assignment algorithm to define their roles, the overall system performance should improve. To explore the potential of this feature, this section examines the extreme case where every user is flexible. We will still base the result on the probability p_1 of matching a random driver. However, because *f* is now undefined, the logic relating p_1 and *r* has to be changed.

The flexible system is assumed to work as follows. Each user submits his/her trip information well in advance without specifying a role. The platform then identifies the candidates for feasible matches for every user, both as a rider and driver. Once done, the platform optimizes all the matches and tells the users what to do; i.e.: ride with someone (and be a "successful rider"), carry someone (and be a "successful driver"), or travel alone (and be an "unsuccessful user").

In order to derive formulae to predict performance we will assume for simplicity that a simple greedy matching algorithm is used. Section 5 will compare the results against simulations that use an optimal matching algorithm.

The greedy algorithm tries to find matches for all the users in the population by processing them individually one-step at a time. After a user is processed, (s)he is labeled as "considered." If a match is found, both users in the match are also labeled as "matched." To perform a step, we select a user from the current pool of unconsidered and unmatched users and perform two sub-steps – if the pool is empty the algorithm ends. In the first sub-step (a) we treat the selected user as a rider and try to match it with a driver. If we succeed, we label them both as "matched" and proceed to the next step. Otherwise, we perform sub-step (b). Now we treat the selected user as a driver and try to match it with a rider. If successful, we label both users as "matched." Then, regardless of this outcome, we label the processed user as "considered", and proceed to the next step/user.

Now note from the just-described algorithm that the only way a user can become a successful driver is by: (i) not becoming a rider prior to being considered for a driver's role in sub-step (b), and then (ii) succeeding in this sub-step. In view of this, we approximate the probability of becoming a successful rider, which we know is $r/2$, as the product of the probabilities of events (i) and (ii).[5] The probability of (i) equals the fraction of users who are not riders $(1 - r/2)$. The probability of (ii) is the probability of success in sub-step (b), which should be close to p_1.[6] Thus, we can write:

$$r/2 \approx (1 - r/2)p_1, \tag{15}$$

or equivalently:

$$r \approx 2p_1/(1 + p_1). \tag{16}$$

As expected, this value is larger than that predicted by (8), due to the flexible roles our users can play. The improvement is notable. For example, in the symmetric case where $f = 0.5$, formulas (8) and (16) reveal that flexibility increases the value of *r* by almost 100% when p_1 is small.

To estimate the impact of flexible roles on distances saved/added we assume that formulas (9) and (10) still apply approximately to our greedy algorithm. This seems reasonable because both matching algorithms ignore distance when assigning a match. Now recall that (9) and (10) give the total changes in distances prorated across all drivers. Therefore, since the fraction of users who are drivers (solo or otherwise) is $(1 - r/2)$ we can express the changes per user as:

$$\delta = l(1 - r/2), \text{ and } \delta' = l'(1 - r/2), \tag{17}$$

where l and l' are given by (9) and (10).

We conclude this section by stressing that the formulas provided here are only rough approximations. Furthermore, they should underestimate slightly the benefits to society due to the extreme simplicity of the algorithm used. To explore this idea, the formulas are now compared against the results of simulations that use a more sophisticated matching algorithm.

5. Numerical verification

The benchmark for comparison is a simulation that matches randomly generated[7] driver and rider trips with an algorithm that maximizes the total vehicle distance savings.[8] All unmatched drivers and riders are recorded as solo travelers. The optimization problem is a standard bipartite matching problem (a linear program) that is solved with commercial package *Gurobi*. No simplifying assumptions are used in the simulation.

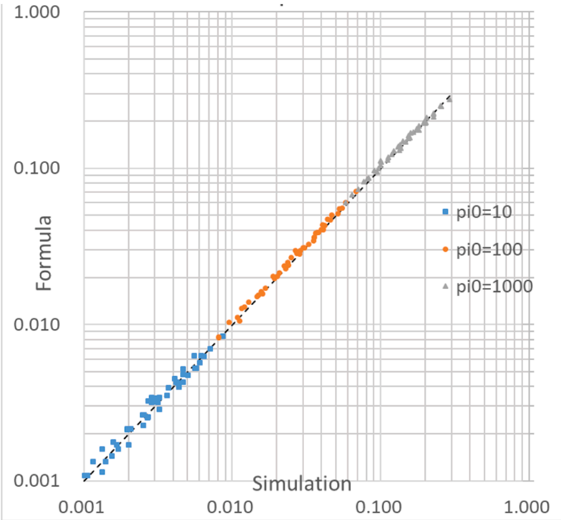
Even though the system is reservation-based, the earliest drivers and riders face considerably fewer options due to the strict detour and waiting time limits. Hence, in each simulation run, we discard the users' experiences from "warm up" or "cool down" periods consisting of about the first 5,000 and last 5,000 trips, but record those of the middle 90,000 users. These data were used to estimate *r*, δ, and δ', and these quantities were then compared against those predicted by expressions (8) and (11).

[5] An approximation is involved because events (i) and (ii) are not strictly independent; e.g., a user with a longer trip length is more likely to satisfy (i) and (ii) simultaneously.

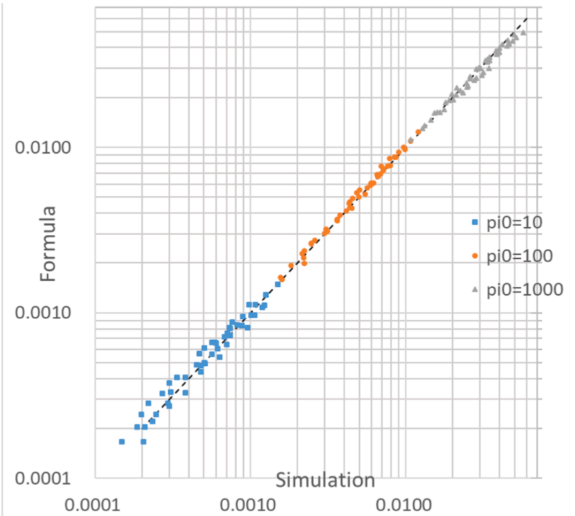
[6] The result is not exact because the users in our algorithm are pre-screened in sub-step (a) and therefore not purely random as was assumed in the derivation of p_1.

[7] Rider and driver trips are generated in space and time from homogeneous and independent Poisson processes inside a unit square, with all operating conditions conformable to what was assumed in the derivation of the formulas.

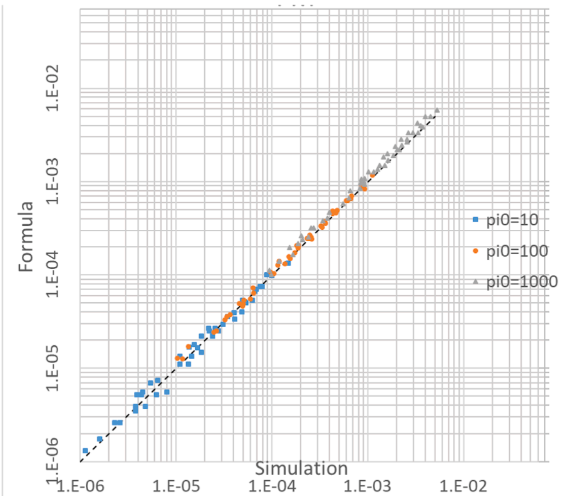
[8] Other objective functions such as maximizing the total number of matches were also tested; the system performance turned out to be very similar.



(a) Probability of successful matches, r



(b) Expected VKT saving per trip



(c) Expected PKT added per trip

(caption on next page)

Fig. 3. Comparison of performance metrics from matching algorithm and formulas under fixed fraction of driver/rider.

For illustration purposes, we focus on medium sized cities like San Francisco (population ~800,000, R~100 km^2, $v \sim 40$ km/hr), where buses, streetcars, and all taxi companies combined carry about 340 thousand trips daily (Bliss, 2018). In such cities, π_1=π_2=0.1 translates into 1.5 min of waiting time and 1 km of detour limit. Assuming that 5% of these trips use carpooling service within about 17 hours, we have $\pi_0 \sim 100$ (when $\pi_1 = 0.1$). The daily ridership in mega cities like NYC, or in small towns like Champaign, IL, is about an order of magnitude larger or smaller. Fig. 3 plots the simulated and predicted values of r, δandδ′ for the following parameter combinations: $f \in \{0.25, 0.5, 0.75\}$,$\pi_0 \in \{1, 10, 100\}$, $\pi_1 \in \{0.025, 0.05, 0.075, 0.1\}$, and $\pi_2 \in \{0.025, 0.05, 0.075, 0.1\}$.

It can be seen from Fig. 3 that, for the selected parameter values, the formulas approximates the r, δ, and δ' values quite accurately, and the differences are generally no more than 3-5%. Also, the smaller the thresholds, the less approximation error (from omitted higher order terms) in the spatial derivation. In Fig. 3(a) and (b), the probability r and VKT savings δ increase almost linearly with π_2, showing benefit of larger thresholds. However, the added PKT δ' tends to increase with π_2 as well. This suggests that increasing the detour bound too much could be counterproductive for society.

A similar simulation is done to verify (16)-(17) as well. The objective is still to maximize the total vehicle distance savings. The only difference is that every user is treated both as a potential rider (i.e., seeking possible pickup from a driver) and as a potential driver. Hence the parameter f is no longer needed. In this case, the optimization model reduces to a minimum-weight matching problem, which still is a linear program and can still be solved by *Gurobi's* package.

Fig. 4 plots the simulated and predicted r, δ and δ' values under similar parameters to those in Fig. 3. While these curves largely follow the same trends, they are generally 3-4 times the values of their counterparts in Fig. 3. This indicates, as expected, that the relaxation of fixed user roles (either as a driver or a rider) significantly increases the likelihood of a successful match.

6. Example

The formulas presented in this paper can help planners anticipate how a profit-maximizing TNC responds to regulations and monetary incentives. City planners could then systematically formulate policies that avoid undesirable surprises; e.g., with a bi-level optimization model. In this type of approach, a lower-level model would predict what the TNC does and – with this model as a constraint – an upper-level model would optimize the city's policy. However, for this (or any other systematic) approach to be feasible, predictive formulas such as those presented in this paper are needed.

To illustrate the idea in a transparent and concrete way, this section presents a simple toy example that contains a minimum set of key modeling elements and relevant formulas. We shall assume that the TNC uses the setup of Sections 2 and 3 with distinct pools of drivers and riders. In practice, the example should be modified to reflect the characteristics of the particular situation at hand.

The example's lower-level model is described first. We start with the TNC's revenues and outlays, which are used to formulate its profit. The framework is as follows. We assume that the TNC charges each successfully matched rider a fee proportional to its trip distance (the rider-kilometers traveled, RKT) at the rate of P [$/km], and then compensates the servicing driver at a lower rate Q [$/km] using the same RKT. The TNC also receives performance-based incentives (or penalties) from the city in which it operates. In this example, we shall assume that the city primarily aims at taking cars off the road, and hence subsidizes the TNC at the rate of s [$/km] for each unit of VKT saved.[9] Thus, the TNC's profit per successful match is:

$$[\text{TNC profit per match}] = (P - Q)[\text{RKT}] + s[\text{VKT saved}].$$

In this formula, the distances in brackets are endogenous variables. The variables s, P, and Q are decision variables chosen by the city (s) and the TNC (P, Q). The TNC also chooses the time and detour guarantees, π_1 and π_2.

The average profit per carpool user can now be expressed by taking the expectation of the above formula, and expressing the result in terms of the variables δ and δ' introduced in Section 2.3 (with the corrections in Section 3, as needed). The result reduces to:

$$[\text{average TNC profit per user}] = (P - Q + s)\delta + (P - Q)\delta'.$$

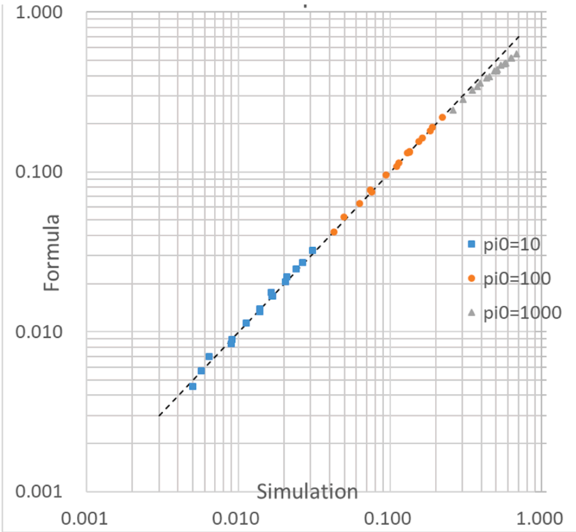
To estimate δ and δ' we need π_0, f and g, see e.g., (11); or equivalently the number of users of both types (riders, R, and drivers, D) that sign up to play.[10] Because these numbers depend on the TNC's fee structure, two demand models must be used to estimate them. To keep things simple we shall assume that the model for D is a truncated linear function of the compensation rate Q and the detour limit π_2; i.e., that$D = \max\{0, -a_1 + a_2Q - a_3\pi_2\}$. For the same reason, we also assume that R is a truncated linear function of the fare rate P and the wait limit π_1; i.e., that$R = \max\{0, b_1 - b_2P - b_3\pi_1\}$. It is reasonable to expect that a_i,$b_i > 0$ for $i = 1, 2, 3$.[11]

We are now ready to write the lower level (TNC profit maximization) problem. The decision variables are (π_1,π_2,P, Q). The objective

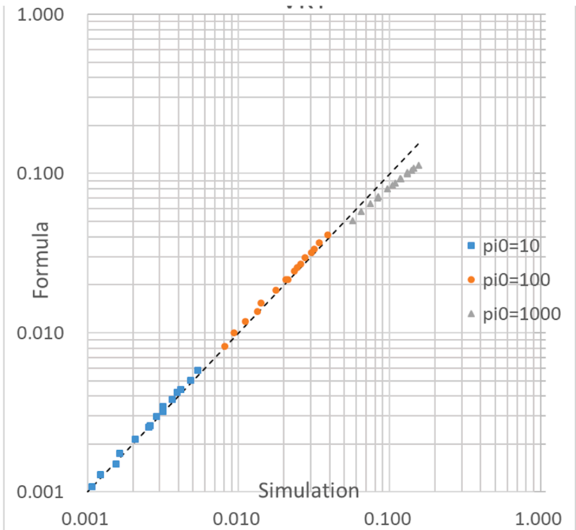
[9] The TNC would have to share (anonymously if necessary) the users' origins and destinations for each matched trip; e.g., through an open-source database and API.

[10] Note that per our setup,$f = R/(R + D)$, $g = D/(R+D)$, and $\pi_0 = R + D$.

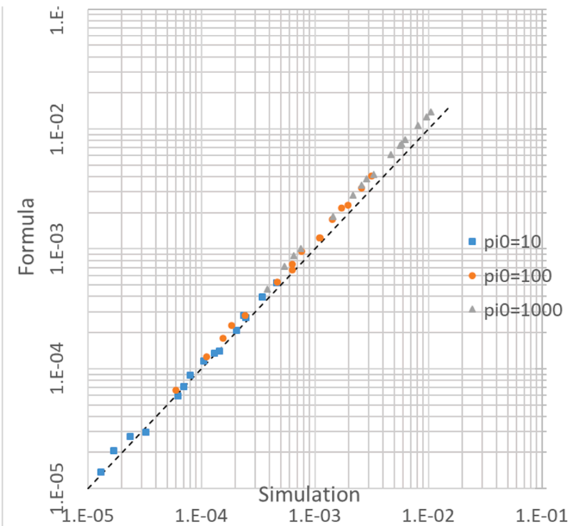
[11] It is reasonable to expect the ratios a_3/a_2 and b_3/b_2 to be roughly equal to the value (willingness to pay) for a trip from the driver and rider perspectives.



(a) Probability of successful matches, r



(b) Expected VKT saving per trip



(c) Expected PKT added per trip

(caption on next page)

Fig. 4. Comparison of performance metrics from matching algorithm and formulas under flexible driver/riders.

function (total profit) is the product of the average profit per user given above, and (R+D). Constraints include the two demand functions and some definitional equalities linking all the auxiliary decision variables. It is as follows:

$$\mathrm{Max}[(P-Q+s)\delta+(P-Q)\delta'](R+D) \tag{18}$$

$$\mathrm{S.t.}\, D=\max\{0,-a_1+a_2Q-a_3\pi_2\}$$

$$R=\max\{0,b_1-b_2P-b_3\pi_1\}$$

$$f=R/(R+D), g=D/(R+D)$$

$$\pi_0=R+D$$

δ, δ′ from (1)–(17)

$$\pi_1,\pi_2,P,Q\geq 0.$$

Note, the only exogenous data for this problem are s (which is chose by the city) and the (a_i, b_i) parameters of the demand function. City planners can make educated guesses for these parameters and/or estimate them from surveys. Thus, although the city planners may not know exactly how a TNC may respond to incentives, they have the rough data required to solve (18). Planners can therefore use (18) as an approximate "reverse engineered" way of anticipating the TNC's response to a planned incentive.

Planners can also use (18) to search for optimal incentives; e.g., by including it as a constraint in an upper-level policy optimization model. In our particular case, the upper-level problem is very simple because we only have one policy instrument (the variable s) but our method also applies to multi-instrument scenarios.

In our particular case, we have assumed that the city wishes to encourage VKT reductions. Therefore we assume that its policy objective is maximizing a weighted difference between the VKT savings, $(R+D)\delta$, and the total subsidy that must be paid, $(R+D)s\delta$. Furthermore, we choose the VKT weight to be $(\frac{b_1}{2b_2})$ because this value approximately equals the average "value of a trip" among the riders, although in an actual application the choice should be made in consultation with the city managers. Given all this, the upper-level problem is:

$$\mathrm{Max}\left(\frac{b_1}{2b_2}-s\right)(R+D)\delta \tag{19}$$

s.t. π_1,π_2,P, Q, D, R, f, g solve (18).

The lower-level TNC problem (18) and the upper-level city's problem (19) are both nonlinear. Hence, the bi-level problem could be difficult to solve. Fortunately, the problem only has six decision variables in the lower level and one in the upper level. Hence, the problem at each level can be solved with an exhaustive search algorithm.

To demonstrate the feasibility of this approach, we consider a city that spans a unit square with the following dimensionless parameters: $a_1=40$, $a_2=4$, $a_3=200$, $b_1=120$, $b_2=2$, $b_3=200$. The parameters were derived by considering a medium sized city like San Francisco. As described in Section 5, when the price P and compensation Q are on the order of \$30, the trip time across our city takes about an hour (under moderate congestion), and when the detour limits π_1 and π_2 are on the order of 0.1, the value of π_0=R+D is on the order of 100. Despite the complexity of some of the component formulas (1)-(17), an exhaustive search algorithm found the optimum in a few minutes. The optimal decision variables are listed in Table 1 below.

The fifth row from the bottom of the table shows that the TNC should be incentivized at rate $s=7$. This rate would ensure that the city achieves its societal goal of reducing VKT, taking into account the TNC's adaptation to the incentive. As a point of interest, the method also returns the values of r, δ and δ', which are auxiliary decision variables. For our example, these performance measures turn out to be: 8.49%, 1.38%, and 0.31%, respectively. For comparison, the last column of the table also shows the do-nothing scenario when the city provides zero subsidy. Note how the TNC's decisions especially the rider price and driver fee are influenced by the subsidy. When the subsidy is zero, the corresponding values of r, δ and δ' are 7.46%, 1.19%, and 0.29%. Note these numbers are averages per trip; the presence of optimal subsidy is shown to notably change the average carpooling system performance.

7. Discussion and conclusion

This paper presented an analytical model in closed form to explore the impacts of detour bounds on the feasibility and performance of reservation-based carpooling services in idealized settings (e.g., square city, homogeneous demand, and dense city streets). We studied two scenarios, one with distinct drivers (who are unwilling to take a ride) and riders (who cannot drive), and the other with users who can both drive and ride. Formulas were provided for the percent of rides that can be matched, as well as for the vehicle distance savings, and passenger distance increase as compared with individual trips. These closed-form results were verified via matching algorithms in both scenarios.

The proposed modeling framework could be used by TNCs to quantify the impacts of service guarantees as well as the size of driver/

Table 1
Optimal solution to the system planning example.

Decision maker	Decision variable	Optimal value	Benchmark
TNC	Unit rider price: P	36	38
	Unit driver fee: Q	30	28
	# of driver: D	46	36
	# of rider: R	22	18
	Wait guarantee: π_1	0.13	0.13
	Detour guarantee: π_2	0.17	0.18
	Objective value	13.43	7.98
City	VKT subsidy: s	7	0
	Matching rate	8.49%	7.46%
	VKT saving (%)	1.38%	1.19%
	PKT increase (%)	0.31%	0.29%
	Objective value	21.54	19.23

rider pools on system operation efficiency. The closed form formulas can be used to design economic incentives (e.g., rider pricing and driver compensation) that can generate the best service quality or yield the maximum profit. The model is also useful to urban policymakers who can estimate beforehand the VKT and PKT outcomes of potential regulations. The simple example in Section 6 shows how the derived formulas can facilitate efforts in this direction.

Future research can be conducted in several directions. This paper only studies reservation-based carpooling services, but not the real-time counterparts where users request service on the go. Analysis of service guarantees for spontaneous carpooling performance should be similar to the analysis just performed (e.g., much as the analysis in Daganzo et al. (2020) for spontaneous ride matching service with third-party chauffeurs was similar to the corresponding reservation service). The simulation experiments in Section 5 show how well the derived closed-form formulas predict performance of the carpooling strategies in the idealized settings assumed in this paper. These tests, however, are limited to the strategies and scenarios used. Clearly, different assumptions on city shape, the demand distribution and street topology may alter the results. We suspect, however, that if these assumptions do not change radically, the results will not either. This, of course should be verified in the future, hopefully with real-world data and in collaboration with TNCs. Such a collaboration would also allow researchers to determine if more sophisticated operational rules should be investigated. Section 6 only presents a toy example to illustrate how the closed-form formulas from this paper could be used by carpooling TNCs or by cities for policy decisions. In the future, it will be interesting to explore additional ways to use the results of this paper for system design, resource allocation, as well as policy making. Finally, it may be of interest to explore how the formulas change if the users are socioeconomically heterogeneous, or if the detour and waiting time guarantees are relative (e.g., proportional to the users' trip lengths).

CRediT authorship contribution statement

Yanfeng Ouyang: Conceptualization, Methodology, Writing – original draft. **Haolin Yang:** Methodology, Data curation, Writing – original draft. **Carlos F. Daganzo:** Conceptualization, Methodology, Writing – review & editing.

Acknowledgment

The very helpful comments from the editors and three anonymous reviewers are gratefully acknowledged. This research was supported in part by the U.S. National Science Foundation via Grant CMMI-1662825.

Appendix A. Derivation of (3) and (4)

This appendix briefly summarizes the derivation of Eqs. (3) and (4). The key step is to identify, for any given X and Y, the formulas for L and L' in all possible combinations of (x, y) and (x_B,y_B) ranges, as shown below.

Range of (x, y)	Range of (x_B,y_B)	L	L'
$0 < x < X$, $0 < y < Y$, $c(x,y) = \frac{\pi_2}{2}$	$x < x_B < X$, $y < y_B < Y$	$x_B - x + y_B - y$	0
	$x - c(x, y) < x_B < x$, $y < y_B < Y$	$x_B - x + y_B - y$	0
	$X < x_B < X + c(x, y)$, $y < y_B < Y$	$x_B - x + y_B - y - 2(x_B - X)$	$2(x_B - X)$
	$x < x_B < X$, $y - c(x, y) < y_B < y$	$x_B - x + y_B - y$	0
	$x < x_B < X$, $Y < y_B < Y + c(x, y)$	$x_B - x + y_B - y - 2(y_B - Y)$	$2(y_B - Y)$

(continued on next page)

(*continued*)

$-\frac{\pi_2}{2} < x < 0$, $0 < y < Y$, $c(x,y) = \frac{\pi_2}{2} + x$	$x < x_B < X$, $y < y_B < Y$	$x_B + x + y_B - y$	$-2x$
	$x - c(x, y) < x_B < x$, $y < y_B < Y$	$x_B + x + y_B - y$	$-2x$
	$X < x_B < X + c(x, y)$, $y < y_B < Y$	$x_B + x + y_B - y - 2(x_B - X)$	$-2x + 2(x_B - X)$
	$x < x_B < X$, $y - c(x, y) < y_B < y$	$x_B + x + y_B - y$	$-2x$
	$x < x_B < X$, $Y < y_B < Y + c(x, y)$	$x_B + x + y_B - y - 2(y_B - Y)$	$-2x + 2(y_B - Y)$
$0 < x < X$, $-\frac{\pi_2}{2} < y < 0$, $c(x,y) = \frac{\pi_2}{2} + y$	$x < x_B < X$, $y < y_B < Y$	$x_B - x + y_B + y$	$-2y$
	$x - c(x, y) < x_B < x$, $y < y_B < Y$	$x_B - x + y_B + y$	$-2y$
	$X < x_B < X + c(x, y)$, $y < y_B < Y$	$x_B - x + y_B + y - 2(x_B - X)$	$-2y + 2(x_B - X)$
	$x < x_B < X$, $y - c(x, y) < y_B < y$	$x_B - x + y_B + y$	$-2y$
	$x < x_B < X$, $Y < y_B < Y + c(x, y)$	$x_B - x + y_B + y - 2(y_B - Y)$	$-2y + 2(y_B - Y)$
$X < x < X + \frac{\pi_2}{2}$, $0 < y < Y$, $c(x,y) = \frac{\pi_2}{2} - (x - X)$	$X < x_B < x$, $y < y_B < Y$	$x - x_B + y_B - y - 2(x - X)$	$2(x - X)$
	$x < x_B < x + c(x, y)$, $y < y_B < Y$	$x - x_B + y_B - y - 2(x - X)$	$2(x - X) + 2(x_B - x)$
	$X - c(x, y) < x_B < X$, $y < y_B < Y$	$x - x_B + y_B - y - 2(x - X) - 2(X - x_B)$	$2(x - X) + 2(X - x_B)$
	$X < x_B < x$, $y - c(x, y) < y_B < y$	$x - x_B + y_B - y - 2(x - X)$	$2(x - X)$
	$X < x_B < x$, $Y < y_B < Y + c(x, y)$	$x - x_B + y_B - y - 2(y_B - Y) - 2(x - X)$	$2(x - X) + 2(y_B - Y)$
$0 < x < X$, $Y < y < Y + \frac{\pi_2}{2}$, $c(x,y) = \frac{\pi_2}{2} - (y - Y)$	$x < x_B < X$, $Y < y_B < y$	$x_B - x + y - y_B - 2(y - Y)$	$2(y - Y)$
	$x - c(x, y) < x_B < x$, $Y < y_B < y$	$x_B - x + y - y_B - 2(y - Y)$	$2(y - Y)$
	$X < x_B < X + c(x, y)$, $Y < y_B < y$	$x_B - x + y - y_B - 2(x_B - X) - 2(y - Y)$	$2(y - Y) + 2(x_B - X)$
	$x < x_B < X$, $y < y_B < y + c(x, y)$	$x_B - x + y - y_B - 2(y - Y)$	$2(y - Y) + 2(y_B - y)$
	$x < x_B < X$, $Y - c(x, y) < y_B < Y$	$x_B - x + y - y_B - 2(y - Y) - 2(Y - y_B)$	$2(y - Y) + 2(Y - y_B)$
All other possibilities		0	0

Next, take expectation of L and L' with respect to (x_B, y_B) and (x, y), which are uniformly distributed within each combination above. The results are the total VKT saving and PKT increase per driver (successfully matched or not). Algebraic manipulation shows that the results, respectively, are

$$\frac{1}{48}\left(4X^2Y^2(X+Y) + 12\pi_2XY(X+Y)^2 + 3\pi_2^2\left(2X^3+7X^2Y+7XY^2+2Y^3\right) + \pi_2^3\left(5X^2+8XY+5Y^2\right) + \pi_2^4(X+Y)\right)$$

$$= \frac{X+Y}{3}N(X,Y) + \frac{1}{48}\pi_2^2\left(7X^2Y+7XY^2\right) + \frac{1}{144}\pi_2^3\left(11X^2+16XY+11Y^2\right) + \frac{1}{48}\pi_2^4(X+Y)$$

and

$\frac{\pi_2^2}{48}\left(18XY(X+Y) + \pi_2(12X^2+16XY+12Y^2) + 3\pi_2^2(X+Y)\right)$.

Then dividing them by the expected number of successfully matched drivers, $N(X, Y)$, we can get Eqs. (3) and (4).

Appendix B. Match feasibility revisited – rider perspective

In Section 2.1, we choose to focus on a generic driver whose trip is characterized by random variables X and Y. Below we derive the counterparts of (7a) and (7b) for an arbitrary rider. The derivations are somewhat different because what a rider requires from a driver is different from the reverse.[12]

The rider's origin and destination are labeled A and B, as in Fig. A1(a). The derivation is first, first when the distance between points

[12] As we saw in Section 2.1, a driver's trip (X, Y) is equally easy to match independent of where in the region the trip takes place. Yet, common sense reveals that the same does not hold for riders. Because the trip of a matched driver must entirely "encompass" that of the rider, a centrally located rider trip is easier to match than a similar trip near the region's edge. As a result, we cannot use the same results as in Section 2.1, and we must consider both the user's origin and destination locations.

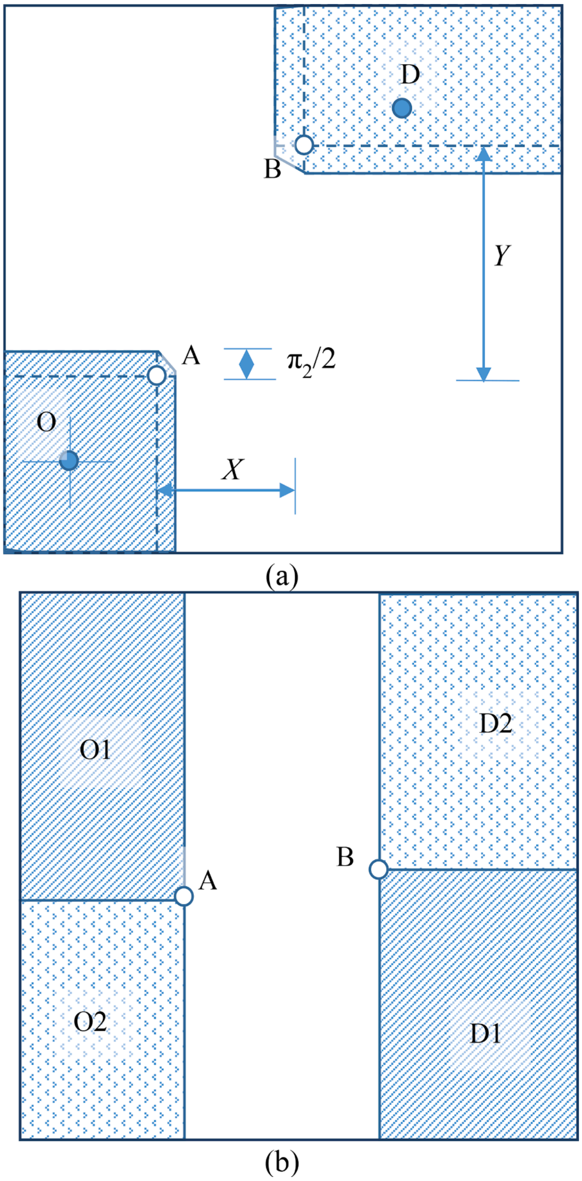


Fig. A1. Origin- and destination-feasible catchment areas (not drawn to proportion) from a rider's perspective.

A and B is larger than $\pi_2/2$. We then sample a random driver with origin O and destination D. For this driver to be origin-feasible, the locus of feasible origins is shown by the shaded polygon near A, which includes a rectangle extended by A and the left-bottom corner of the city, plus the periphery with a width of $\pi_2/2$. To be destination-feasible, the driver's destination D must be in the shaded polygon near point B, which includes the rectangle extended by B and the upper-right corner of the city, plus the periphery. The thickness of the periphery now depends on the location of point O and the amount of detour distance that has occurred before the rider's pickup.

Analysis shall be very similar to that of Section 2.1. A coordinate system can be set up at the left-bottom corner of the city, while the coordinates of points A and B are still marked by (x, y) and (x_B, y_B) respectively. For convenience, the lengths of the rider's trip in horizontal and vertical directions are still denoted by variables: $X = x_B - x$, $Y = y_B - y$. The thickness of destination feasible area equals $\pi_2/2$ if point O is inside the bottom-left rectangle, or reduced otherwise.

When either the horizontal or the vertical distance between points A and B is small (e.g., smaller than $\pi_2/2$), delivering the rider even in the opposite direction does not impose a violation to the detour limit. Hence, more drivers can be feasible, and the additional catchment areas are shown in Fig. A1(b).

Combining all the above, it can be shown that the conditional probability that a driver is both origin- and destination-feasible, given (x, y, X, Y), is as follows:

$$M(x,y,X,Y) = xy(X+x-1)(Y+y-1) + [x(X+x-1)(Y+y-1) + y(X+x-1)(Y+y-1) - xy(X+x-1) - xy(Y+y-1)]\frac{\pi_2}{2}$$
$$- (x+y)(2-X-Y-x-y)\pi_2^2 \Big/ 8.$$

Clearly $M(x, y, X, Y)$is random. We know that X and Y are mutually independent and identically follow the symmetric triangular distribution on [0, 1], while x and y are uniformly distributed on $[0, 1-X]$ and $[0, 1-Y]$ respectively. Algebraic manipulation shows that:

$$E[M(x,y,X,Y)] = \frac{1}{144} + \frac{\pi_2}{12} + O(\pi_2^2).$$

$$Var[M(x,y,X,Y)] = \frac{7}{186624} + \frac{7\pi_2}{4320} + O(\pi_2^2).$$

The expected number of temporally valid drivers is now $(1-f)\,\pi_0\pi_1$. For notation convenience, we let g=1 $-f$. Hence, the number of spatiotemporally feasible drivers that a rider expects to see is $(1-f)\pi_0\pi_1 M(x, y, x_B, y_B)$, and its mean m and variance ψ are given as (12a) and (12b).

References

Bliss, L. (2018). "To measure the Uber effect cities get creative." CityLab, posted on January 12, 2018.

Chan, N.D., Shaheen, S.A., 2012. Ridesharing in North America: past, present, and future. Transp. Rev. 32, 93–112.

Daganzo, C.F., Ouyang, Y., 2019. Public Transportation Systems: Principles of System Design, Operations Planning and Real-Time Control. World Scientific Publishing Company, Singapore.

Daganzo, C.F., Ouyang, Y., Yang, H., 2020. Analysis of ride-sharing with service time and detour guarantees. Transp. Res. Part B 140, 130–150, 2020.

Furuhata, M., Dessouky, M., Ordonez, F., Brunet, M.E., Wang, X., Koenig, S., 2013. Ridesharing: the state-of-the-art and future directions. Transp. Res. Part B 57, 28–46.

Ghilas, V., Demir, E., Van Woensel, T., 2016. The pickup and delivery problem with time windows and scheduled lines. INFOR 54 (2), 147–167.

Graziotin, D. (2013). "An analysis of issues against the adoption of dynamic carpooling." arXiv:1306.0361 [cs.CY]; 10.6084/m9.figshare.710636.

Hosni, H., Naoum-Sawaya, J., Artail, H., 2014. The shared-taxi problem: formulation and solution methods. Transp. Res. Part B 70, 303–318.

Johnson, W.C. Mathematical and Physical Principles of Engineering Analysis. New York McGraw-Hill, 1944.

Ke, J., Yang, H., Li, X., Wang, H., Ye, J., 2020. Pricing and equilibrium in on-demand ride-pooling markets. Transp. Res. Part B 139, 411–431.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2015. The benefits of meeting points in ride-sharing systems. Transp. Res. Part B 82, 36–53.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2016. Making dynamic ride-sharing work: the impact of driver and rider flexibility. Transp. Res. Part E 91, 190–207.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2018. Enhancing urban mobility: integrating ride-sharing and public transit. Comput. Oper. Res. 90, 12–21.

Yang, H., Qin, X., Ke, J., Ye, J., 2020. Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. Transp. Res. Part B 131, 84–105.

Zhang, K., Nie, M., 2019. To pool or not to pool: Equilibrium, pricing and regulation. Pricing and Regulation (December 3, 2019).

Zhang, K., & Nie, M. (2020). Inter-platform competition in a regulated ride-hail market with pooling. Available at SSRN 3721402.