Sequential Individual Rationality in Dynamic Ridesharing

Ragavendran Gopalakrishnan
Cornell University, rg584@cornell.edu
https://people.orie.cornell.edu/ragad3/

Theja Tulabandhula
University of Illinois Chicago, tt@theja.org

Koyel Mukherjee
IBM Research India, koyelmjee@gmail.com

In dynamic ridesharing systems, both operational policies (e.g., ride-matching) and economic policies (e.g., pricing or cost sharing) impact the Quality of Service (QoS) perceived by users. Recent field experiments have found that firms benefit from *proactively* compensating users whose QoS expectations are violated. This motivates a broader analytical study of how behavioral perceptions of QoS impact operational and economic policy design in ridesharing systems. We introduce a novel, QoS-centric framework consisting of the following key elements: (a) users' *state-dependent* utility model that bridges operational effects (detours) and economic effects (prices or cost shares), and serves as the input to a choice model, (b) dynamic notion of QoS, called *sequential individual rationality*, defined on the *sequence* of (dis)utilities from successive stages of a shared ride, that is guided by appropriate behavioral drivers such as reference effect, loss aversion, and recency effect, and (c) formulation of QoS-sensitive economic objectives (profit or fairness) by endogenizing users' choices and QoS constraints. Our framework can be used to extract key operational insights from QoS-sensitive economic objectives, as illustrated in two different ridesharing environments: (i) commercial ridesharing (real-time), which involves pricing exclusive and shared service in order to maximize profit (taking into account penalties for QoS-violations), and (ii) community carpooling (static and dynamic), which involves designing fair cost sharing schemes.

In the commercial setting, we characterize a ride's *optimal shareable region*, and show, perhaps surprisingly, that it may be optimal for a QoS-sensitive service provider to violate QoS and suffer an associated penalty, no matter how strong the users' loss aversion. In the carpooling setting, we characterize routes that admit (nonnegative) budget-balanced, QoS-compliant cost sharing schemes, resulting in a ride's *QoS-compliant shareable region*. We also define *sequential fairness* and characterize a family of fair, QoS-compliant cost sharing schemes that bring out insightful structural properties, including a surprisingly strong requirement that commuters *must* compensate each other for the detour-inconveniences they cause.

Key words: ridesharing; carpooling; pricing; cost sharing; individual rationality; fairness; quality of service; behavioral operations

---

1. Introduction. Urban transportation is facing a host of urgent challenges. The [62] projects that by 2050, 68% of the world's population will live in urban cities (compared with 55% today), and that by 2030, there will be 43 "megacities" whose population exceeds 10 million (compared with 33 today). The same report highlights that *sustainable* urbanization is key to successful economic, social, and environmental development. Ridesharing has emerged as a popular solution that aims to combat ever-increasing congestion along road networks around the world. The potential decrease in the number of Vehicle-Miles Travelled (VMT) could significantly reduce carbon emissions, making ridesharing all the more desirable from a sustainability perspective. The term "ridesharing", in both popular culture and academic literature, has become a buzzword that refers to any *ride-booking* or *ride-hailing* service such as Lyft and Uber, even if there is only one user taking the ride and there is no *demand* sharing involved.[1] Throughout this paper, we use this term to denote only those settings in which two or more users share rides.

[1] The Associated Press has criticized this abuse of the term [21].

Two broad settings facilitate urban ridesharing: In *commercial* ridesharing, on-demand service providers such as Lyft and Uber offer "pooled" versions of their ride-hailing services, (e.g., LyftLine and UberPool), in real time [10]. On the other hand, *community* carpooling programs, e.g., those run by medium to large organizations, encourage groups of regular commuters to/from a common location (such as the workplace) to travel together, using either their personal vehicles or those provided by the company or a third-party contractor. Community carpooling has traditionally been a *static* problem, where the carpooling groups and schedules are determined ahead of time and remain unchanged over long periods [14]. However, *dynamic*, real-time carpooling solutions can be more efficient and increase participation levels [35]. Ridesharing systems are complex, and consist of several key elements which can be classified as follows:

(a) *Operational* modules include (static) group formation, (dynamic) matching, routing, and fleet/supply management.
(b) *Economic* modules include pricing (commercial ridesharing) or cost sharing (peer-to-peer or community carpooling).
(c) *Behavioral* modules include modeling user choice/response and service quality.

Existing literature on designing ridesharing systems treats operational objectives (e.g., minimizing VMT) and economic objectives (e.g., profit/welfare maximization, fair cost sharing) completely independently of each other [41, 25]. Moreover, when modeling Quality-of-Service (QoS) constraints, the focus has largely been on operational measures such as fixed detours [28, 49], or abstract measures such as ride quality ratings [54]. The impact of economic QoS measures such as (ex-post) individual rationality, as well as behavioral factors, on ridesharing operations is only beginning to be understood in experimental and empirical work [15, 43].

We address this gap by proposing a novel, layered analytical framework for designing ridesharing systems, in which the *behavioral* elements, especially QoS (Section 3), play the focal role. In particular, the users' utility and choice functions, and appropriate QoS notions are modeled first, building upon which the *economic* objectives are laid out. Subsequent analysis then yields insightful characterizations that translate to *operational* constraints, which can be passed on to any existing operational optimization framework. *To our knowledge, we are the first to investigate, analytically, the consequences of behavior-integrated, QoS-sensitive economic objectives on the operations of ridesharing systems:*

(a) **Commercial Ridesharing:** In Section 4, we focus on a real-time setting in which a commercial service provider offers both exclusive and shared rides, and must set the corresponding prices depending on the estimated additional delay/inconvenience for a shared ride. We begin with a random utility model for users, a discrete choice model on these utilities, and a notion of QoS that incorporates reference effects and loss aversion into the traditional concept of ex-post individual rationality. A QoS-sensitive service provider's profit optimization problem is then formulated by internalizing the users' choices, taking into account any penalties for QoS violations. In Theorem 1, we characterize the *optimal shareable region* for a shareable ride, wherein the optimal prices induce a nonzero probability of ridesharing. These spatial limits help prune the feasible space for any system-wide operational optimization problem. We investigate the dependence of the shareable region on the QoS-sensitivity of the service provider and the degree of loss aversion of the user. Theorem 2 shows, perhaps surprisingly, that it may be optimal for the service provider to violate QoS and suffer the associated penalty, no matter how strong the QoS-sensitivity. Finally, we provide closed form expressions for the optimal prices in Theorem 3.
(b) **Community Carpooling:** In Section 5, we consider both static and dynamic scenarios of community carpooling, wherein a group of commuters (from within a larger pool of participants) carpool together to a common destination and share the operational cost among themselves. Here, we begin with a *disutility* model for users, a static notion of QoS, and a stronger, dynamic

notion of QoS that incorporates recency effects into the traditional concept of individual rationality. In Theorems 4-6, we characterize routes (sequences of pickup locations) that admit (nonnegative) budget-balanced, QoS-compliant cost sharing schemes. In the dynamic scenario, this characterization defines a *QoS-compliant shareable region* for an existing shareable ride, wherein a new commuter can be feasibly accommodated. Theorems 7-10 then furnish constant and sublinear bounds on the worst-case QoS-compliant detour for homogeneous carpooling commuters. Finally, we introduce a dynamic notion of fairness and characterize a family of fair, QoS-compliant cost sharing schemes in Theorem 11. Our characterization exposes several practical structural properties of such schemes, including a surprisingly strong requirement that commuters *must* compensate each other for the detour-inconveniences they cause.

Finally, we conclude in Section 6 with a discussion on how our framework could potentially be generalized to be more broadly applicable to dynamic shared service systems in which the quality of shared service, as perceived by customers based on their experience (e.g., waiting in queues, delays/interruptions during service), plays a central role in designing operational and economic policies (e.g., matching/routing, staffing, pricing).

2. Related literature. Our work presents a framework for understanding the consequences of behavior-integrated, QoS-centric economic objectives on designing the operational policies of ridesharing systems, in both commercial (pricing) and peer-to-peer or community carpooling (cost sharing) contexts. As such, this contribution sits within the intersection of several related research streams.

There is a vast literature on dynamic vehicle routing [23, 16, 2, 50] that focuses on operational optimization problems in ridesharing systems. Our work can be thought of as feeding additional operational constraints into these problems that ensure that the resulting operational policies are compatible with optimal QoS-aware economic policies.

The pricing-related literature on commercial “ridesharing” is limited to ride-hailing settings [6, 8, 7, 59], that is, the pricing of pooled or shared rides is not considered. The only exception that we are aware of is the work of [25], which is close in spirit to our analysis in the commercial setting. Here, the authors characterize the optimal pricing policy of a service provider as a function of the demand rate, delay-sensitivity of users waiting for service, and their inconvenience costs due to ridesharing. While we focus only on modeling the detour-inconvenience from shared service and demand-side behavior, their model of disutility integrates the (non-detour-related) inconvenience effects from both waiting and shared service, and they consider supply-side (driver) behavior as well. However, this complexity limits their analysis to a simplistic scenario with a maximum capacity of 2, and a single common source and destination, thus excluding the possibility of detours.

Carpooling programs, especially those run by employers, have been studied for decades [20, 40]. The cost sharing problem in this context has garnered relatively little attention—in most existing schemes, individual passengers are asked to post what they are willing to pay in advance [12], or share the total cost proportionately according to the distances travelled [22, 3]. Such methods ignore the real-time costs and delays incurred during the ride (as in the first instance), or are insensitive to the disproportionate delays encountered during the ride (as in the second instance).

Recent work has studied cost sharing when passengers have significant autonomy in choosing rides or forming carpooling groups, e.g., cost sharing schemes based on the concept of kernel in cooperative game theory [9], second-price auction based solutions [31], and market based ride-matching models with deficit control [65]. Fair cost sharing in ridesharing has also been studied under a mechanism design framework by [29], where an individually rational VCG-based payment scheme is modified to recover budget-balance at the cost of incentive compatibility, and by [41], where customers are offered an additive, detour-based discount, and the allocations and pricing are determined through an auction. Our work differs from all the above in that our framework is *intentionally* agnostic to the operational/mechanistic aspects of the system such as ride-matching or group formation. Moreover, all these works lack any integration of QoS in a real-time or dynamic setting that is motivated from a behavioral standpoint.

While there exist some previous works on ridesharing that simultaneously address individual rationality and detour limits [29, 57, 49], they treat them as *independent* constraints. In contrast, operational constraints such as detour limits are a *consequence* of our QoS-centric framework.

Variations of individual rationality (IR) involving informational aspects are well studied in the economics literature, e.g., *ex-ante*, *interim*, and *ex-post* IR in mechanism design [39], and sequential IR in bargaining and repeated games [19]. [15] find that offering *proactive* compensation to users for ex-post IR violations increases a firm's net profit using field experiments in commercial ridesharing. However, we are the first to analytically adapt IR for dynamic ridesharing by infusing behavioral considerations, and subsequently build a companion notion of fairness.

An extensive literature on cooperative game theory and fair division [37, 27] offers various cost sharing schemes that can be analyzed in our framework in the static carpooling setting. Our view of fairness in the dynamic setting relies on how the total *incremental* benefit due to ridesharing is allocated among the commuters at each stage of the ride. We believe the two approaches are quite different, but not independent; we defer a discussion regarding possible connections to our concluding remarks in Section 6.

Modeling the quality of service in ridesharing is complex; see Section 3.1.2 and Table 5 in [47] for a survey of operational metrics used in the Operations Management literature. [36] study the impact of operational quality measures on the ridesharing system's costs; however, our approach to modeling QoS is utilitarian and behavioral (see Section 3). [32] emphasize the importance of considering user satisfaction as the primary objective in designing operational policies of ridesharing systems, and use deep learning techniques to model user satisfaction from data.

Finally, we survey literature that motivates the behavioral approach to modeling QoS when service is experienced *sequentially* [18]. [55] analyze how customers react to a sequence of services/service-levels in an experiential setting. In our dynamic ridesharing scenario, these "service-level epochs" correspond to additional detours resulting from subsequent addition of users to an existing ride. In the commercial context, the service provider informs the user in advance the estimated additional delay due to ridesharing, which triggers benchmark/reference effects, and subsequent loss aversion [60] with respect to the announced benchmark. Such behavior among delay-sensitive customers is well-studied in the queueing literature [64]. In the carpooling context, absent an "anchor" estimate at the time of joining a ride, commuters compare the impact of a future disruption (detour due to the addition of a new commuter) against their most recently updated disutility. Behavioral models that support this assumption include memory decay [17, 34] and recency effects [30].

3. Modeling Quality-of-Service. In this section, we briefly explain our approach to modeling Quality-of-Service (QoS), which is central to achieving the goals of our framework. There are two important aspects:

(a) **Utilitarian:** We choose to model QoS as a property on the *(dis)utilities* of the users. This is critical, because a user's utility function bridges the ridesharing system's operational module (by incorporating the effect of detours), the economic module (by incorporating the effect of prices or cost shares), and the behavioral module (by serving as the basis of the choice model). This is why our framework is so effective in extracting operational insights from analyzing QoS-aware economic objectives.

(b) **Behavioral:** In a dynamic ridesharing system, users experience a *sequence* of (dis)utilities at every stage of a shared ride (due to additional detours from new users and/or updated cost shares). It is not apriori clear how traditional notions of utilitarian QoS concepts such as (ex-post) individual rationality apply in such settings. Thus, we turn to literature that studies appropriate models of human behavior when service is experienced sequentially absent a benchmark, e.g., [18, 55], in the carpooling context, and when the service provider announces a pre-specified service-level estimate, e.g., [64], in the commercial context.

Guided by these factors, we adapt the concept of individual rationality (IR) for dynamic shared service systems, and name it *Sequential Individual Rationality* (SIR), which, broadly speaking, requires some form of IR to hold at every stage of the shared service experience. The exact notion that we adopt depends on the appropriate behavioral justification:

(a) In the *commercial ridesharing* context (Section 4.1.5), we assume that the benchmark/reference effect is the dominating behavioral driver, along with some degree of loss aversion. Thus, we define SIR as requiring *ex-post* IR to hold at every stage of the ride. To be precise, every time a new user joins the ride, the updated utility (based on the increased detour) of the existing users must not fall below that of their original alternative (based on which they made their choices to opt for a shared ride). Loss aversion is taken into account by appropriately scaling the penalty that the service provider suffers for violating SIR.

(b) In the *community carpooling* context (Section 5.1.3), since there is no explicit benchmark, we assume that the recency effect is the dominating behavioral driver, and define SIR as requiring the sequence of disutilities experienced to be nonincreasing. That is, every stage of the ride should be IR with respect to the previous (most recent) stage.

We acknowledge that there are certainly other ways of modeling QoS when guided by utilitarian and behavioral considerations, and we invite future research to explore appropriate alternative models in ridesharing or other dynamic shared service systems.

4. QoS-Aware profit maximization in commercial ridesharing. We consider a setting where a commercial ridesharing service provider seeks to design pricing and ride-matching policies that maximize profit on a per-ride basis. In our model, we focus on the revenue and costs associated with *active* users, by considering only the time they spend in service (inside the vehicle). This focus simplifies the model by removing the dependence on the availability of vehicles and fleet (supply) management controls; these constraints can be considered within a system-wide operational optimization problem. Perhaps more importantly, it allows us to isolate the effect of purely service-related QoS (e.g., the effects of detours on active users) on operational policy design.

Users interact with the service provider through an interface on a mobile device, similarly to popular ridesharing services. The interaction consists of the following two stages:

(a) **Stage 1:** User j inputs their source coordinate (S_j) and destination coordinate (D_j), and receives a menu of service options. For simplicity, we assume just two options—an exclusive ride with no detours at a price of p_j^x, and a (possibly) shared ride with an estimated detour of $\hat{\delta}_j$ at a price of p_j^s.

(b) **Stage 2:** The user evaluates these options and performs one of three actions: requesting the exclusive service, requesting the shared service, or neither. Under the former two actions, the user is assigned an appropriate vehicle, which could result in either initiating a new ride (possible under either service), or modifying an existing ride (possible only under the shared service).

We assume that the prices p_j^x and p_j^s are *upfront* prices (rather than the more traditional *per-mile* and/or *per-minute* pricing), which is in accordance with most major commercial ridesharing systems. For our analysis, we consider real-time, or "Ride Now" requests, and require fast, time-sensitive responses from the service provider.[2] Thus, it is important for the service provider to

[2] Scheduled ride options have only been launched recently, e.g., as recently as 2016 for Uber, and are typically available to only a subset of the users and in limited geographical regions.

respond quickly in the first stage, whereas the second stage, especially when a shared ride is requested, may take slightly longer due to a possibly batched combinatorial optimization. For instance, Uber's price estimates are not indicative of real-time availability [61]. *Our focus is on the decision problem of the service provider in **Stage 1**, that is, what values of p_j^x, p_j^s, and $\hat{\delta}_j$ should be returned to the user?*

The service provider, in the first stage, considers each new user sequentially and immediately, independent of other new users. While this could lead to suboptimal solutions, it avoids a computationally intensive, likely combinatorial optimization which would be infeasible for a large-scale system. Moreover, greedy, myopic profit-maximization strategies have been empirically shown to be close to optimal on New York City taxi data [10]. To be precise, when a user j inputs their source and destination coordinates, the service provider considers all possible existing shareable rides that could *feasibly* detour from their existing route to serve this additional user. Then, for each of these feasible rides, the provider computes the optimal values of p_j^x, p_j^s, and $\hat{\delta}_j$, as well as the corresponding optimal incremental profit, if user j were to be added to these rides. The ride offering the maximum optimal incremental profit is then chosen as a *tentative* match, and the corresponding optimal prices and detour estimate are returned to the user. A more intensive optimization can be carried out in the second stage, if and when the user actually requests a shared ride, which could end up finding a better match, perhaps with other shared ride requests that were also received recently, e.g., [42] develop reoptimization methods for dynamic vehicle routing.

In summary, when a new user j's first stage query (S_j, D_j) is received, the provider must answer the following questions, with respect to each existing shareable ride:

- *Feasibility:* Would adding user j to the ride maximize the expected incremental profit?
- *Pricing:* What are the prices (p_j^x, p_j^s) that maximize the expected incremental profit?
- *Detour Estimate:* What is the detour estimate $\hat{\delta}_j$ to be provided?

The rest of this section is organized as follows. We introduce aspects of the users' and service provider's models in Sections 4.1 and 4.2, respectively. Section 4.3 presents the key results from the service provider's optimization problem, wherein Theorems 1 and 3 address feasibility and optimal pricing, respectively. In Section 4.4, we discuss whether the service provider has an incentive to be strategic when considering truthful revelation of the detour estimate. Finally, Section 4.5 presents some illustrative numerical examples.

For any two spatial coordinates A and B, we let $d(A,B)$ denote the shortest distance from A to B. An exclusive ride always provides service along a shortest route.

4.1. Model for ridesharing users.

4.1.1. User's utility. User i has a valuation $v_i > 0$ per mile for exclusive service. These valuations are independently and identically distributed across users, according to a distribution with cumulative distribution function F_v and corresponding density function f_v. For shared service, the user's valuation depreciates by a factor $k_i(\delta_i)$, a decreasing function of δ_i, the *fractional* detour experienced by user i. To be precise, δ_i is the *additional* distance travelled by user i due to sharing service (over and above the shortest distance $d(S_i, D_i)$), as a fraction of $d(S_i, D_i)$. We let $k(0) = \overline{k} \leq 1$ to model fixed, non-detour-related inconveniences from sharing. Thus, the utility function of user i is given by:

$$U_i(\texttt{choice}_i; p_i^x, p_i^s, \delta_i) = \begin{cases} v_i d(S_i, D_i) - p_i^x, & \texttt{choice}_i = \texttt{Exclusive} \\ k_i(\delta_i) v_i d(S_i, D_i) - p_i^s, & \texttt{choice}_i = \texttt{Shared} \\ 0, & \texttt{choice}_i = \texttt{Declined} \end{cases} \tag{1}$$

At the time of making the choice, user i does not know the actual detour δ_i that they would experience. Instead, they only know the estimated detour $\hat{\delta}_i$. Thus, the users set $\texttt{choice}_i$ to maximize $U_i(\texttt{choice}_i; p_i^x, p_i^s, \hat{\delta}_i)$. Moving forward, we define

$$\hat{U}_i^s(p_i^s) = U_i(\texttt{Shared}; \cdot, p_i^s, \hat{\delta}_i) = k_i(\hat{\delta}_i) v_i d(S_i, D_i) - p_i^s \tag{2}$$

to be the *estimated* utility of user i when choosing `Shared`, and

$$U_i^s(p_i^s, \delta_i) = U_i(\texttt{Shared}; \cdot, p_i^s, \delta_i) = k_i(\delta_i) v_i d(S_i, D_i) - p_i^s \tag{3}$$

to be the *actual* utility of user i when choosing `Shared`.

We assume that the depreciation functions k_i of the users and the distribution F_v of their exclusive per-mile valuations are known to the service provider; however, the realized valuations v_i are private information to the users.

4.1.2. User's choice. Intuitively, users with 'low' v_i would choose `Declined`, those with 'high' v_i would choose `Exclusive`, and those with 'intermediate' v_i would choose `Shared`. We now formalize this threshold behavior of the user choice.

If a user i chooses `Shared`, then, it implies that $\hat{U}_i^s(p_i^s)$ (defined in (2)) is greater than the utility from choosing `Exclusive` or `Declined`:

$$k_i(\hat{\delta}_i) v_i d(S_i, D_i) - p_i^s > \max\{0, v_i d(S_i, D_i) - p_i^x\}. \tag{4}$$

Simplifying the above inequality yields

$$\underline{v}_i < v_i < \overline{v}_i, \tag{5}$$

where the lower and upper bounds, $\underline{v}_i$ and $\overline{v}_i$ are given by:

$$\underline{v}_i = \frac{p_i^s}{k_i(\hat{\delta}_i) d(S_i, D_i)}, \qquad \overline{v}_i = \frac{p_i^x - p_i^s}{(1 - k_i(\hat{\delta}_i)) d(S_i, D_i)}. \tag{6}$$

An immediate necessary condition for (5) to be satisfied for some v_i is that $\underline{v}_i < \overline{v}_i$, which yields:

$$p_i^s < k_i(\hat{\delta}_i) p_i^x, \tag{7}$$

which imposes a constraint on the prices that the service provider considers, should it be feasible to offer a shared ride option to user i. Moreover, when the above constraint is violated, i.e., when $p_i^s \geq k_i(\hat{\delta}_i) p_i^x$, the exact value of p_i^s does not affect a user's choice between `Exclusive` and `Declined`, since that choice would be completely determined by p_i^x. This observation relieves the service provider from explicitly considering $p_i^s > k_i(\hat{\delta}_i) p_i^x$, simplifying the search space. Thus, without loss of generality, we assume that

$$p_i^s \leq k_i(\hat{\delta}_i) p_i^x. \tag{8}$$

A similar analysis for the choices `Exclusive` and `Declined`, under (8), results in the following user choice function:

$$\texttt{choice}_i^*(v_i; \underline{v}_i, \overline{v}_i) = \begin{cases} \texttt{Declined}, & v_i \leq \underline{v}_i \\ \texttt{Shared}, & \underline{v}_i < v_i < \overline{v}_i \\ \texttt{Exclusive}, & v_i \geq \overline{v}_i \end{cases} \tag{9}$$

4.1.3. A new user's impact on an existing shareable ride. Suppose a new user j is being considered for addition to an existing shareable ride with $j-1$ users in the vehicle. (We assume that $j-1 \geq 1$; the bootstrapping problem of computing optimal prices to be offered to a user to initiate a new shareable ride does not involve any existing passengers.) Without loss of generality, we assume that the indices of the existing users are in the order in which they are scheduled to be dropped off according to the existing route plan, with ties broken arbitrarily (e.g., when two or more existing users share a common destination). Let D_0 denote the current location of the vehicle. If user j is added to the existing ride, we assume that the new route plan leaves unchanged the relative order in which the existing users are scheduled to be dropped off. (This enables the routing optimization to be quick, by limiting to a quadratic number of possibilities.) Let $t_j^- < j-1$ (respectively, $t_j^+ \leq j$) denote the largest (respectively, smallest) among the indices of existing users who are dropped off immediately before picking up (respectively, after dropping off) user j, according to the new route plan. If nobody gets dropped off before j is picked up, define $t_j^- = 0$. (If j is the last user to be dropped off, define $t_j^+ = j$.) Define the following quantities:

$$\Delta_j^s = d\left(D_{t_j^-}, S_j\right) + d\left(S_j, D_{t_j^-+1}\right) - d\left(D_{t_j^-}, D_{t_j^-+1}\right) \tag{10}$$

$$\Delta_j^d = \begin{cases} d\left(S_j, D_j\right) + d\left(D_j, D_1\right) - d\left(S_j, D_1\right), & t_j^+ = 1 \\ d\left(D_{t_j^+-1}, D_j\right) + d\left(D_j, D_{t_j^+}\right) - d\left(D_{t_j^+-1}, D_{t_j^+}\right), & 1 < t_j^+ < j \\ d\left(D_{j-1}, D_j\right), & t_j^+ = j \end{cases} \tag{11}$$

Δ_j^s and Δ_j^d are the *source detour* and *destination detour* from the current route to serve the additional user j, respectively. Not all existing users experience both of these detours, as we discuss next.

Suppose δ_i^{j-1} denotes the fractional detour that would be incurred by an existing user $i < j$ according to the existing route plan, if user j is not added to the shared ride. Then,

$$\delta_i^j = \delta_i^{j-1} + \frac{\mathbb{1}\{i > t_j^-\}\Delta_j^s + \mathbb{1}\{i \geq t_j^+\}\Delta_j^d}{d(S_i, D_i)} \tag{12}$$

is the fractional detour that would be incurred by user i according to the new route plan if user j is added to the shared ride. Let δ_j^j denote the fractional detour that user j would experience according to the new route plan.

4.1.4. Individual Rationality (IR). A shared ride is individually rational (IR) for a user, if their utility from the shared ride is nonnegative. There are different notions of IR in the literature; the one we focus on is called *ex-post IR*, and means that the *actual* utility of the user at the end of the shared ride, given by (3), is nonnegative, that is, $U_i^s(p_i^s, \delta_i) \geq 0$ for all i. Since U_i^s is a decreasing function of δ_i, this property is always satisfied when the service provider ensures that $\delta_i \leq \hat{\delta}_i$ for all i, because that would, in turn, ensure that $U_i^s(p_i^s, \delta_i) \geq \hat{U}_i^s(p_i^s)$, which is nonnegative because the user chose `Shared`.

Motivated by recent experiments [15] highlighting the benefits to a service provider of *proactively* compensating users whose ex-post IR constraints may have been violated, we adopt ex-post IR as an indicator of the provider's Quality-of-Service (QoS).

4.1.5. Sequential Individual Rationality (SIR). In the commercial ridesharing context, our notion of *sequential IR* (SIR) requires that the service provider *sustain* ex-post IR for all the users, at every stage of a shared ride. In other words, whenever a new user j is considered for addition to an existing shareable ride with $j-1$ users in the vehicle, the service provider must ensure that $U_i^s(p_i^s, \delta_i^j) \geq 0$ for all $i \leq j$, where δ_i^j is given by (12). At first glance, such a notion may seem unnecessarily strong; however, in our model, SIR is *necessary* to ensure ex-post IR. This is because, for a fixed p_i^s that was committed to at the time that user i joined the shared ride, $U_i^s(p_i^s, \delta_i)$ is a decreasing function of δ_i, which, in turn, is nondecreasing as new users are added.

Of course, this concern about violating ex-post IR is exactly what was addressed by [15] in their experimental study. Their findings motivate us to consider, in the rest of this section, the consequences of violating SIR in the interim (during the shared ride), but restoring ex-post IR (at the end of the shared ride) through an appropriate monetary compensation to the user (penalty to the provider). Perhaps surprisingly, our results (Section 4.3) indicate that the service provider's profit-optimal, QoS-aware policy is one that may violate SIR, and quite liberally at that.

4.2. Model for QoS-aware service provider. A QoS-aware service provider, when evaluating the potential for profit from adding a new user j to an existing shareable ride, must be aware of how it would impact the utilities of the existing users $i < j$. As discussed earlier, while ensuring that $\delta_i^j \le \hat{\delta}_i$ for all $i \le j$ would guarantee SIR-compliance, such a policy may be too restrictive. For example, allowing δ_i^j to slightly exceed $\hat{\delta}_i$ for a user i, would result in $U_i^s(p_i^s, \delta_i^j) < \hat{U}_i^s(p_i^s)$, but it may not necessarily violate their ex-post IR constraint, since it is still possible that $U_i^s(p_i^s, \delta_i^j) > 0$ if the user's valuation v_i were large enough. Although v_i is private information, the service provider can infer that v_i lies in a range specified by (5). Hence, the provider could consider a risky policy in which δ_i^j exceeds $\hat{\delta}_i$, but then, proactively compensate the users for the potential violation of their ex-post IR constraints, e.g., by means of a discount coupon [15]. We call this compensation an *incremental penalty.*

4.2.1. Incremental penalty. A service provider's incremental penalty depends on the value of δ_i^j relative to δ_i^{j-1} and $\hat{\delta}_i$. Thus, we consider the following three cases:

- $\delta_i^{j-1} \le \delta_i^j \le \hat{\delta}_i$: Here, the actual utility of user i, while reduced due to the incremental detour caused by the addition of j, nevertheless stays above their expected utility, and hence, is nonnegative. Thus, the service provider incurs no incremental penalty.
- $\delta_i^{j-1} \le \hat{\delta}_i < \delta_i^j$: Here, the addition of user j would result in exceeding the expected detour promised to user i, which could result in their actual utility falling below their expected utility. Still, the service provider incurs an incremental penalty only if the actual utility becomes negative. Thus,

$$\Delta\texttt{Penalty}_i^j = \begin{cases} 0, & U_i^s(p_i^s, \delta_i^j) \ge 0 \\ -U_i^s(p_i^s, \delta_i^j), & U_i^s(p_i^s, \delta_i^j) < 0 \end{cases} \tag{13}$$

- $\hat{\delta}_i < \delta_i^{j-1} \le \delta_i^j$: Here, the expected detour promised to user i has already been exceeded, and the addition of user j would result in a further excess. Thus,

$$\Delta\texttt{Penalty}_i^j = \begin{cases} 0, & U_i^s(p_i^s, \delta_i^j) \ge 0 \\ U_i^s(p_i^s, \delta_i^{j-1}) - U_i^s(p_i^s, \delta_i^j), & U_i^s(p_i^s, \delta_i^{j-1}) < 0 \\ -U_i^s(p_i^s, \delta_i^j), & U_i^s(p_i^s, \delta_i^{j-1}) \ge 0 \;\&\&\; U_i^s(p_i^s, \delta_i^j) < 0 \end{cases} \tag{14}$$

Combining all the above cases, we obtain:

$$\Delta\texttt{Penalty}_i^j = \min\{0, U_i^s(p_i^s, \delta_i^{j-1})\} - \min\{0, U_i^s(p_i^s, \delta_i^j)\}. \tag{15}$$

4.2.2. Expected incremental penalty. Using the prior distribution F_v, and the inferred range of v_i from (5), the *expected* incremental penalty can be computed as follows:

$$\begin{aligned} Exp\Delta\texttt{Penalty}_i^j &= \mathbb{E}_{v_i}\left\{\Delta\texttt{Penalty}_i^j \mid \texttt{choice}_i^* = \texttt{Shared}\right\} \\ &= \frac{\int_{\min\{\underline{v}_i^{j-1}, \underline{v}_i\}}^{\min\{\underline{v}_i^{j-1}, \overline{v}_i\}} U_i^s(p_i^s, \delta_i^{j-1}) dF_v(v_i) - \int_{\min\{\underline{v}_i^j, \underline{v}_i\}}^{\min\{\underline{v}_i^j, \overline{v}_i\}} U_i^s(p_s^i, \delta_i^j) dF_v(v_i)}{F_v(\overline{v}_i) - F_v(\underline{v}_i)}, \end{aligned} \tag{16}$$

where $\underline{v}_i^{j-1}$ and $\underline{v}_i^j$ are given by:

$$\underline{v}_i^{j-1} = \frac{p_i^s}{k_i(\delta_i^{j-1})d(S_i, D_i)}, \qquad \underline{v}_i^j = \frac{p_i^s}{k_i(\delta_i^j)d(S_i, D_i)}. \tag{17}$$

4.2.3. Maximum incremental penalty. Compensating user i an amount equal to $Exp\Delta\texttt{Penalty}_i^j$ would restore their ex-post IR property *in expectation*, but if the user's realized value of v_i were small enough, they could still be left unsatisfied. Thus, the service provider may alternatively consider providing user i with the maximum possible amount by which their ex-post IR constraint could have been violated, denoted by $Max\Delta\texttt{Penalty}_i^j$. It can be seen that this is given by the value of $\Delta\texttt{Penalty}_i^j$ from (15) evaluated at:

$$v_i^{Max\Delta\texttt{Penalty}} = \min\left\{\overline{v}_i, \max\left\{\underline{v}_i, \underline{v}_i^{j-1}\right\}\right\}. \tag{18}$$

4.2.4. Service provider's incremental profit. We can now write down the incremental profit to a QoS-aware service provider considering adding a new user j to an existing shareable ride with $j-1$ users in the vehicle. Let c denote the cost per mile that the service provider incurs. Then, the incremental profit is given by:

$$\begin{aligned} &\Delta P_j(p_j^x, p_j^s; \hat{\delta}_j, \texttt{choice}_j) \\ &\quad = \begin{cases} p_j^x - c\, d(S_j, D_j), & \texttt{choice}_j = \texttt{Exclusive} \\ p_j^s - c\left(\Delta_j^s + \Delta_j^d\right) - \beta \sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j, & \texttt{choice}_j = \texttt{Shared} \\ 0, & \texttt{choice}_j = \texttt{Declined}, \end{cases} \end{aligned} \tag{19}$$

where $Any\Delta\texttt{Penalty}_i^j$ refers to either the expected or maximum incremental penalty (depending on the service provider's intent) that would be incurred as compensation to user i for a possible violation of their ex-post IR constraint, due to sharing the ride with user j. $\beta \geq 0$ is a parameter that incorporates the following effects:

- Higher the value of β, more QoS-aware the service provider is.
- Higher the value of β, more *loss averse* the existing users are towards considering the monetary compensation offered at the end of the ride as restoring their ex-post IR.

Note that any detour estimate computed exogenously by the service provider for user j must satisfy $\hat{\delta}^j \geq \delta_j^j$. We assume that a sufficiently QoS-aware service provider would communicate this estimate truthfully to user j, and therefore, there is no need to consider a term $Any\Delta\texttt{Penalty}_j^j$ in (19). We formally show, in Appendix EC.2, that our assumption is indeed valid, as long as $\beta \geq 1$.

Thus, the *expected* incremental profit is the sum of the incremental profits when user j chooses `Exclusive` and `Pooled`, weighted by the respective probabilities of these choices, given the distribution F_v, and the user choice function (9):

$$\begin{aligned} Exp\Delta P_j(p_j^x, p_j^s) &= (1 - F_v(\overline{v}_j))\Delta P_j(p_j^x, p_j^s; \hat{\delta}_j, \texttt{Exclusive}) \\ &\quad + \left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right)\Delta P_j(p_j^x, p_j^s; \hat{\delta}_j, \texttt{Shared}). \end{aligned} \tag{20}$$

Of particular interest is the probability of user j choosing `Shared`, which is given by:

$$ProbSharing_j(p_j^x, p_j^s) = \mathbb{P}\left\{\texttt{choice}_j^* = \texttt{Shared}\right\} = F_v(\overline{v}_j) - F_v(\underline{v}_j). \tag{21}$$

A QoS-aware service provider's objective is therefore to maximize $Exp\Delta P_j(p_j^x, p_j^s)$, subject to $p_j^x \geq cd(S_j, D_j)$ and $0 \leq p_j^s \leq k_j(\hat{\delta}_j)p_j^s$.

4.3. Optimal policy for a QoS-aware service provider. To simplify the exposition of the results, we assume that the density function f_v has an unbounded support, in particular, f_v is continuous in $[0,\infty)$. (Our results extended to distributions with finite support.) We assume that the distribution F_v is *regular*, a standard assumption in the literature [38]. This means that the function ϕ_v, given by

$$\phi_v(x) = x - \frac{1-F_v(x)}{f_v(x)}, \tag{22}$$

is strictly increasing in $[0,\infty)$, and hence invertible. Many common distributions satisfy this assumption in practice; in particular, all distributions with nonincreasing hazard rate are regular.

Our first result concerns the uniqueness of the optimal prices, $p_j^{x,*}$ and $p_j^{s,*}$.

LEMMA 1. *If F_v is regular, then, $p_j^{x,*} \in [0,\infty)$, $p_j^{s,*} \in \left[0, k_j(\hat{\delta}_j)p_j^{x,*}\right]$ are unique.*

Lemma 1 follows from the observation that the objective (20) is concave when F_v is regular.

Our second result concerns the optimal probability of user j choosing `Shared`, given by $ProbSharing_j^* = ProbSharing_j(p_j^{x,*}, p_j^{s,*})$. We provide a characterization of the *optimal shareable region*, that is, the possible locations of S_j and D_j (relative to the current location of the vehicle and the drop-off locations of the existing users) for which $ProbSharing_j^* > 0$.

THEOREM 1. *$ProbSharing_j^* > 0$ if and only if*

$$c\left(k_j(\hat{\delta}_j)d(S_j,D_j) - (\Delta_j^s + \Delta_j^d)\right) > \beta \sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j. \tag{23}$$

Theorem 1 serves as an important operational tool for a QoS-aware, profit-maximizing service provider to determine the *feasibility* of an existing shareable ride in accommodating a new request. The left hand side of (23) is the difference in the provider's operating costs when j is served exclusively (scaled down by $k_j(\hat{\delta}_j)$) and when j is added to the existing ride, and the right hand side is the resulting penalty to be paid to the existing users. The proof of Theorem 1 is deferred to Appendix EC.1.

Our next result exposes a surprising property of the optimal shareable region. For any source coordinate S_j, define the following two regions:

- $\mathcal{R}_j^{SIR}(S_j)$ is the collection of destination coordinates D_j for which the right hand side of (23) vanishes for all β. In other words, adding any request within this region to the existing shareable ride will not violate SIR for any existing user.
- $\mathcal{R}_j^{OPT(\beta)}(S_j)$ is the optimal shareable region, consisting of all destination coordinates D_j for which (23) is satisfied, that is, $ProbSharing_j^* > 0$.

We can now state our result:

THEOREM 2. *For any S_j, if $\mathcal{R}_j^{OPT(0)}(S_j) \nsubseteq \mathcal{R}_j^{SIR}(S_j)$, then, for all $\beta > 0$, $\mathcal{R}_j^{OPT(\beta)}(S_j) \nsubseteq \mathcal{R}_j^{SIR}(S_j)$.*

Theorem 2 states that if there exists a request (S_j, D_j) for which it is optimal for a QoS-agnostic service provider (with $\beta = 0$) to violate SIR, then, there exists a request for which it is optimal for a QoS-aware service provider (with $\beta > 0$) to violate SIR, no matter how large their QoS-sensitivity β. Informally, as β increases, the region $\mathcal{R}_j^{OPT(\beta)}(S_j)$ keeps shrinking, and as $\beta \to \infty$, it converges to a finite region $\mathcal{R}_j^{OPT(\infty)}(S_j)$ that is contained within $\mathcal{R}_j^{SIR}(S_j)$. This behavior is illustrated through a numerical example in Section 4.5. The proof of Theorem 2 is deferred to Appendix EC.1.

Our final result provides closed form expressions for the optimal prices.

Theorem 3. *The optimal prices are given by*

$$p_j^{x,*} = \begin{cases} \left((1-k_j(\hat{\delta}_j))\phi_v^{-1}\left(\frac{\Delta c^x - \Delta c^s}{(1-k_j(\hat{\delta}_j))d(S_j,D_j)}\right) + k_j(\hat{\delta}_j)\phi_v^{-1}\left(\frac{\Delta c^s}{k_j(\hat{\delta}_j)d(S_j,D_j)}\right)\right)d(S_j,D_j), & ProbSharing_j^* > 0 \\ \phi_v^{-1}(c)\, d(S_j,D_j), & ProbSharing_j^* = 0 \end{cases}$$
$$p_j^{s,*} = \begin{cases} \phi_v^{-1}\left(\frac{\Delta c^s}{k_j(\hat{\delta}_j)d(S_j,D_j)}\right)k_j(\hat{\delta}_j)d(S_j,D_j), & ProbSharing_j^* > 0 \\ \phi_v^{-1}(c)\, k_j(\hat{\delta}_j)d(S_j,D_j), & ProbSharing_j^* = 0, \end{cases} \tag{24}$$

where $\Delta c^x = cd(S_j,D_j)$ *and* $\Delta c^s = c(\Delta_j^s + \Delta_j^d) + \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j$ *are the incremental costs incurred by the service provider when serving user* j *exclusively and shared, respectively.*

The proof of Theorem 3 is deferred to Appendix EC.1.

4.4. Strategic concerns regarding detour estimates. Our results in the previous section require that the service provider provide a new user j with a detour estimate $\hat{\delta}_j \geq \delta_j^j$. However, one may wonder if the service provider might stand to gain by "luring" user j with a false, smaller detour estimate $\hat{\delta}_j < \delta_j^j$, and then suffering a penalty equal to $\beta Any\Delta\mathtt{Penalty}_j^j$. We omit a detailed discussion of this issue due to space limitations, but it can be shown (see Appendix EC.2) that a sufficiently QoS-aware service provider ($\beta \geq 1$) cannot gain from lying, e.g., shading the detour estimate.

4.5. Numerical examples. In this section, we illustrate the spatial properties of the optimal shareable region visually, using a small numerical example in the Euclidean space $\mathbb{R}^2$. First, we consider a simple scenario where all the users are traveling to a common destination, D. A shared ride is 'bootstrapped' by the first passenger, whose source is S_1. The grey shaded ellipse-shaped region in Figure 1 (*left*) depicts the "SIR-feasible region" $\mathcal{R}_j^{SIR}(D)$ (for $j=2$), the collection of source coordinates S_2 from which a second user can be added to the ride without violating SIR for the first user. In the same figure, the dashed curves represent the boundaries of the optimal shareable regions $\mathcal{R}_j^{OPT(\beta)}(D)$ (for $j=2$), for $\beta = 1, 20$. Then, Figure 1 (*center* and *right*) shows how these regions change as the ride progresses, for $j=3$ and $j=4$, respectively, for a random selection of S_2 (and subsequently S_3).

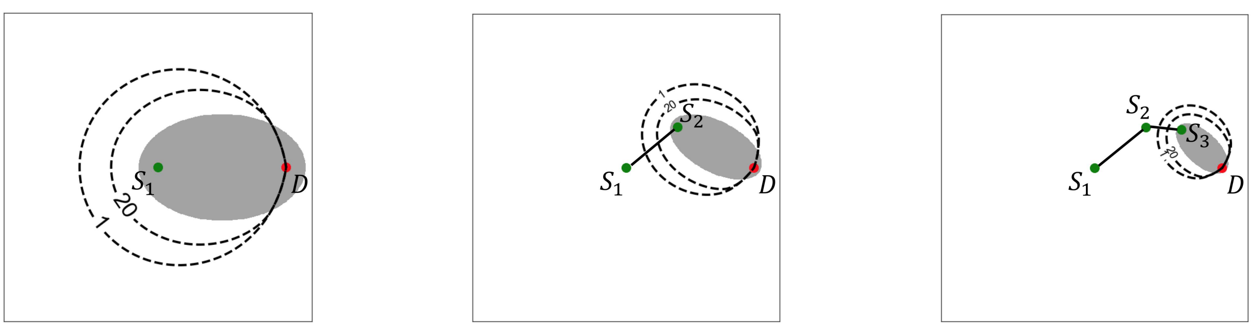

Figure 1. Evolution of the optimal shareable region (interior of the dashed curves, for $\beta = 1, 20$) from within which it is profitable to add a subsequent user to the ride, as the ride progresses and more users are added. For reference, the grey shaded area shows the SIR-feasible region.

First, observe that there are points within $\mathcal{R}_j^{SIR}(D)$ that are not in the optimal shareable region. Thus, even though the provider incurs no penalty by adding a user from such points, it would be suboptimal to do so. Next, observe that the portion of $\mathcal{R}_j^{OPT(\beta)}(D)$ that is outside $\mathcal{R}_j^{SIR}(D)$ is smaller for $\beta = 20$ than for $\beta = 1$. This (partially) illustrates the convergence argument that supports Theorem 2.

Next, we consider a more complex scenario where users have different sources and destinations. In Figure 2, the interpretations of the dashed curves and the grey shaded regions are the same as before, except that they depict possible locations of D_2 (*left*) and D_3 (*right*), respectively. The bottom half of Figure 2 is the "zoomed out" version of the top half, that demonstrates that $\mathcal{R}_j^{OPT(\beta)}(S_j)$ (for $j=2,3$) are *closed* regions. The shapes that define the regions in Figure 2 are more complicated than those in Figure 1 due to the spatial discontinuities associated with the order in which the users are dropped off.

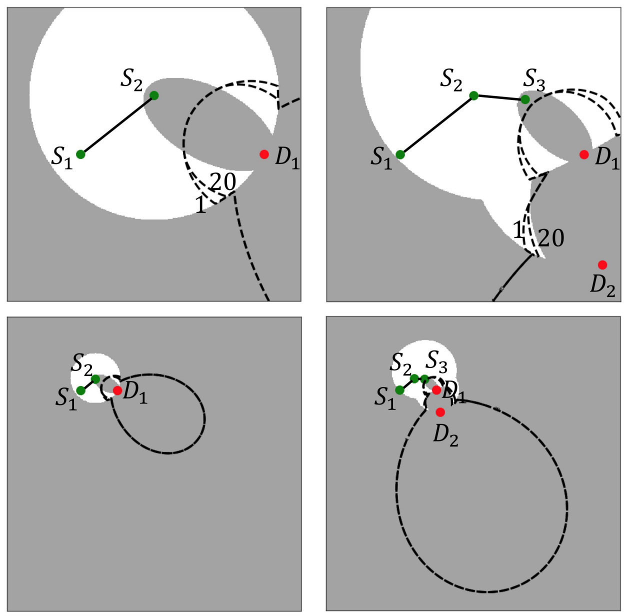

FIGURE 2. Evolution of the optimal shareable region (interior of the dashed curves, for $\beta = 1, 20$) to within which it is profitable to drop off the subsequent users, as the ride progresses and more users are added. For reference, the grey shaded area shows the SIR-feasible region.

5. Fair cost sharing in community carpooling. In this section, we consider a different scenario for dynamic ridesharing, namely, community carpooling. Typically, this involves a large pool of commuters who wish to share a ride to/from a common destination/source. For example, it is common for employers to facilitate and encourage carpooling among their employees [20, 44], due to tax benefits they often enjoy as a result [51]. However, participation levels within organizations is consistently low [14]. While improving the carpooling experience by forming more efficient carpooling groups should increase participation, perhaps the biggest obstacle is the perceived loss of flexibility due to relying on a fixed set of other commuters. While the pool of participants does not change significantly over a short period of time, everyday demand for travel within the pool can be quite dynamic due to diverse/flexible schedules and needs, which renders a static carpooling solution quite inefficient [24].

Solving the carpooling problem consists of the following stages:

(a) **Group Formation:** The pool of participating commuters is partitioned into smaller groups of commuters that carpool together, based on several factors, including spatio-temporal constraints.

(b) **Routing:** Each carpooling group decides the best route (order of pickups), based on the constraints of the individual commuters within the group.

(c) **Cost Sharing:** Commuters within each group decide how to share the operational cost of the ride among themselves.

Typically, these stages take place in a top-down sequence, that is, group formation and routing first, followed by cost sharing [33, 63]. In fact, most organizational programs only go so far as to facilitate group formation [2, 49], leaving the groups to solve the routing and cost sharing problems on their own. This may lead to inefficient schemes that involve, e.g., commuters within a group taking turns driving their own vehicle in order to avoid the cost sharing issue altogether, since small-scale monetary transactions between friends or colleagues may be perceived as 'awkward' [13].

Since our objective is to incorporate Quality of Service (QoS) and fairness considerations while constructing a more flexible and dynamic solution, we take a bottom-up approach. Accordingly, the focus of this section is on the design of *QoS-aware* and *fair* cost sharing schemes, and the constraints that such a requirement imposes on the routing, which, in turn, restrict the space of feasible partitions of the commuter pool into carpooling groups. We argue that emphasizing key aspects of commuters' carpooling experience as starting points of the solution design process yields a better carpooling solution. Perhaps more importantly, designing the cost sharing scheme first allows it to be agnostic to the type of scenario (static or dynamic) that it would be applied in. We note that modern technology allows such cost shares to be tracked automatically behind the scenes, and be settled at a later time, perhaps at regular (e.g., monthly) intervals [1].

We begin with a description of our model for cost sharing, introduce the appropriate notion of QoS, and present results that characterize routes for which QoS-aware cost sharing schemes exist. We then move on to introducing an appropriate notion of fairness, and characterize QoS-aware cost sharing rules that are also fair. Finally, we discuss how our results apply to both static and dynamic carpooling scenarios.

5.1. Model for cost sharing in carpooling. As discussed above, our approach is to focus on the design of fair cost sharing schemes for a fixed set of commuters N carpooling to a common destination D, and a fixed route r_N (ordered sequence of the commuters). Let $N = \{1, 2, \ldots, j\}$, and, without loss of generality, let $r_N = (1, 2, \ldots, j)$, that is, the commuters are indexed in the order in which they appear in the route r_N. Let S_i denote the source coordinates of commuter $i \in N$. The initial state of a ride involves the first commuter, with $i = 1$, driving their vehicle from S_1 towards D. For any two spatial coordinates A and B, we let $d(A, B)$ denote the shortest distance from A to B. We assume that the operational cost of a ride involving j commuters is proportional to the total distance traveled by the vehicle according to the route r_N, and is given by

$$\mathcal{OC}(N; r_N) = c\left(\sum_{i=1}^{j-1} d(S_i, S_{i+1}) + d(S_j, D)\right), \tag{25}$$

where c is the operational cost per mile, which is either set by the first commuter, or by the system (depending on the characteristics of the vehicle).

Let f denote the cost sharing scheme according to which $\mathcal{OC}(N; r_N)$ is shared among the j commuters. In particular, $f(i, N; r_N)$ denotes the share of $\mathcal{OC}(N; r_N)$ borne by commuter $i \in N$ under route r_N. A cost sharing scheme f is said to be budget-balanced if,

$$\sum_{i=1}^{j} f(i, N; r_N) = \mathcal{OC}(N; r_N). \tag{26}$$

Our goal is to design budget-balanced cost sharing schemes f that are also *QoS-aware* and *fair*. Since we adopt a utilitarian approach to modeling QoS and fairness, we introduce the utility model for commuters first.

5.1.1. Disutility and detour-inconvenience. The *disutility* of commuter $i \in N$ according to the route r_N is given by

$$\mathcal{DU}(i,N;r_N) = f(i,N;r_N) + \mathcal{IC}(i,N;r_N), \tag{27}$$

where $\mathcal{IC}(i,N;r_N)$ denotes the "inconvenience cost" due to the detour endured by commuter $i \in N$, caused by commuters $i+1,\dots,j$ that joined the ride after i did (according to the route r_N). In our model, we let this term be proportional to the length of the detour:

$$\mathcal{IC}(i,N;r_N) = \alpha_i \left(\sum_{k=i}^{j-1} d(S_k,S_{k+1}) + d(S_j,D) - d(S_i,D) \right), \tag{28}$$

where α_i is a parameter that denotes the *detour-sensitivity* of commuter i. Note that $\mathcal{IC}(j,N;r_N) = 0$, since the last commuter to join the ride suffers no detour. An equivalent expression for the inconvenience cost is given by

$$\mathcal{IC}(i,N;r_N) = \alpha_i \sum_{k=i+1}^{j} \delta_k, \tag{29}$$

where $\delta_k = d(S_{k-1},S_k) + d(S_k,D) - d(S_{k-1},D)$ denotes the *incremental* detour due to commuter $k > 1$ joining the ride.

5.1.2. Individual Rationality (IR). A cost sharing scheme f is Individually Rational (IR) for a commuter $i \in N$, if their disutility from the shared ride is not more than that from an alternative, which we assume to be driving their own vehicle to the destination. We say that f is IR on route r_N, if it is IR for all commuters at the end of the ride:

$$\mathcal{DU}(i,N;r_N) \le cd(S_i,D), \quad \forall\, i \in N. \tag{30}$$

Substituting for the disutility from (27), and using (29) for the inconvenience cost, we can state the following equivalent definition. A cost sharing scheme f is IR on a route r_N, if

$$f(i,N;r_N) + \alpha_i \sum_{k=i+1}^{j} \delta_k \le cd(S_i,D), \quad \forall\, i \in N. \tag{31}$$

5.1.3. Sequential Individual Rationality (SIR). In the context of a static carpooling solution, the carpooling group N, the route r_N, and the cost shares of each carpooling commuter $f(i,N;r_N)$ are known in advance; therefore, IR is an acceptable indicator of QoS. However, in dynamic carpooling, the group N is only revealed sequentially (according to the route r_N) over time. This results in the commuters experiencing a corresponding sequence of disutilities. In particular, when commuter $k \in N$ joins the ride, the *updated* disutilities of commuters $1 \le i \le k$ are given by $\mathcal{DU}(i,N(k);r_{N(k)})$, where $N(k) = \{1,2,\dots,k\}$ and $r_{N(k)} = (1,2,\dots,k)$ denotes the partial route up to and including k. Therefore, the sequence of disutilities experienced by a commuter $i \in N$ is given by $\mathcal{SDU}(i,N;r_N) = \big(cd(S_i,D), \mathcal{DU}(i,N(i);r_{N(i)}), \mathcal{DU}(i,N(i+1);r_{N(i+1)}),\dots,\mathcal{DU}(i,N;r_N)\big)$. (For convenience, the sequence of disutilities is prefixed with the disutility from commuter i's alternative.)

We say that a cost sharing scheme f is *sequentially IR* or SIR on route r_N, if, for all $i \in N$, $\mathcal{SDU}(i,N;r_N)$ is nonincreasing, that is, for all $i \in N$,

$$\begin{aligned} \mathcal{DU}(i,N(i);r_{N(i)}) &\le cd(S_i,D), \text{ and} \\ \mathcal{DU}(i,N(k+1);r_{N(k+1)}) &\le \mathcal{DU}(i,N(k);r_{N(k)}) \ \forall\, k \in \{i,i+1,\dots,j-1\}. \end{aligned} \tag{32}$$

Using (27) and (29) for the disutilities and inconvenience cost yields the following equivalent definition. A cost sharing scheme f is SIR on a route r_N, if, for all $i \in N$,

$$\begin{gathered} f(i, N(i); r_{N(i)}) \leq cd(S_i, D), \text{ and} \\ f(i, N(k+1); r_{N(k+1)}) + \alpha_i \delta_{k+1} \leq f(i, N(k); r_{N(k)}) \ \forall \ k \in \{i, i+1, \dots, j-1\}. \end{gathered} \tag{33}$$

Note that while SIR guarantees IR, it is much stronger, and ensures that the entire *experience* of carpooling is favorable to all commuters.

5.1.4. Routes admitting QoS-aware cost sharing schemes. Not all routes admit budget-balanced, QoS-aware cost sharing schemes. Intuitively, this is because routes that include large detours may induce prohibitively large disutilities that violate IR/SIR.

Definition 1. A route r_N for a carpooling group N is *IR-feasible* (*SIR-feasible*) if there exists a budget-balanced cost sharing scheme f that is IR (SIR) on r_N.

From here on, whenever it is understood from context, we drop the explicit dependence of all the quantities on the route to simplify notation. Before moving on to the results, we present an illustrative example.

Example 1. Consider $j = 3$ commuters, picked up from their sources S_1, S_2, S_3 (in that order), travelling to a common destination D. The progression of the route $r_{N(k)}$, as the commuters are picked up one by one, is depicted in Figure 3. Given the final route r_N, the distances traveled by commuters are $d(S_1, S_2) + d(S_2, S_3) + d(S_3, D), d(S_2, S_3) + d(S_3, D)$, and $d(S_3, D)$, respectively. The total distance traveled by the carpooling vehicle is $d(S_1, S_2) + d(S_2, S_3) + d(S_3, D)$. The operational cost is thus $\mathcal{OC}(N) = c(d(S_1, S_2) + d(S_2, S_3) + d(S_3, D))$. Therefore, if f is a budget-balanced cost sharing scheme, $f(1, N) + f(2, N) + f(3, N) = c(d(S_1, S_2) + d(S_2, S_3) + d(S_3, D))$.

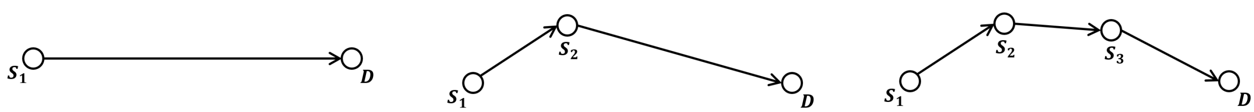

Figure 3. Route progress while picking up commuters traveling to a common destination.

The incremental detours due to commuters 2 and 3 are:

$$\delta_2 = d(S_1, S_2) + d(S_2, D) - d(S_1, D), \quad \delta_3 = d(S_2, S_3) + d(S_3, D) - d(S_2, D).$$

The inconvenience costs incurred by each commuter due to other commuters are:

$$\mathcal{IC}(1, N) = \alpha_1(\delta_2 + \delta_3), \quad \mathcal{IC}(2, N) = \alpha_2 \delta_3, \quad \mathcal{IC}(3, N) = 0.$$

Thus, a budget-balanced cost sharing scheme f is IR on route r_N if

$$f(1, N) + \alpha_1(\delta_2 + \delta_3) \leq cd(S_1, D), \quad f(2, N) + \alpha_2 \delta_3 \leq cd(S_2, D), \text{ and} \quad f(3, N) \leq cd(S_3, D).$$

A necessary condition for the route r_N to be IR-feasible is therefore obtained by summing up these inequalities, using budget-balance of f, and simplifying:

$$\left(1 + \frac{\alpha_1}{c}\right) \delta_2 + \left(1 + \frac{\alpha_1}{c} + \frac{\alpha_2}{c}\right) \delta_3 \leq d(S_2, D) + d(S_3, D).$$

The SIR constraints are stronger, since they require relative IR at every stage (when a subsequent commuter joins the ride):

$$f(1,N)+\alpha_1(\delta_2+\delta_3)\le f(1,N(2))+\alpha_1\delta_2\le cd(S_1,D),$$
$$f(2,N)+\alpha_2\delta_3\le f(2,N(2))\le cd(S_2,D),$$
$$f(3,N)\le cd(S_3,D).$$

A necessary condition for the route r_N to be SIR-feasible is therefore obtained by summing up these inequalities (at each stage), using budget-balance of f, and simplifying:

$$\left(1+\frac{\alpha_1}{c}\right)\delta_2\le d(S_2,D)\quad\text{and}\quad\left(1+\frac{\alpha_1}{c}+\frac{\alpha_2}{c}\right)\delta_3\le d(S_3,D).$$

The necessary conditions for IR/SIR can be interpreted as imposing upper bounds on the incremental detours at every stage of the ride. Perhaps surprisingly, they also turn out to be sufficient, as we show formally in the next section.

5.2. Characterizing IR/SIR-feasibile routes. The intuition gained from Example 1 suggests that routes with large detours are unlikely to be IR/SIR-feasible, that is, no budget-balanced cost sharing scheme would be IR/SIR on such routes. Theorems 4 and 5 provide formal characterizations of IR/SIR-feasible routes.

THEOREM 4. *The route $r_N=(1,2,\ldots,j)$ for a carpooling group $N=\{1,2,\ldots,j\}$ is IR-feasible if and only if*

$$\sum_{i=2}^{j}\left(1+\sum_{k=1}^{i-1}\frac{\alpha_k}{c}\right)\delta_i\le\sum_{i=2}^{j}d(S_i,D). \tag{34}$$

THEOREM 5. *The route $r_N=(1,2,\ldots,j)$ for a carpooling group $N=\{1,2,\ldots,j\}$ is SIR-feasible if and only if*

$$\left(1+\sum_{k=1}^{i-1}\frac{\alpha_k}{c}\right)\delta_i\le d(S_i,D),\quad\forall\ i\in\{2,3,\ldots,j\}. \tag{35}$$

Theorems 4 and 5 provide necessary and sufficient conditions for the *existence* of a budget-balanced cost sharing scheme that is IR/SIR on a route, by establishing upper bounds on (a linear combination of) the *incremental* detours due to successive commuters. If it is desired, for practical reasons, that the cost sharing scheme also be nonnegative, that is, no commuter gets paid to carpool, then, in addition, the *total* detour experienced by each commuter must also be bounded above. Theorem 6 formalizes this “add-on” condition.

THEOREM 6. *The route $r_N=(1,2,\ldots,j)$ for a carpooling group $N=\{1,2,\ldots,j\}$ admits a nonnegative, budget-balanced cost sharing scheme that is IR (respectively, SIR) on r_N if and only if, in addition to* (34) *(respectively,* (35)*),*

$$\frac{\alpha_i}{c}\left(\sum_{k=i+1}^{j}\delta_k\right)\le d(S_i,D),\quad\forall\ i\in\{1,2,\ldots,j-1\}. \tag{36}$$

The intuition for the proofs of Theorems 4, 5, and 6 can be gleaned from a more careful analysis of Example 1. The formal proofs are deferred to Appendix EC.3.

5.2.1. Incremental detours on SIR-feasible routes. We now take a closer look at the upper bound on the permissible incremental detour due to the addition of commuter i to the ride on an SIR-feasible route, given by (35), namely

$$\delta_i \leq \frac{d(S_i, D)}{\left(1 + \sum_{k=1}^{i-1} \frac{\alpha_k}{c}\right)}.$$

This bound diminishes with increasing i and increasing proximity of commuter i to the destination, which means that as more commuters are picked up, the permissible additional detour to pick up yet another passenger keeps shrinking, which is natural. For the passengers in Example 1, Fig. 4 shows the evolution of the "SIR-feasible region" (points from which the next passenger can be picked up so that the resultant route is SIR-feasible) in Euclidean space, for different values of $\frac{\alpha_i}{c}$ for $i = 1, 2$.

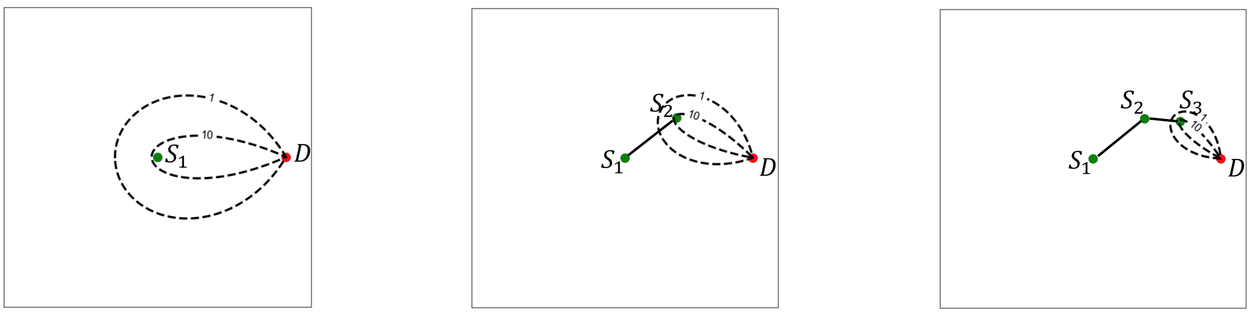

FIGURE 4. Evolution of the SIR-feasible regions (interior of the dashed curves) as more commuters are added to the ride, when $\frac{\alpha_1}{c} = \frac{\alpha_2}{c} = 1, 10$. Note that the regions keep shrinking with every subsequent commuter.

5.2.2. Bounds on total distance traveled along SIR-feasible routes. It may be useful to understand how the QoS-aware routes characterized by Theorems 4, 5, and 6 fare with respect to the maximum possible distance traveled by a commuter i (as a fraction of their distance to the destination, $d(S_i, D)$), which can be thought of as a worst-case measure of a commuter's inconvenience. We call this measure the *starvation factor* of commuter i. The starvation factor of a route is the maximum starvation factor among all the commuters. Intuitively, the starvation factor is a decreasing function of the ratios $\frac{\alpha_i}{c}$, since QoS-aware routes ensure that passengers that are more sensitive to detours suffer smaller detours. Our goal in this section is to quantify this intuition.

Let $\mathcal{I}(n)$ denote the space of all carpooling problem instances of size n (consisting of n source coordinates and a common destination point from an underlying metric space). Given an instance $p \in \mathcal{I}(n)$, let $\mathcal{R}(p)$ denote the set of all QoS-aware routes for this instance.

Given a QoS-aware route $r \in \mathcal{R}(p)$, let

$$\gamma_r(i) = \frac{\sum_{k=i}^{n-1} d(S_k, S_{k+1}) + d(S_n, D)}{d(S_i, D)} = 1 + \frac{\sum_{k=i+1}^{n} \delta_k}{d(S_i, D)} \tag{37}$$

denote the starvation factor of passenger i along route r, and let $\gamma_r = \max_{i \leq n} \gamma_r(i)$ denote the starvation factor of the route r.

DEFINITION 2. The QoS-aware *starvation factor* over all instances of size n is

$$\gamma(n) = \max_{p \in \mathcal{I}(n)} \min_{r \in \mathcal{R}(p)} \gamma_r.$$

It is straightforward to derive an upper bound for $\gamma(n)$ from Theorem 6, when a requirement of nonnegativity of the cost sharing scheme is imposed:

$$\gamma(n) \leq 1 + \frac{c}{\min_{i \leq n} \alpha_i}. \tag{38}$$

However, when the cost sharing scheme is not restrained to be nonnegative, characterizing the starvation factor on SIR-feasible routes is nontrivial, since it involves working with the individual bounds on the incremental detours from Theorem 5. Here, we show:

1. **Upper Bounds:** (Theorems 7-9) The worst starvation factor among SIR-feasible routes, $\max_{p \in \mathcal{I}(n)} \max_{r \in \mathcal{R}(p)} \gamma_r$, is (i) $\Theta(2^n)$ when $\frac{\alpha_i}{c} \to 0$, (ii) $\Theta(\sqrt{n})$ when $\frac{\alpha_i}{c} = 1$, and (iii) 1 when $\frac{\alpha_i}{c} \to \infty$, for all $i \leq n$. As upper bounds for $\gamma(n)$, these are not necessarily tight, since an instance for which an SIR-feasible route has the worst starvation factor may also admit other SIR-feasible routes with smaller starvation factors.
2. **Lower Bounds:** (Theorem 10) $\gamma(n)$ is no smaller than (i) $\Theta(n)$ when $\frac{\alpha_i}{c} \to 0$, and (ii) $\Theta(\log n)$ when $\frac{\alpha_i}{c} = 1$, for all $i \leq n$. These lower bounds are tight.

It is interesting to note that the gap between the upper and lower bounds narrows down and vanishes as $\frac{\alpha_i}{c}$ increases to ∞. (From (37), 1 is always a trivial lower bound for the starvation factor of any route.) The proofs are involved, and are deferred to Appendix EC.3.

We begin by establishing an almost obvious result that when commuters are infinitely inconvenienced by even the smallest of detours, (frankly, why would such passengers even consider carpooling?) the only SIR-feasible routes are those with zero detours, which implies a starvation factor of 1.

Theorem 7. *If $\frac{\alpha_i}{c} \to \infty$ for all $i \leq n$, then $\gamma_r = 1$ for any SIR-feasible route r.*

Next, we consider commuters who value their time more than c, and show that the worst they would have to endure is a *sublinear* starvation factor, in particular, $\Theta(\sqrt{n})$. This is tight when $\alpha_i = c$ for all $i \leq n$, in the sense that there exists an SIR-feasible route with $\Theta(\sqrt{n})$ starvation factor. However, as the α_i keep increasing beyond c, this bound becomes looser, culminating in a $\Theta(\sqrt{n})$ gap when $\alpha_i \to \infty$, as evidenced by Theorem 7.

Theorem 8. *If $\frac{\alpha_i}{c} \geq 1$ for all $i \leq n$, then $\gamma_r \leq 2\sqrt{n}$ for any SIR-feasible route r.*

Even though it may be unrealistic, as an academic exercise, we investigate an upper bound on γ_r when the passengers are completely unaffected by detours, that is, $\frac{\alpha_i}{c} \to 0$ for all $i \leq n$. Not surprisingly, it turns out that the starvation factor can be exponentially large in such a scenario, as the next theorem shows.

Theorem 9. *If $\frac{\alpha_i}{c} \to 0$ for all $i \leq n$, then $\gamma_r \leq 2^n$ for any SIR-feasible route r.*

The upper bounds of Theorems 8-9 on γ_r are tight, as discussed next; however, by Definition 2, they also serve as upper bounds on $\gamma(n)$, in which capacity, they may not necessarily be tight. This is because, an instance for which an SIR-feasible route has the worst starvation factor may also admit better SIR-feasible routes. For example, Figure 5 depicts an instance in one-dimensional Euclidean space for which the route $(S_1, S_2, \ldots, S_n, D)$ is SIR-feasible (satisfying (35) with equality) and has a starvation factor of $\Theta(\sqrt{n})$. (The same instance with the distances appropriately modified illustrates the $\Theta(2^n)$ starvation factor of Theorem 9.) However, note that the reverse route $(S_n, S_{n-1}, \ldots, S_1, D)$ is also SIR-feasible and has a starvation factor of 1.

Finally, we establish a tight lower bound on $\gamma(n)$ for arbitrary $\alpha_i > 0$, by exhibiting an instance with a unique SIR-feasible route with the desired starvation factor.

Theorem 10. $\gamma(n) \geq \sum_{i=1}^{n} \left(1 + \sum_{k=1}^{i-1} \frac{\alpha_k}{c}\right)^{-1}$.

It is easy to observe that the lower bound of Theorem 10 simplifies to $\Theta(\log n)$ when $\frac{\alpha_i}{c} = 1$, and $\Theta(n)$ when $\frac{\alpha_i}{c} \to 0$, for all $i \leq n$.

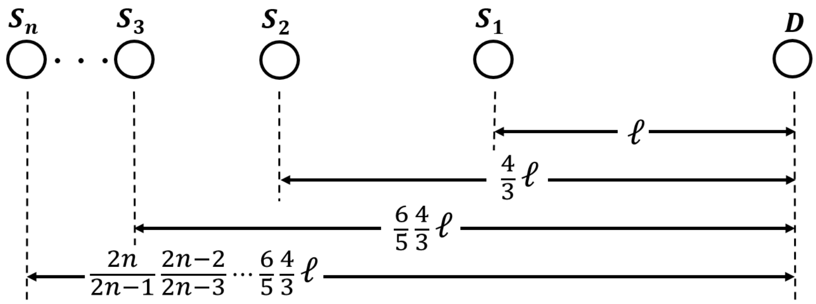

FIGURE 5. Carpooling instance with a route $(S_1, S_2, \ldots, S_n, D)$ whose starvation factor is $\Theta(\sqrt{n})$. If the distances $d(S_i, D)$, $i \leq n$, were $2^{i-1}\ell$ instead, then the starvation factor of the same route would be $\Theta(2^n)$.

5.3. The benefit of carpooling and sequential fairness. Under a cost sharing scheme that is IR, the decrease in disutility to a customer (the difference between the right and left hand sides of (31)) can be viewed as their benefit from carpooling. Further, it can be seen that the total benefit of sharing, obtained by summing the individual benefits, is independent of the cost sharing scheme, as long as it is budget-balanced. This observation exposes an underlying "duality"—a cost sharing scheme can, in fact, be viewed as a *benefit sharing scheme*. Such a view invites defining cost sharing schemes based on traditional notions of fairness, e.g., a fair cost sharing scheme should induce a distribution of the total benefit among the carpooling commuters suitably proportionately. There is a vast literature within cooperative game theory discussing fair cost sharing [27].

We extend this notion to budget-balanced cost sharing schemes that are SIR by investigating how they distribute the total *incremental* benefit due to each subsequent commuter arriving into the system, leading to a natural definition of *sequential fairness*.

DEFINITION 3. When commuter $i \in N$ joins the ride, the *incremental benefit* to commuters $k \leq i$ is given by

$$\mathcal{IB}(k,i,N) = \begin{cases} \mathcal{DU}(k, N(i-1)) - \mathcal{DU}(k, N(i)), & k < i \\ cd(S_i, D) - \mathcal{DU}(i, N(i)), & k = i \end{cases} = \begin{cases} f(k, N(i-1)) - f(k, N(i)) - \alpha_k \delta_i, & k < i \\ cd(S_i, D) - f(i, N(i)), & k = i. \end{cases} \quad (39)$$

DEFINITION 4. When commuter $i \in N$ joins the ride, the *total incremental benefit* to commuters $k \leq i$ is given by

$$\begin{aligned} \mathcal{TIB}(i,N) = \sum_{k=1}^{i} \mathcal{IB}(k,i,N) &= \sum_{k=1}^{i-1} f(k, N(i-1)) - \sum_{k=1}^{i} f(k, N(i)) + cd(S_i, D) - \delta_i \sum_{k=1}^{i-1} \alpha_k \\ &= c\left(d(S_i, D) - \left(1 + \sum_{k=1}^{i-1} \frac{\alpha_k}{c}\right)\delta_i\right). \end{aligned} \quad (40)$$

We take a very general, but minimal, approach to defining sequential fairness. All that is required of a cost sharing scheme to be sequentially fair is that, when commuter $i \in N$ joins the ride, the portion of the total incremental benefit that is enjoyed by a commuter $k < i$ is proportional to the incremental inconvenience cost to k due to i. This is formalized in the following definition.

DEFINITION 5. Given a vector $\vec{\beta} = (\beta_2, \beta_3, \ldots, \beta_j)$, where $0 \leq \beta_i \leq 1$ for $2 \leq i \leq j$, a budget-balanced, SIR cost sharing scheme f is $\vec{\beta}$*-sequentially fair* if, for all $2 \leq i \leq j$,

$$\frac{\mathcal{IB}(k,i,N)}{\mathcal{TIB}(i,N)} = \begin{cases} \beta_i \frac{\mathcal{IC}(k,N(i)) - \mathcal{IC}(k,N(i-1))}{\sum_{m=1}^{i-1}(\mathcal{IC}(m,N(i)) - \mathcal{IC}(m,N(i-1)))}, & k < i \\ 1 - \beta_i, & k = i \end{cases} = \begin{cases} \beta_i \frac{\alpha_k}{\sum_{m=1}^{i-1} \alpha_m}, & k < i \\ 1 - \beta_i, & k = i. \end{cases} \quad (41)$$

Here, $1 - \beta_i$ denotes the fraction of the total incremental benefit enjoyed by commuter i as a result of joining the ride, and β_i denotes the remaining fraction, which is split among the previous commuters in proportion to their α_k values.

It turns out that the requirements imposed by Definition 5, while perhaps appearing to be quite lenient, are sufficient for a strong and meaningful characterization of sequentially fair cost sharing schemes, as we discuss next.

5.3.1. Characterizing sequentially fair cost sharing schemes. We begin with a theorem that provides an exact characterization of budget-balanced sequentially fair cost sharing schemes.

THEOREM 11. *Given a vector $\vec{\beta} = (\beta_2, \beta_3, \ldots, \beta_j)$, where $0 \le \beta_i \le 1$ for $2 \le i \le j$, a budget-balanced cost sharing scheme f is $\vec{\beta}$-sequentially fair if and only if, for all $2 \le i \le j$,*

- *The cost to commuter i is given by*

$$f(i, N(i)) = \beta_i \big[cd(S_j, D)\big] + (1-\beta_i)\left[c\left(1 + \sum_{m=1}^{i-1} \frac{\alpha_k}{c}\right)\delta_i\right]. \tag{42}$$

- *The incremental "discount" to each previous commuter $k < i$ is given by*

$$f(k, N(i-1)) - f(k, N(i)) = \beta_i \left[\frac{\alpha_k}{\sum_{m=1}^{i-1} \alpha_m}\left(cd(S_i, D) - c\delta_i\right)\right] + (1-\beta_i)\big[\alpha_i \delta_i\big]. \tag{43}$$

We omit the proof, since it is a straightforward substitution of equations (39)-(40) in Definition 5 and rearrangement of the terms. The characterization of Theorem 11 reveals elegant structural properties of sequentially fair cost sharing schemes:

(a) **Online Implementation for Dynamic Carpooling:** When a new commuter i is picked up, their estimated cost is given by $f(i, N(i))$, which is their final payment if there are no more commuters. At the same time, each existing commuter $k < i$ obtains a discount in the amount of $f(k, N(i-1)) - f(k, N(i))$ that brings down their previous cost estimates. This suggests a novel *reverse-meter* design for a dynamic carpooling mobile application on each commuter's smartphone that keeps track of their estimated costs, as the ride progresses. Starting with $f(i, N(i))$ when commuter i begins their ride, the estimate would keep decreasing every time a detour begins to pick up the next commuter. Such a visually compelling interface would encourage increased participation in carpooling programs.

(b) **Convex Combination of Extreme Schemes:** For each i, $2 \le i \le j$, the cost sharing scheme is a convex combination of the following two extreme schemes:
   - *The total incremental benefit is fully enjoyed by the new commuter i, i.e., $\beta_i = 0$.* Here, from (42)-(43), the new commuter i (a) pays an amount $c\delta_i$ that corresponds to the increase in the operational cost, and (b) pays each existing commuter $k < i$ an amount $\alpha_k \delta_i$ that corresponds to the incremental inconvenience cost they suffered.
   - *The total incremental benefit is fully enjoyed by the existing commuters $k < i$, i.e., $\beta_i = 1$.* Here, from (42)-(43), the new commuter i pays $cd(S_j, D)$, the same as it would have cost them if they had driven their own car to the destination. From this, a portion $c\delta_i$ that corresponds to the increase in the operational cost is set aside, and what is left is split among the existing commuters in proportion to their α_k values.

   Note that the new commuter i pays the least in the former scheme ($\beta_i = 0$) and the most in the latter scheme ($\beta_i = 1$).

(c) **Transfers Between Commuters:** From the previous observation, it follows that a new commuter must, at minimum, fully compensate existing commuters for the incremental inconvenience costs that resulted from the detour to serve them, which can be viewed as internal transfers between passengers. Even though it may be reasonable to expect this (in an axiomatic sense) from a fair cost sharing scheme, it is remarkable that our notion of sequential fairness *mandates* this property.

In designing a sequentially fair cost sharing scheme, $\vec{\beta}$ can be chosen strategically to incentivize commuters to participate in dynamic carpooling, e.g., setting β_2 large enough to encourage bootstrapping when there is a shortage of available rides to meet the demand. We end this section with an example.

EXAMPLE 2. Let $\alpha_i = c = 1 \;\forall\; i \in N$. For $1 \le k \le i \le j$, the cost sharing scheme $f^{\mathtt{XC}}$ is:

$$f^{\mathtt{XC}}(k, N(i)) = \left(\sum_{m=k+1}^{i} \frac{d(S_{m-1}, S_m)}{m-1} + \frac{d(S_i, D)}{i} \right) + (k-1)\delta_k - \left(\sum_{m=k+1}^{i} \delta_m \right).$$

The first two terms correspond to dividing the operational cost of each segment equally among the commuters traveling along that segment. The third term corresponds to commuter k compensating each of the $k-1$ previous commuters, for the incremental detour they suffered. The last term corresponds to the net compensation received by commuter k from all future commuters, for the incremental detours that k suffered.

Intuitively, it could be argued that $f^{\mathtt{XC}}$ is a "fair" cost sharing scheme. In our framework, it can be shown that for $\vec{\beta} = \left(\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{j}\right)$, it is a $\vec{\beta}$-sequentially fair cost sharing scheme:

From (42)-(43), we get

$$\frac{\mathcal{IB}(i,i,N)}{\mathcal{TIB}(i,N)} = \frac{d(S_i, D) - f^{\mathtt{XC}}(i, N(i))}{d(S_i, D) - i\delta_i} = \frac{d(S_i, D) - \left(\frac{d(S_i,D)}{i} + (i-1)\delta_i\right)}{d(S_i, D) - i\delta_i} = \frac{i-1}{i} = 1 - \frac{1}{i},$$

as desired. Also, for $k < i$, we get

$$\begin{aligned} \frac{\mathcal{IB}(k,i,N)}{\mathcal{TIB}(i,N)} &= \frac{f^{\mathtt{XC}}(k, N(i-1)) - f^{\mathtt{XC}}(k, N(i)) - \delta_i}{d(S_i, D) - i\delta_i} = \frac{\frac{d(S_{i-1},D)}{i-1} - \left(\frac{d(S_{i-1},S_i)}{i-1} + \frac{d(S_i,D)}{i}\right)}{d(S_i, D) - i\delta_i} \\ &= \frac{\frac{d(S_i,D)}{i(i-1)} - \frac{1}{i-1}\delta_i}{d(S_i, D) - i\delta_i} = \frac{1}{i}\frac{1}{i-1}. \end{aligned}$$

---

6. Concluding remarks. By thrusting an individual user to the center, our framework infuses behavioral QoS models into traditional economic objectives to unearth key operational insights at the microscopic unit of a single ride. The natural next step in this bottom-up approach is to understand how these operational constraints interact across multiple rides, over a large network, and with varying demand patterns. A real-world, data-driven simulation of a ridesharing system (see, e.g., [56]) that incorporates, e.g., the SIR-feasible routing constraints (35) would be a good starting point. Revisiting computational questions surrounding traditional Vehicle Routing Problems (VRPs) in light of these constraints is also worth exploring. (We present an extended discussion of new algorithmic questions inspired by SIR-feasibility in Appendix EC.4.)

From a broader perspective, our work can perhaps be viewed as a connecting piece in the complex puzzle of guiding and facilitating sustainable urbanization. How strongly can utilitarian and behavioral QoS-awareness at a unit level influence key trade-offs between commercial (profit), societal (welfare), and environmental (vehicle-miles) objectives at the system-level? What coordinated individual incentive schemes (that affect users' utilities) and industry-wide policy interventions (that affect commercial objectives) can a government entity implement to best regulate such trade-offs?

We believe that our framework can be extended more generally to other dynamic resource and service sharing systems such as contact centers [5], cloud computing [4], and shared logistics in supply chain distribution networks [11]. Operational and economic policies impact the quality of shared service in such systems, wherein users may experience a sequence of utilities every time the state of the system changes (e.g., due to new arrivals/departures, addition/removal of capacity). Human behavioral effects induced by the environment determine whether users are frustrated or satisfied with their temporal utility sequence. Appropriate notions of QoS can then capture these effects and internalize them into the operational/economic performance analysis to yield optimal QoS-aware policies for the system.

As an example, consider the following question: What QoS-aware and/or fair routing and staffing policies would result when taking into account the waiting experience of customers in a multi-server queueing system? Our framework would direct one to begin with a review of relevant literature from behavioral operations management on how customers perceive waiting in queues (see, e.g., [53]) to model "state-dependent" utilities for the users and define an appropriate (probabilistic) variant of SIR. The same question can also be asked from the point of view of the service experience of human servers (see, e.g., [26]).

The 'duality' between cost sharing and benefit sharing in our framework (Section 5.3) is worth a deeper analysis. While the space of cost sharing schemes that the two views accommodate are no different from each other, there is a crucial difference in approaching their design. In particular, a budget-balanced cost sharing scheme need only recover the operational costs; see (26). The inconvenience costs experienced by the commuters are a separate artifact of our QoS-focused framework, which only explicitly affect the design of cost sharing schemes when viewed through the lens of benefit sharing and sequential fairness. What *traditional* fairness properties does a sequentially fair cost sharing scheme possess? For example, under what conditions, if any, is it equivalent to the Shapley value, or is in the core of a cost sharing game?

Finally, we note that throughout, we have assumed knowledge of key elements of users' (dis)utility ($k_i(\cdot)$ in commercial ridesharing, α_i in community carpooling). In reality, they most likely need to be estimated empirically, or elicited directly from the users. In the latter case, users' reports may not be accurate due to privacy or strategic concerns. It would be interesting to study the trade-offs between efficiency, fairness, budget-balance, and incentive compatibility in such scenarios, by suitably integrating our framework with that of online mechanism design [48, 58].

References

[1] Acker, Amelia, Dhiraj Murthy. 2018. Venmo: Understanding mobile payments as social media. *Proceedings of the 9th International Conference on Social Media and Society*. 5–12.

[2] Agatz, Niels, Alan Erera, Martin Savelsbergh, Xing Wang. 2012. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research* **223**(2) 295–303.

[3] Agatz, Niels AH, Alan L Erera, Martin WP Savelsbergh, Xing Wang. 2011. Dynamic ride-sharing: A simulation study in metro Atlanta. *Transportation Research Part B: Methodological* **45**(9) 1450–1464.

[4] Ahn, Jeongseob, Changdae Kim, Jaeung Han, Young-ri Choi, Jaehyuk Huh. 2012. Dynamic virtual machine scheduling in clouds for architectural shared resources. *Proceedings of the 4th USENIX conference on Hot Topics in Cloud Computing*.

[5] Aksin, Zeynep, Mor Armony, Vijay Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.

[6] Banerjee, Siddhartha, Ramesh Johari, Carlos Riquelme. 2015. Pricing in ride-sharing platforms: a queueing-theoretic approach. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. 639–639.

[7] Benjaafar, S., H. Bernhard, C. Courcoubetis. 2017. Drivers, Riders and Service Providers: The Impact of the Sharing Economy on Mobility. Available at SSRN: `https://ssrn.com/abstract=3035478`.

[8] Bimpikis, K., O. Candogan, S. Daniela. 2016. Spatial Pricing in Ride-Sharing Networks. Working Paper. Available at SSRN: `https://ssrn.com/abstract=2868080`.

[9] Bistaffa, Filippo, Alessandro Farinelli, Georgios Chalkiadakis, Sarvapali Ramchurn. 2015. Recommending fair payments for large-scale social ridesharing. *Proceedings of the 9th ACM Recommender Systems Conference (RecSys)*. 139–146.

[10] Biswas, Arpita, Ragavendran Gopalakrishnan, Theja Tulabandhula, Koyel Mukherjee, Asmita Metrewar, Raja Subramaniam Thangaraj. 2017. Profit optimization in commercial ridesharing. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 1481–1483.

[11] Bowersox, Donald J, David J Closs, Theodore P Stank. 2000. Ten mega-trends that will revolutionize supply chain logistics. *Journal of Business Logistics* **21**(2) 1.

[12] Cao, Bin, Louai Alarabi, Mohamed F Mokbel, Anas Basalamah. 2015. SHAREK: A scalable dynamic ride sharing system. *Proceedings of the 16th IEEE International Conference on Mobile Data Management (MDM)*. 4–13.

[13] Caraway, Monica, Daniel A. Epstein, Sean A. Munson. 2017. Friends don't need receipts: The curious case of social awareness streams in the mobile payment app Venmo. *Proceedings of the ACM Humam-Computer Interaction* **1**(CSCW) 28:1–28:17.

[14] Chan, Nelson D, Susan A Shaheen. 2012. Ridesharing in North America: Past, present, and future. *Transport Reviews* **32**(1) 93–112.

[15] Cohen, M., Fiszer M.-D., B. J. Kim. 2018. Frustration-based promotions: Field experiments in ride-sharing. Working Paper. Available at SSRN: https://ssrn.com/abstract=3129717.

[16] Cordeau, Jean-François, Gilbert Laporte, Martin WP Savelsbergh, Daniele Vigo. 2006. Vehicle routing. *Handbooks in operations research and management science: Transportation*, vol. 14, chap. 6. Elsevier, 367–428.

[17] Das Gupta, Aparupa, Uday S Karmarkar, Guillaume Roels. 2015. The design of experiential services with acclimation and memory decay: Optimal sequence and duration. *Management Science* **62**(5) 1278–1296.

[18] Dixon, Michael J, Liana Victorino. 2019. The sequence of service: An affect perspective to service scheduling. *Handbook of Service Science, Volume II*. Springer, 49–76.

[19] Esteban, Joan. 1991. *The social viability of money (competitive equilibria and the core of overlapping generations economies)*, vol. 372. Springer-Verlag.

[20] Ferguson, Erik. 1990. The influence of employer ridesharing programs on employee mode choice. *Transportation* **17**(2) 179–207.

[21] Freed, Benjamin. 2015. Why You Shouldn't Call Uber and Lyft "Ride-Sharing". URL http://www.washingtonian.com/blogs/capitalcomment/transportation/why-you-shouldnt-call-uber-and-lyf

[22] Geisberger, Robert, Dennis Luxen, Sabine Neubauer, Peter Sanders, Lars Volker. 2010. Fast detour computation for ride sharing. *Proceedings of the 10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, OASIcs-OpenAccess Series in Informatics*, vol. 14. 88–99.

[23] Ghiani, Gianpaolo, Francesca Guerriero, Gilbert Laporte, Roberto Musmanno. 2003. Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research* **151**(1) 1–11.

[24] He, Wen, Kai Hwang, Deyi Li. 2014. Intelligent carpool routing for urban ridesharing by mining GPS trajectories. *IEEE Transactions on Intelligent Transportation Systems* **15**(5) 2286–2296.

[25] Jacob, J., R. Roet-Green. 2018. Ride Solo or Pool: The Impact of Sharing on Optimal Pricing of Ride-Sharing Services. Working Paper. Available at SSRN: https://ssrn.com/abstract=3008136.

[26] Jaeker, J. A. B., A. L. Tucker. 2017. Past the Point of Speeding Up: The Negative Effects of Workload Saturation on Efficiency and Patient Severity. *Management Science* **63**(4) 1042–1062.

[27] Jain, K., M. Mahdian. 2007. Cost sharing. N. Nisan, T. Roughgarden, É. Tardos, V. V. Vazirani, eds., *Algorithmic Game Theory*, chap. 15. Cambridge University Press, 385–410.

[28] Jung, Jaeyoung, R Jayakrishnan, Ji Young Park. 2013. Design and modeling of real-time shared-taxi dispatch algorithms. *Proceedings of the Transportation Research Board 92nd Annual Meeting*.

[29] Kamar, Ece, Eric Horvitz. 2009. Collaboration and shared plans in the open world: Studies of ridesharing. *IJCAI*. 187–194.

[30] Kanoria, Yash, Ilan Lobel, Jiaqi Lu. 2018. Managing customer churn via service mode control. *SSRN id 3188226* .

[31] Kleiner, Alexander, Bernhard Nebel, Vittorio Amos Ziparo. 2011. A mechanism for dynamic ride sharing based on parallel auctions. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*. 266–272.

[32] Levinger, C., A. Azaria, N. Hazon. 2018. Human Satisfaction as the Ultimate Goal in Ridesharing. Working Paper. Available at: https://arxiv.org/abs/1807.00376.

[33] Li, Shu, Fan Fei, Dou Ruihan, Shui Yu, Wanchun Dou. 2016. A dynamic pricing method for carpooling service based on coalitional game analysis. *International Conference on Smart City, Data Science and Systems, and High Performance Computing and Communications*. 78–85.

[34] Li, Yifu, Tinglong Dai, Xiangtong Qi. 2018. Service design under acclimation and non-homogeneous memory decay. *SSRN id 3194758* .

[35] Mallus, M., G. Colistra, L. Atzori, M. Murroni, V. Pilloni. 2017. Dynamic carpooling in urban areas: Design and experimentation with a multi-objective route matching algorithm. *Sustainability* **9**(2) 254.

[36] Molenbruch, Y., K. Braekers, A. Caris. 2017. Operational effects of service level variations for the dial-a-ride problem. *Central European Journal of Operations Research* **25**(1) 71–90.

[37] Moulin, Hervé. 2004. *Fair division and collective welfare*. MIT press.

[38] Myerson, Roger B. 1981. Optimal auction design. *Mathematics of Operations Research* **6**(1) 58–73.

[39] Narahari, Y, Ramasuri Narayanam, Dinesh Garg, Hastagiri Prakash. 2009. Foundations of mechanism design. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions, Part of the Series on Advanced Information and Knowledge Processing*. Springer, 1–131.

[40] Neoh, Jun Guan, Maxwell Chipulu, Alasdair Marshall. 2017. What encourages people to carpool? An evaluation of factors with meta-analysis. *Transportation* **44**(2) 423–447.

[41] Nguyen, Duc Thien. 2013. Fair cost sharing auction mechanisms in last mile ridesharing. Ph.D. thesis, Singapore Management University.

[42] Ninikas, George, Ioannis Minis. 2014. Reoptimization strategies for a dynamic vehicle routing problem with mixed backhauls. *Networks* **64**(3) 214–231.

[43] Olszewski, R., P. Pałka, A. Turek. 2018. Solving "Smart City" Transport Problems by Designing Carpooling Gamification Schemes with Multi-Agent Systems: The Case of the So-Called "Mordor of Warsaw". *Sensors (Basel)* **18**(1) 141.

[44] Örmeci, E Lerzan, F Sibel Salman, Eda Yücel. 2014. Staff rostering in call centers providing employee transportation. *Omega* **43** 41–53.

[45] Papadimitriou, C. H. 1977. The Euclidean travelling salesman problem is NP-complete. *Theoretical Computer Science* **4**(3) 237–244.

[46] Papadimitriou, C. H. 1994. *Computational Complexity*. Addison-Wesley.

[47] Paquette, J., J.-F. Cordeau, G. Laporte. 2009. Quality of service in dial-a-ride operations. *Computers & Industrial Engineering* **56**(4) 1721–1734.

[48] Parkes, D. C. 2007. Online mechanisms. N. Nisan, T. Roughgarden, É. Tardos, V. V. Vazirani, eds., *Algorithmic Game Theory*, chap. 15. Cambridge University Press, 411–439.

[49] Pelzer, Dominik, Jiajian Xiao, Daniel Zehe, Michael H Lees, Alois C Knoll, Heiko Aydt. 2015. A partition-based match making algorithm for dynamic ridesharing. *IEEE Transactions on Intelligent Transportation Systems* **16**(5) 2587–2598.

[50] Pillac, Victor, Michel Gendreau, Christelle Guéret, Andrés L Medaglia. 2013. A review of dynamic vehicle routing problems. *European Journal of Operational Research* **225**(1) 1–11.

[51] Potter, Stephen, Marcus Enoch, Tom Rye, Colin Black, Barry Ubbels. 2006. Tax treatment of employer commuting support: An international review. *Transport Reviews* **26**(2) 221–237.

[52] Prins, Christian, Philippe Lacomme, Caroline Prodhon. 2014. Order-first split-second methods for vehicle routing problems: A review. *Transportation Research Part C: Emerging Technologies* **40** 179–200.

[53] Rafaeli, A., G. Barron, K. Haber. 2002. The Effects of Queue Structure on Attitudes. *Journal of Service Research* **5**(2) 125–139.

[54] Rajkumar, Arun, Koyel Mukherjee, Theja Tulabandhula. 2017. Learning to partition using score based compatibilities. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 574–582.

[55] Roels, Guillaume. 2019. Optimal structure of experiential services: Review and extensions. *Handbook of Service Science, Volume II*. Springer, 105–146.

[56] Santi, Paolo, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H Strogatz, Carlo Ratti. 2014. Quantifying the benefits of vehicle pooling with shareability networks. *PNAS* **111**(37) 13290–13294.

[57] Santos, Douglas Oliveira, Eduardo Candido Xavier. 2013. Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 13. 2885–2891.

[58] Shen, Wen, Cristina V. Lopes, Jacob W. Crandall. 2016. An online mechanism for ridesharing in autonomous mobility-on-demand systems. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 475–481.

[59] Taylor, T. A. 2018. On-Demand Service Platforms. *Manufacturing & Service Operations Management* **20**(4) 704–720.

[60] Tversky, Amos, Daniel Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics* **106**(4) 1039–1061.

[61] Uber. 2018. Uber API price estimates. https://developer.uber.com/docs/riders/references/api/v1.2/estimates Accessed: 2018-11-02.

[62] United Nations. 2018. World Urbanization Prospects: The 2018 Revision. Available at: https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf.

[63] Wang, Xing, Niels Agatz, Alan Erera. 2017. Stable matching for dynamic ride-sharing systems. *Transportation Science* **52** 739–1034.

[64] Yu, Q., G. Allon, A. Bassamboo. 2017. The Reference Effect of Delay Announcements: A Field Experiment. Working Paper. Available at SSRN: https://ssrn.com/abstract=2945535.

[65] Zhao, Dengji, Dongmo Zhang, Enrico H Gerding, Yuko Sakurai, Makoto Yokoo. 2014. Incentives in ridesharing with deficit control. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. 1021–1028.

Sequential Individual Rationality in Dynamic Ridesharing: Technical Appendix

Ragavendran Gopalakrishnan, Theja Tulabandhula, and Koyel Mukherjee

In this technical appendix, we provide proofs for the results stated in the main body of the manuscript titled: "Sequential Individual Rationality in Dynamic Ridesharing". The proofs of these results are in the order in which they appear in the main body.

EC.1. Proofs from Section 4.3. Before presenting the proofs, we develop some of the common technical machinery, beginning with the derivatives of the valuation thresholds $\underline{v}_j$ and $\overline{v}_j$, given by (6). Their first order partial derivatives with respect to the prices p_j^x and p_j^s are given by:

$$\frac{\partial \underline{v}_j}{\partial p_j^x} = 0, \qquad \frac{\partial \underline{v}_j}{\partial p_j^s} = \frac{1}{k_j(\hat{\delta}_j)d(S_j, D_j)}, \tag{EC.1}$$

$$\frac{\partial \overline{v}_j}{\partial p_j^x} = -\frac{\partial \overline{v}_j}{\partial p_j^s} = \frac{1}{(1-k_j(\hat{\delta}_j))d(S_j, D_j)}. \tag{EC.2}$$

Using the above, we derive the first order partial derivatives of the service provider's expected incremental profit $Exp\Delta P_j$ (given by (20)), with respect to the prices p_j^x and p_j^s:

$$\frac{\partial Exp\Delta P_j}{\partial p_j^x} = -f_v(\overline{v}_j)\left(\phi_v(\overline{v}_j) - \frac{c\left(d(S_j,D_j) - \Delta_j^s - \Delta_j^d\right) - \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j}{(1-k_j(\hat{\delta}_j))d(S_j,D_j)}\right), \tag{EC.3}$$

$$\frac{\partial Exp\Delta P_j}{\partial p_j^s} = -\frac{\partial Exp\Delta P_j}{\partial p_j^x} - f_v(\underline{v}_j)\left(\phi_v(\underline{v}_j) - \frac{c\left(\Delta_j^s + \Delta_j^d\right) + \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j}{k_j(\hat{\delta}_j)d(S_j,D_j)}\right). \tag{EC.4}$$

Let $\underline{v}_j^*$ and $\overline{v}_j^*$ denote the valuation thresholds evaluated at the optimal prices $p_j^{x,*}$ and $p_j^{s,*}$. From Lemma 1, $p_j^{x,*} \in [0, \infty)$, and $p_j^{s,*} \in \left[0, k_j(\hat{\delta}_j)p_j^{x,*}\right]$ are unique. Therefore, $0 \le \underline{v}_j^* \le \overline{v}_j^* < \infty$.

EC.1.1. Proof of Theorem 1. From (21), $ProbSharing_j^* > 0$ if and only if $\underline{v}_j^* < \overline{v}_j^*$. Thus, the optimal prices $p_j^{x,*}$ and $p_j^{s,*}$ must be interior maximizers, and $\underline{v}_j^*$ and $\overline{v}_j^*$ must satisfy $\frac{\partial Exp\Delta P_j}{\partial p_j^x} = 0$ and $\frac{\partial Exp\Delta P_j}{\partial p_j^s} = 0$ simultaneously. Since ϕ_v is a strictly increasing function, $\underline{v}_j^* < \overline{v}_j^*$ if and only if $\phi_v(\underline{v}_j^*) < \phi_v(\overline{v}_j^*)$. From (EC.3)-(EC.4), this yields

$$\frac{c\left(\Delta_j^s + \Delta_j^d\right) + \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j}{k_j(\hat{\delta}_j)d(S_j,D_j)} < \frac{c\left(d(S_j,D_j) - \Delta_j^s - \Delta_j^d\right) - \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j}{(1-k_j(\hat{\delta}_j))d(S_j,D_j)}$$

$$\iff \quad c\left(k_j(\hat{\delta}_j)d(S_j,D_j) - (\Delta_j^s + \Delta_j^d)\right) > \beta\sum_{i=1}^{j-1} Any\Delta\mathtt{Penalty}_i^j.$$

∎

EC.1.2. Proof of Theorem 3. When $ProbSharing_j^* > 0$, the expressions for $\underline{v}_j^*$ and $\overline{v}_j^*$ are obtained from setting $\frac{\partial Exp\Delta P_j}{\partial p_j^x} = 0$ and $\frac{\partial Exp\Delta P_j}{\partial p_j^s} = 0$, after which the corresponding optimal prices $p_j^{x,*}$ and $p_j^{s,*}$ can be extracted from (6).

When $ProbSharing_j^* = 0$, the service provider's expected incremental profit, from (20), simplifies to

$$Exp\Delta P_j(p_j^x) = \left(1 - F_v(p_j^x)\right)\left(p_j^x - c\, d(S_j, D_j)\right),$$

which is only a function of p_j^x. It is straightforward to solve the first order condition to obtain $p_j^{x,*} = \phi_v^{-1}(c)d(S_j, D_j)$. Since $ProbSharing_j^* = 0$, $\underline{v}_j^* = \overline{v}_j^*$, and so, $p_j^{s,*} = k_j(\hat{\delta}_j)p_j^{x,*} = \phi_v^{-1}(c)\, k_j(\hat{\delta}_j)d(S_j, D_j)$. ∎

EC.1.3. Proof of Theorem 2. Suppose we are given a destination point D_j^0 such that $D_j^0 \in \mathcal{R}_j^{OPT(0)}(S_j)$, but $D_j^0 \notin \mathcal{R}_j^{SIR}(S_j)$. This means that, for the request (S_j, D_j^0), from (23), we have

$$c\left(k_j(\hat{\delta}_j)d(S_j, D_j^0) - (\Delta_j^s + \Delta_j^d)\right) > 0, \tag{EC.5}$$

and $\sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j > 0$. In order to prove Theorem 2, given any $\beta > 0$, we need to exhibit the existence of a destination point D_j^β such that $D_j^\beta \in \mathcal{R}_j^{OPT(\beta)}(S_j)$, but $D_j^\beta \notin \mathcal{R}_j^{SIR}(S_j)$.

Let S_j be relabeled as D_0. It can be shown that the "greedy sequential insertion" routing algorithm (outlined in the beginning of Section 4 and in Section 4.1.3) ensures the following property. If D_j^0 is inserted between D_ℓ and $D_{\ell+1}$ (for some $0 \le \ell < j-1$) in the new route plan, then $d(D_k, D_j^0) \ge d(D_k, D_\ell)$ for all $0 \le k \le \ell$. This property guarantees the existence of a *trajectory* of destination points D_j (starting from D_j^0) along which $d(S_j, D_j)$ is constant while Δ_j^d and $\sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j$ decrease as the trajectory approaches the boundary of $\mathcal{R}_j^{SIR}(S_j)$, at which point, $\sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j = 0$. Thus, given any $\beta > 0$, there exists a point D_j^β on this trajectory (sufficiently close to the boundary of $\mathcal{R}_j^{SIR}(S_j)$) for which

$$c\left(k_j(\hat{\delta}_j)d(S_j, D_j^\beta) - (\Delta_j^s + \Delta_j^d)\right) > \beta \sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j, \tag{EC.6}$$

since $\sum_{i=1}^{j-1} Any\Delta\texttt{Penalty}_i^j$ can be made arbitrarily close to 0 along the trajectory while still staying outside $\mathcal{R}_j^{SIR}(S_j)$. ∎

EC.2. Strategic concerns regarding detour estimates. We now investigate whether the service provider might have an incentive to communicate to user j, an estimated detour $\hat{\delta}_j < \delta_j^j$ to "lure" the user into requesting a shared ride, knowing that it would cost the service provider a penalty. Intuitively, it must be that there exists a threshold for the QoS-sensitivity $\beta^\dagger > 0$ such that when $\beta < \beta^\dagger$, the provider should benefit from lying, whereas when $\beta > \beta^\dagger$, such behavior would not be profitable. In this section, we show that $\beta^\dagger < 1$.

When $\hat{\delta}_j < \delta_j^j$, the service provider's inremental profit when user j chooses `Shared`, $\Delta P_j(p_j^x, p_j^s; \hat{\delta}_j, \texttt{Shared})$ (given by (19)), must include an additional penalty term of $\beta Max\Delta\texttt{Penalty}_j^j$. From (15), $\Delta\texttt{Penalty}_j^j = -\min\{0, U_j^s(p_j^s, \delta_j^j)\}$. Then, from (18), the value of $Max\Delta\texttt{Penalty}_j^j$ is obtained by substituting $v_j = \underline{v}_j$, and $\underline{v}_j$ is given by (6). This yields

$$Max\Delta\texttt{Penalty}_j^j = -U_j^s(p_j^s, \delta_j^j) = p_j^s - k_j(\delta_j^j)\underline{v}_j d(S_j, D_j) = p_j^s\left(1 - \frac{k_j(\delta_j^j)}{k_j(\hat{\delta}_j)}\right). \tag{EC.7}$$

Thus, when $\hat{\delta}_j < \delta_j^j$, the service provider's expected incremental profit $Exp\Delta P_j$ (from (20)) is given by

$$\begin{aligned} Exp\Delta P_j(p_j^x, p_j^s, \hat{\delta}_j) = &\ (1 - F_v(\overline{v}_j))\left(p_j^x - c\ d(S_j, D_j)\right) \\ &+ \left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right)\left(p_j^s - c\left(\Delta_j^s + \Delta_j^d\right) - \beta \sum_{i=1}^{j} Max\Delta\mathtt{Penalty}_i^j\right) \\ = &\ (1 - F_v(\overline{v}_j))\left(p_j^x - c\ d(S_j, D_j)\right) \\ &+ \left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right)\left(p_j^s - \beta p_j^s\left(1 - \frac{k_j(\delta_j^j)}{k_j(\hat{\delta}_j)}\right) - c\left(\Delta_j^s + \Delta_j^d\right) - \beta \sum_{i=1}^{j-1} Max\Delta\mathtt{Penalty}_i^j\right) \end{aligned} \tag{EC.8}$$

Notice that $Exp\Delta P_j$ now depends on $\hat{\delta}_j$, in addition to p_j^x, p_j^s. Its first order partial derivatives with respect to p_j^x, p_j^s, and $\hat{\delta}_j$ are given by

$$\frac{\partial Exp\Delta P_j}{\partial p_j^x} = -f_v(\overline{v}_j)\left(\phi_v(\overline{v}_j) + \beta \underline{v}_j\ \frac{k_j(\hat{\delta}_j) - k_j(\delta_j^j)}{1 - k_j(\hat{\delta}_j)} - \frac{c\left(d(S_j, D_j) - \Delta_j^s - \Delta_j^d\right) - \beta \sum_{i=1}^{j-1} Max\Delta\mathtt{Penalty}_i^j}{(1 - k_j(\hat{\delta}_j))d(S_j, D_j)}\right), \tag{EC.9}$$

$$\begin{aligned} \frac{\partial Exp\Delta P_j}{\partial p_j^s} = -\frac{\partial Exp\Delta P_j}{\partial p_j^x}\left(1 + \frac{f_v(\underline{v}_j)}{f_v(\overline{v}_j)} \frac{1 - k_j(\hat{\delta}_j)}{k_j(\hat{\delta}_j)}\right) - f_v(\underline{v}_j)\left(\phi_v(\underline{v}_j) + \phi_v(\overline{v}_j)\frac{1 - k_j(\hat{\delta}_j)}{k_j(\hat{\delta}_j)} - \frac{c}{k_j(\hat{\delta}_j)}\right) \\ - \beta\left(\frac{k_j(\hat{\delta}_j) - k_j(\delta_j^j)}{k_j(\hat{\delta}_j)}\right)\left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right), \end{aligned} \tag{EC.10}$$

$$\begin{aligned} \frac{\partial Exp\Delta P_j}{\partial \hat{\delta}_j} = &\ k_j'(\hat{\delta}_j)d(S_j, D_j)\left((\overline{v}_j - \underline{v}_j)\frac{\partial Exp\Delta P_j}{\partial p_j^x} - \underline{v}_j\frac{\partial Exp\Delta P_j}{\partial p_j^s}\right) \\ &- k_j'(\hat{\delta}_j)d(S_j, D_j)\left((\overline{v}_j - \underline{v}_j)(1 - F_v(\overline{v}_j)) + (\beta - 1)\underline{v}_j\left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right)\right). \end{aligned} \tag{EC.11}$$

When $ProbSharing_j^* > 0$, the optimal prices $p_j^{x,*}$ and $p_j^{s,*}$ must be interior maximizers, and thus, $\frac{\partial Exp\Delta P_j}{\partial p_j^x} = 0$ and $\frac{\partial Exp\Delta P_j}{\partial p_j^s} = 0$ hold simultaneously, under which, (EC.11) becomes

$$\frac{\partial Exp\Delta P_j}{\partial \hat{\delta}_j} = -k_j'(\hat{\delta}_j)d(S_j, D_j)\left((\overline{v}_j - \underline{v}_j)(1 - F_v(\overline{v}_j)) + (\beta - 1)\underline{v}_j\left(F_v(\overline{v}_j) - F_v(\underline{v}_j)\right)\right).$$

Since k_j is a decreasing function, the above partial derivative is increasing in β, and positive when $\beta \geq 1$. Therefore, $\beta^\dagger$ must be less than 1. ∎

EC.3. Proofs from Section 5.

EC.3.1. Proof of Theorem 5. From the SIR constraints (33), we have that for all $l \in N$ (omitting the dependence on route for simplicity):

$$f(l, N(l)) \leq cd(S_l, D), \text{ and} \tag{EC.12}$$

$$f(l, N(m+1)) + \alpha_l \delta_{m+1} \leq f(l, N(m))\ \forall\ m \in \{l, l+1, \ldots, j-1\}. \tag{EC.13}$$

For any $i \in \{2, 3, ..., j\}$, the "only if" direction can be seen to hold by adding all $m = i-1$ related inequalities in (EC.13) and the inequality corresponding to $l = i$ in (EC.12):

$$\sum_{k=1}^{i} f(k, N(i)) - \sum_{k=1}^{i-1} f(k, N(i-1)) + \sum_{k=1}^{i-1} \alpha_k \delta_i \leq cd(S_i, D).$$

Using budget-balance to simplify the first two terms, we get

$$c\left(d(S_{i-1},S_i)+d(S_i,D)-d(S_{i-1},D)\right)+\sum_{k=1}^{i-1}\alpha_k\delta_i \le cd(S_i,D)$$

$$\Rightarrow c\delta_i+\sum_{k=1}^{i-1}\alpha_k\delta_i \le cd(S_i,D) \tag{EC.14}$$

$$\Rightarrow \left(1+\sum_{k=1}^{i-1}\frac{\alpha_k}{c}\right)\delta_i \le d(S_i,D).$$

Next, we prove the "if" direction. Assuming that (35) holds, it suffices to exhibit a budget-balanced cost sharing scheme f, under which all the SIR constraints given by (EC.12) and (EC.12) are satisfied.

For $1\le k\le j$, and $1\le i\le k$, we construct $f(i,N(k))$ recursively, so that (EC.12) and (EC.13) are satisfied. The base case follows from budget-balance, that is, $f(i,\{i\})=cd(S_i,D)$ for all $i\in N$. Assume that for some $2\le k\le j$, we have defined $f(i,N(k-1))$ for all $1\le i\le k-1$. Then, we set

$$f(i,N(k))=f(i,N(k-1))-\alpha_i\delta_k,\quad 1\le i\le k-1$$

$$f(k,N(k))=c\left(\sum_{l=1}^{k-1}d(S_l,S_{l+1})+d(S_k,D)\right)-\sum_{i=1}^{k-1}f(i,N(k)).$$

By construction, it follows that (EC.13) is satisfied, and f is budget-balanced. It remains to be shown that (EC.12) is also satisfied.

$$\begin{aligned}
f(k,N(k)) &= c\left(\sum_{l=1}^{k-1}d(S_l,S_{l+1})+d(S_k,D)\right)-\sum_{i=1}^{k-1}\left(f(i,N(k-1))-\alpha_i\delta_k\right)\\
&\overset{(\dagger)}{=} c\left(\sum_{l=1}^{k-1}d(S_l,S_{l+1})+d(S_k,D)-\sum_{l=1}^{k-2}d(S_l,S_{l+1})-d(S_{k-1},D)+\sum_{i=1}^{k-1}\frac{\alpha_i}{c}\delta_k\right)\\
&= c\left(d(S_{k-1},S_k)+d(S_k,D)-d(S_{k-1},D)+\sum_{i=1}^{k-1}\frac{\alpha_i}{c}\delta_k\right)\\
&= c\left(\delta_k+\sum_{i=1}^{k-1}\frac{\alpha_i}{c}\delta_k\right)\\
&\le cd(S_k,D),
\end{aligned}$$

where the last step follows from (35), and step (†) follows from budget-balance. ■

EC.3.2. Proof of Theorem 7. First, we note that in the limit, when $\frac{\alpha_i}{c}\to\infty$ for all $i\in N$, the SIR-feasibility constraints (35) reduce to

$$d(S_{j-1},S_j)+d(S_j,D)-d(S_{j-1},D)\le 0,\quad 2\le j\le n.$$

Since the points are from an underlying metric space, distances satisfy the triangle inequality, which means

$$d(S_{j-1},S_j)+d(S_j,D)-d(S_{j-1},D)\ge 0,\quad 2\le j\le n.$$

Therefore, it must be that

$$d(S_{j-1},S_j)+d(S_j,D)-d(S_{j-1},D)= 0,\quad 2\le j\le n.$$

By summing up the last $n-i$ equations, i.e., $i+1 \le j \le n$, we get

$$\sum_{j=i}^{n-1} d(S_j, S_{j+1}) + d(S_n, D) - d(S_i, D) = 0,$$

from which we obtain

$$\gamma_r = \max_{i \in N} \left(\frac{\sum_{j=i}^{n-1} d(S_j, S_{j+1}) + d(S_n, D)}{d(S_i, D)} \right) = 1.$$

This completes the proof. ∎

EC.3.3. Proof of Theorem 8. First, we note that under the constraint $\frac{\alpha_i}{c} \ge 1$ for all $i \in N$, the SIR-feasibility constraints (35) imply

$$d(S_{j-1}, S_j) + d(S_j, D) - d(S_{j-1}, D) \le \frac{d(S_j, D)}{j}, \quad 2 \le j \le n. \tag{EC.15}$$

We begin by deriving an upper bound on the starvation factor of the i-th passenger, $1 \le i < n$, along any SIR-feasible route. (Note that the starvation factor of the last passenger to be picked up is always 1.) First, we sum up the last $n-i$ inequalities of (EC.15), i.e., $i+1 \le j \le n$, to obtain

$$\sum_{j=i}^{n-1} d(S_j, S_{j+1}) + d(S_n, D) - d(S_i, D) \le \sum_{j=i}^{n} \frac{d(S_j, D)}{j}. \tag{EC.16}$$

Next, we derive upper bounds for each $d(S_j, D)$, $i < j \le n$, in terms of $d(S_i, D)$. The j-th SIR-feasibility constraint from (EC.15) can be rewritten as

$$d(S_j, D) - \frac{d(S_j, D)}{j} \le d(S_{j-1}, D) - d(S_{j-1}, S_j).$$

We know that $d(S_{j-1}, S_j) + d(S_{j-1}, D) \ge d(S_j, D)$, since all points are from an underlying metric space and therefore, distances are symmetric and satisfy the triangle inequality. Using this inequality above, we get

$$\begin{aligned}
& d(S_j, D) - \frac{d(S_j, D)}{j} \le d(S_{j-1}, D) - (d(S_j, D) - d(S_{j-1}, D)) \\
\implies & (2j-1) d(S_j, D) \le 2j d(S_{j-1}, D) \\
\implies & d(S_j, D) \le \frac{2j}{2j-1} d(S_{j-1}, D).
\end{aligned}$$

Unraveling the recursion yields

$$d(S_j, D) \le \left(\prod_{k=i+1}^{j} \frac{2k}{2k-1} \right) d(S_i, D) = \frac{C_j}{C_i} d(S_i, D),$$

where, for $m \ge 1$, $C_m = \prod_{k=1}^{m} \frac{2k}{2k-1}$. We can evaluate C_j as follows:

$$C_j = \prod_{k=1}^{j} \frac{2k}{2k-1} = \prod_{k=1}^{j} \frac{(2k)^2}{2k(2k-1)} = \frac{2^{2j}(j!)^2}{(2j)!} = \frac{2^{2j}}{\binom{2j}{j}}.$$

We then use a known lower bound for the central binomial coefficient, $\binom{2j}{j} \geq \frac{2^{2j-1}}{\sqrt{j}}$, to obtain $C_j \leq 2\sqrt{j}$. This yields $d(S_j,D) \leq \frac{2\sqrt{j}}{C_i} d(S_i,D)$. Substituting in (EC.16), we get

$$\sum_{j=i}^{n-1} d(S_j,S_{j+1}) + d(S_n,D) - d(S_i,D) \leq \sum_{j=i}^{n} \frac{2}{C_i}\frac{\sqrt{j}}{j} d(S_i,D) = \frac{2}{C_i}\left(\sum_{j=i}^{n}\frac{1}{\sqrt{j}}\right) d(S_i,D)$$
$$\implies \sum_{j=i}^{n-1} d(S_j,S_{j+1}) + d(S_n,D) \leq \left(1 + \frac{2}{C_i}\left(\sum_{j=i}^{n}\frac{1}{\sqrt{j}}\right)\right) d(S_i,D).$$

This results in the desired upper bound for the starvation factor of the i-th passenger along any SIR-feasible route:

$$\gamma_r(i) \leq 1 + \frac{2}{C_i}\left(\sum_{j=i}^{n}\frac{1}{\sqrt{j}}\right).$$

The starvation factor of a route is the maximum starvation factor of all its passengers:

$$\gamma_r = \max_{i\in N}\gamma_r(i) \leq \max_{1\leq i<n}\left(1+\frac{2}{C_i}\left(\sum_{j=i}^{n}\frac{1}{\sqrt{j}}\right)\right) = 1 + \frac{2}{C_1}\left(\sum_{j=1}^{n}\frac{1}{\sqrt{j}}\right) = 1 + \sum_{j=1}^{n}\frac{1}{\sqrt{j}},$$

since C_i is increasing in i and $C_1 = 2$. The final step is to show that for all $n \geq 1$, $\sum_{j=1}^{n}\frac{1}{\sqrt{j}} \leq 2\sqrt{n} - 1$. The proof is by induction. The base case (for $n=1$) is satisfied with equality. Assume that the statement is true for some $k \geq 1$. Then, for $k+1$, we have, $\sum_{j=1}^{k+1}\frac{1}{\sqrt{j}} \leq 2\sqrt{k} - 1 + \frac{1}{\sqrt{k+1}} = \frac{\sqrt{4k(k+1)}+1}{\sqrt{k+1}} - 1 \leq \frac{\sqrt{4k(k+1)+1}+1}{\sqrt{k+1}} - 1 = \frac{(2k+1)+1}{\sqrt{k+1}} - 1 = 2\sqrt{k+1} - 1$, which completes the inductive step. Using this bound, we get $\gamma_r \leq 2\sqrt{n}$, as desired. This completes the proof. ∎

EC.3.4. Proof of Theorem 9. First, we note that in the limit, when $\frac{\alpha_i}{c} \to 0$ for all $i \in N$, the SIR-feasibility constraints (35) reduce to

$$d(S_{j-1},S_j) + d(S_j,D) - d(S_{j-1},D) \leq d(S_j,D), \quad 2 \leq j \leq n. \tag{EC.17}$$

Our proof technique is exactly the same as that for Theorem 8. We begin by deriving an upper bound on the starvation factor of the i-th passenger, $1 \leq i < n$, along any SIR-feasible route, by summing up the last $n-i$ inequalities of (EC.17) to obtain

$$\sum_{j=i}^{n-1} d(S_j,S_{j+1}) + d(S_n,D) - d(S_i,D) \leq \sum_{j=i}^{n} d(S_j,D). \tag{EC.18}$$

Next, we derive upper bounds for each $d(S_j,D)$, $i < j \leq n$, in terms of $d(S_i,D)$. The j-th SIR-feasibility constraint from (EC.17) can be rewritten as $d(S_{j-1},S_j) \leq d(S_{j-1},D)$. Using this in the triangle inequality $d(S_j,D) \leq d(S_{j-1},S_j) + d(S_{j-1},D)$, we get $d(S_j,D) \leq 2d(S_{j-1},D)$. Unraveling this recursion then yields $d(S_j,D) \leq 2^{j-i}d(S_i,D)$. Substituting this in (EC.18),

$$\sum_{j=i}^{n-1} d(S_j,S_{j+1}) + d(S_n,D) - d(S_i,D) \leq \sum_{j=i}^{n} 2^{n-i} d(S_i,D) = \sum_{j=0}^{n-i} 2^j d(S_i,D) = \left(2^{n-i+1}-1\right) d(S_i,D)$$
$$\implies \sum_{j=i}^{n-1} d(S_j,S_{j+1}) + d(S_n,D) \leq 2^{n-i+1} d(S_i,D).$$

Thus, the starvation factor of the i-th passenger along any SIR-feasible route is upper bounded as $\gamma_r(i) \leq 2^{n-i+1}$. Finally,

$$\gamma_r = \max_{i\in N}\gamma_r(i) \leq \max_{1\leq i<n} 2^{n-i+1} = 2^n.$$

This completes the proof. ∎

EC.3.5. Proof of Theorem 10. To reduce notational clutter, we let $z_j = \left(1 + \frac{1}{c}\sum_{k=1}^{j-1}\alpha_k\right)^{-1}$, for $1 \le j \le n$. We exhibit an instance of size n for which there is a unique SIR-feasible path whose starvation factor is exactly $\sum_{j=1}^{n} z_j$.

This instance is depicted in Fig. EC.1. Here, $d(S_j, D) = \ell$ for $1 \le j \le n$, and $S_{j-1}S_k > S_{j-1}S_j = z_j\ell$ for $2 \le j < k \le n$. It is straightforward to see that the route $(S_1, S_2, \ldots, S_n, D)$ is SIR-feasible from (35), since for $2 \le j \le n$, we have $d(S_{j-1}, S_j) + d(S_j, D) - d(S_{j-1}, D) = z_j\ell + \ell - \ell = z_j d(S_j, D)$, by construction. Thus, the starvation factor for this route is given by $\sum_{j=2}^{n} z_j + 1 = \sum_{j=1}^{n} z_j$, as desired.

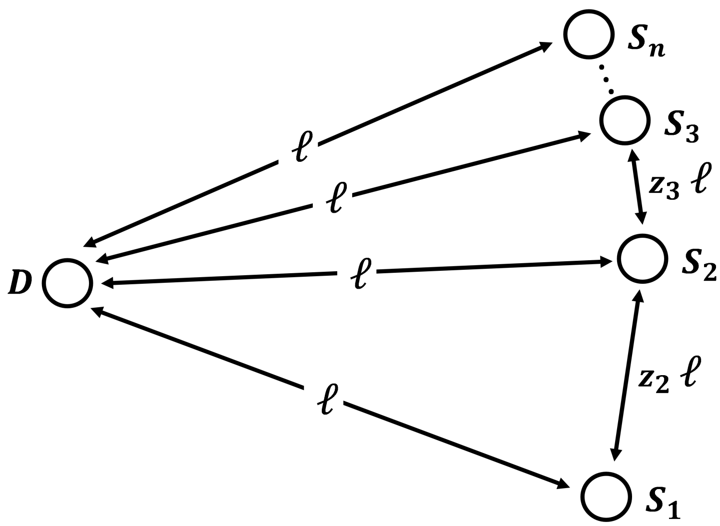

Figure EC.1. An instance to establish lower bound on the starvation factor.

It remains to be shown that no other route is SIR-feasible. First, we note that the SIR-feasibility constraints (35) for this example simplify to

$$d(S_{j-1}, S_j) \le z_j\ell, \quad 2 \le j \le n, \tag{EC.19}$$

where $z_2 > z_3 > \ldots > z_n$, and S_j refers to the j-th pickup point along the route. The proof is by induction. First, consider the pickup point S_1, whose distance from S_2 is $z_2\ell$, and from any other pickup point is strictly greater than $z_2\ell$, by construction. From (EC.19), it can be seen that no two pickup points that are more than $z_2\ell$ apart can be visited in succession, and that the only way to visit two pickup points that are exactly $z_2\ell$ apart is to visit them first and second. Thus, any SIR-feasible route must begin by visiting S_1 and S_2 first. This logic can be extended to build the unique SIR-feasible route that we analyzed above. ∎

EC.4. New algorithmic problems The SIR-feasibility constraints (35) can be considered as additional constraints to the routing optimization problem. For instance, vehicle routing problems with various operational objectives, ridesharing with multiple pickups and dropoff points, online routing problems can all benefit from incorporating SIR-feasibility constraints while performing route optimization. As a concrete example, consider the following ride matching and routing problem:

Given n pickup points and a common dropoff point in a metric space, (a) does there exist an allocation of pickup points to $1 \le m \le n$ vehicles, each with capacity $\lceil \frac{n}{m} \rceil \le c \le n$, such that there exists an SIR-feasible route for each vehicle? And (b) if so, what is the allocation and corresponding routes that minimize the total "vehicle-miles" traveled?

We do not know whether the feasibility problem (a) can be solved in polynomial time, even when $m = 1$ and $\alpha_i = \alpha_j$ for all $1 \le i, j \le n$, where it reduces to finding a sequence of the pickup points that satisfies the inequalities (35). The "Markovian" nature of these inequalities (each inequality only depends on adjacent pickup points in the route) suggests that it may be worth trying to come up with a polynomial time algorithm for the feasibility problem. In Section EC.4.1, we show that this problem is NP-hard when not restricted to a metric space, which implies that any polynomial time algorithm, if one exists, must necessarily exploit the properties of a metric space. However, even if one succeeds in this endeavor, we show in Section EC.4.2 that the optimization (b) over all SIR-feasible routes is NP-hard.

Like SIR-feasibility, there might be other constraints on the ordering of the pickup points (for instance, due to hard requirements on pickup times). Studying such variants might help understand how to tackle SIR-feasibility constraints. For example, it is known that finding the optimal allocation (minimizing the total vehicle-miles traveled) of passengers to vehicles without any restriction on the order of pickups is NP-hard [16]. On the other hand, as we show in Section EC.4.3, the problem is polynomial time solvable if a strict total ordering is imposed and the capacity of each vehicle is unrestricted. It then becomes an interesting future direction to investigate what kinds of order constraints retain polynomial time solvability of the problem.

EC.4.1. Determining existence of SIR-feasible routes is hard. In this section, we present Theorem EC.1, which shows that determining whether an SIR-feasible route exists is NP-hard in general, by a reduction from the undirected Hamiltonian path problem.[3]

Definition EC.1. Given a set $\mathcal{N}$ of n pickup points, and a common dropoff point in an underlying (possibly non-metric) space, and positive coefficients $c, \alpha_1, \alpha_2, \ldots, \alpha_n$, `SIR-Feasibility` is the problem of determining whether an SIR-feasible route of length n exists, that is, whether there exists a sequence of the pickup points that satisfies the SIR-feasibility constraints (35).

Theorem EC.1. *`SIR-Feasibility` is NP-hard.*

Proof. Given an instance of the Hamiltonian path problem in the form of a simple, undirected graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$, we construct an instance of `SIR-Feasibility` as follows. Let P_j denote a pickup point corresponding to vertex $v_j \in V$. Let $N = \{P_1, P_2, \ldots, P_n\}$ denote the set of pickup points, and D denote the common dropoff point. Then, we set the pairwise distances to

$$P_iP_j = \begin{cases} \frac{\ell}{n}, & (v_i, v_j) \in E \\ \ell, & \textit{otherwise}, \end{cases}$$

where $\ell > 0$ is any constant. We also set $P_iD = \ell$ for all i, and $c = \alpha_1 = \alpha_2 = \ldots = \alpha_n$, so that the SIR-feasibility constraints are given by (35). Then, there is a one-to-one correspondence between the set of Hamiltonian paths in G and the set of SIR-feasible routes in the corresponding instance of `SIR-Feasibility`, as follows:

1. Given a Hamiltonian path through a sequence of vertices $(u_1, u_2, \ldots, u_n)$ in G, let the corresponding sequence of pickup points be $(S_1, S_2, \ldots, S_n)$. Then, the route $(S_1, S_2, \ldots, S_n, D)$ is SIR-feasible, since the SIR-feasibility constraints (EC.15) reduce to $d(S_{j-1}, S_j) \le \frac{\ell}{j}$ for $2 \le j \le n$, which are true, by construction.
2. Given an SIR-feasible route $(S_1, S_2, \ldots, S_n, D)$, let the corresponding sequence of vertices in G be $(u_1, u_2, \ldots, u_n)$. Since the route is SIR-feasible, it must be that $d(S_{j-1}, S_j) \le \frac{\ell}{j}$ for $2 \le j \le n$. By construction, this means that $d(S_{j-1}, S_j) = \frac{\ell}{n}$, implying that $(u_{j-1}, u_j) \in E$ for $2 \le j \le n$. Thus, the corresponding path is Hamiltonian.

[3] Given an undirected graph, a Hamiltonian path is a path in the graph that visits each vertex exactly once. The undirected Hamiltonian path problem is to determine, given an undirected graph, whether a Hamiltonian path exists. It is known to be NP-hard.

Hence, any algorithm for `SIR-Feasibility` can be used to solve the undirected Hamiltonian path problem with a polynomial overhead in running time. Since the latter is NP-hard, so is the former. This completes the proof. ∎

However, it can be easily seen that `SIR-Feasibility` is not hard in certain special cases and in certain metric spaces. Consider an input graph, where the pickup points and the dropoff point are embedded on a line, and $\alpha_i = c$ for all $i \in \mathcal{N}$. Without loss of generality, we assume that the pickup points $\{S_1, \ldots, S_n\}$ appear in the same order on the line, so that S_1 and S_n are the two end points. Clearly, if the destination D occurs before S_1 (respectively, after S_n), the instance is SIR-feasible. This is because the route starting from S_n (respectively, S_1) and ending at D, visiting all the pickup points along the way incurs zero detour for everyone, and is thus SIR-feasible. In fact, such a route also traverses the minimum distance among all feasible routes. However, consider the case where D is located at some intermediate location. Such an instance will never be SIR-feasible. To see this, first consider an instance where $n = 2$, and $S_1 < D < S_2$. Let $S_1D = x$, $S_2D = y$; hence $S_1S_2 = x + y$. We analyze the SIR-feasiblity constraints (35) for each of two cases. If S_1 is visited before S_2, then SIR-feasibility requires that $x + y + y - x \le \frac{y}{2}$, which is impossible. Similarly, if S_2 is visited before S_1, then SIR-feasibility requires that $x + y + x - y \le \frac{x}{2}$, which is also impossible. Now, when $n > 2$ and D is located at an intermediate point, any feasible route must, at some point, "jump over" D from some S_i to another S_j, at which stage the analysis would be the same as that for $n = 2$, and is therefore not SIR-feasible. A similar phenomenon can be observed when the underlying metric is a tree rooted at D and the pickup points are located at the leaves, and $\alpha_i = c$ for all $i \in \mathcal{N}$. It can be shown that instances where the pickup points are spread across more than one subtree rooted at D cannot be SIR-feasible, and when the pickup points are all part of a single subtree rooted at D, SIR-feasibility can be checked in polynomial time. We leave open the problem of determining whether `SIR-Feasibility` is hard in general metric spaces.

EC.4.2. Optimizing over SIR-feasible routes is hard. Given an undirected weighted graph, the problem of determining an optimal Hamiltonian cycle[4] (one that minimizes the sum of the weights of its edges) is a well known problem called the *Traveling Salesperson Problem*, abbreviated as *TSP*. A slight variant of this problem, known as *Path-TSP*, is when the traveling salesperson is not necessarily required to return to the starting point or depot, in which case we only seek an optimal Hamiltonian path. These problems are NP-hard [46]. Special cases of the above problems arise when the graph is complete and the edge weights correspond to distances between vertices from a metric space. These variants, which we call *Metric-TSP* and *Metric-Path-TSP*, respectively, are also NP-hard, e.g., [45] showed the hardness for the Euclidean metric.

Definition EC.2. Given a set $\mathcal{N}$ of n pickup points, a common dropoff point in an underlying metric space, and positive coefficients $\alpha_{op}, \alpha_1, \alpha_2, \ldots, \alpha_n$, `Opt-SIR-Route` is the problem of finding an SIR-feasible route of length n of minimum total distance.

Theorem EC.2. *`Opt-SIR-Route` is NP-hard.*

Proof. Given an instance of *Metric-Path-TSP* in the form of a complete undirected graph $G = (V, E)$ and distances $d(v_i, v_j)$ for each $v_i, v_j \in V$ from a metric space, we construct an instance of `Opt-SIR-Route` as follows. Let P_j denote a pickup point corresponding to vertex $v_j \in V$. Let $N = \{P_1, P_2, \ldots, P_n\}$ denote the set of pickup points, and D denote the common dropoff point.

We set the pairwise distances P_iP_j to be equal to $d(v_i, v_j)$ for all $v_i, v_j \in V$. We also set $P_iD = L$ for all i, where

$$L > n \left(\max_{1 \le i < j \le n} P_iP_j \right)$$

[4] A Hamiltonian cycle is a Hamiltonian path that is a cycle. In other words, it is a cycle in the graph that visits each vertex exactly once.

is any constant. We also set $c = \alpha_1 = \alpha_2 = \ldots = \alpha_n$, so that the SIR-feasibility constraints are given by (35). It is easy to see that for any route $(S_1, S_2, \ldots, S_n, D)$, these SIR-feasibility constraints reduce to $d(S_{j-1}, S_j) \leq \frac{\ell}{j}$ for $2 \leq j \leq n$, which are true, by construction and our choice of L. Thus, all $n!$ routes in our constructed instance of `Opt-SIR-Route` are SIR-feasible. Moreover, by construction, the distance traveled along any route is exactly L more than the weight of the path determined by the corresponding sequence of vertices in G. This implies that any optimal SIR-feasible route is given by a sequence of pickup points corresponding to an optimal Hamiltonian path in G, followed by a visit to D. Hence, any algorithm for `Opt-SIR-Route` can be used to solve *Metric-Path-TSP* with a polynomial overhead in running time. Since the latter is NP-hard, so is the former. This completes the proof. ∎

EC.4.3. Optimal allocation of totally ordered passengers to uncapacitated vehicles. In this section, we present a polynomial time algorithm for optimal allocation of passengers to vehicles (minimizing the total vehicle-miles traveled), given a total order on the pickups, and when the capacity of any vehicle is unrestricted. To the best of our knowledge, this result is new; see [52] for a survey on related problem variants.

Our result relies on reducing the allocation problem to a minimum cost flow problem on a flow network with integral capacities. We are given the set $\mathcal{N}$ of passengers (that is, the set of n ordered pickup locations) traveling to a common dropoff location D. Without loss of generality, we let the indices in $\mathcal{N}$ reflect the position in the pickup order, that is, $u \in \mathcal{N}$ is the u-th pick up from location S_u. For convenience, we index the destination D as $n+1$. Let the unknown optimal assignment use $1 \leq m' \leq n$ vehicles (we address how to find it later). A directed acyclic flow network (see Figure EC.2) is then constructed as follows:

(1) s and t denote the source and sink vertices, respectively.
(2) For each passenger/pickup location $u \in \mathcal{N}$, we create two vertices and an edge: an entry vertex u_{in}, an exit vertex u_{out}, and an edge of cost 0 and capacity 1 directed from u_{in} to u_{out}. We also create a vertex $n+1$ corresponding to the dropoff location.
(3) We create n edges, one each of cost 0 and capacity 1 from the source vertex s to each of the entry vertices u_{in}, $u \in \mathcal{N}$.
(4) We create n edges, one each of cost $S_u D$ and capacity 1 from each of the exit vertices u_{out}, $u \in \mathcal{N}$, to the dropoff vertex $n+1$.
(5) To encode the pickup order, for each $1 \leq u < v \leq n$ we create an edge of cost $(S_u S_v - L)$ and capacity 1 directed from u_{out} to v_{in}, where L is a sufficiently large number satisfying $L > 2 \max_{u,v \in \mathcal{N} \cup \{n+1\}} S_u S_v$.
(6) We add a final edge of cost 0 and capacity m' from the dropoff vertex $n+1$ to the sink vertex t, thereby limiting the maximum flow in the network to m' units.

Since all the edge capacities are integral, the integrality theorem guarantees an integral minimum cost maximum flow, and we assume access to a poly-time algorithm to compute it in a network with possibly negative costs on edges. Notice that we do have negative edge costs (step (5) of the above construction); however, our network is a directed acyclic graph, owing to the fact that there is a total ordering on the pickup locations. Hence, there are no negative cost cycles.

Before presenting the full proof, we briefly outline the steps involved:

- Any integral maximum flow from s to t must be comprised of m' vertex-disjoint paths between the source vertex s and the dropoff vertex $n+1$.
- Any integral minimum cost flow must cover all the $2n$ pickup vertices, that is, a unit of flow enters every entry vertex u_{in}, and a unit of flow exits each exit vertex u_{out}, $u \in \mathcal{N}$.
- The partition of $\mathcal{N}$ according to the m' vertex-disjoint paths between s and $n+1$ in an integral minimum cost maximum flow corresponds to the optimal allocation of the n totally ordered passengers among m' uncapacitated vehicles.

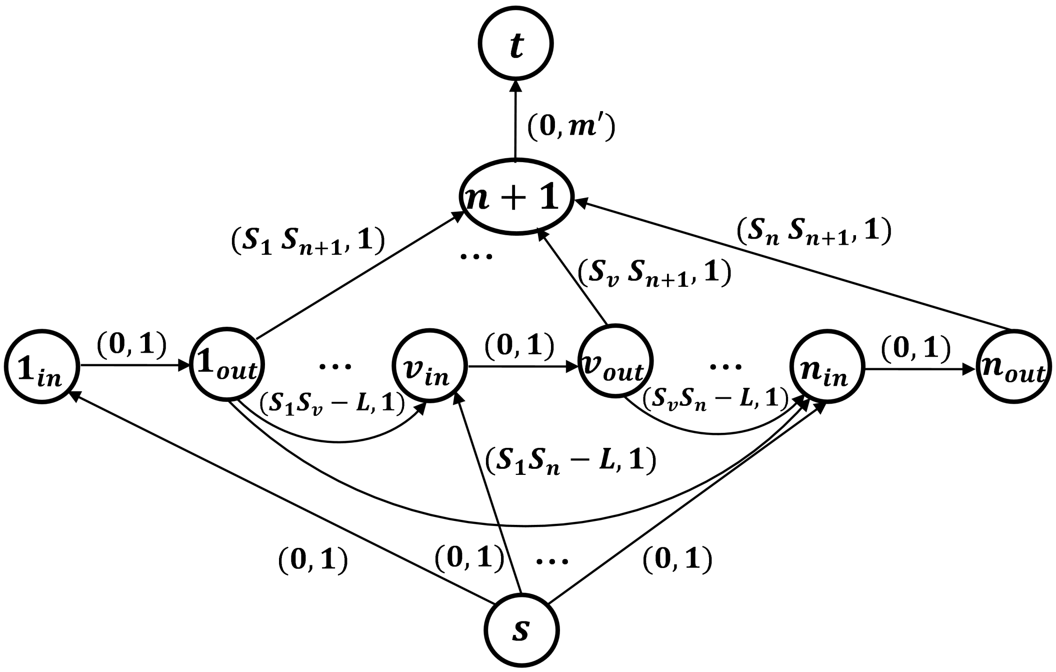

FIGURE EC.2. Illustration of the directed acyclic flow network, a minimum cost maximum flow on which corresponds to an assignment of n totally ordered passengers to m' uncapacitated vehicles. Each of the edge labels correspond to a tuple consisting of edge cost and edge capacity.

Finally, we argue that the overall optimal assignment can be obtained by computing the optimal assignments using the above reduction for each $1 \le m' \le n$ and choosing the one with the overall minimum cost, which completes the reduction. Next, we present the detailed proofs of the above steps.

LEMMA EC.1. *Any integral maximum flow from s to t must be comprised of m' vertex-disjoint paths between the source vertex s and the dropoff vertex $n+1$.*

Proof. First, we observe that any integral *feasible* flow from s to t in the network is comprised of vertex-disjoint paths between the source vertex s and the dropoff vertex $n+1$, each carrying one unit of flow. This is because, every entry vertex u_{in} has only one outgoing edge, namely, the one directed to its corresponding exit vertex u_{out}, which has unit capacity. (Similarly, every exit vertex only has one incoming edge, of unit capacity.) Thus, once a unit of flow is routed through u_{in} and u_{out} by some path, another path cannot route any additional flow through these vertices. Since the maximum flow on the network is m' units, any integral feasible maximum flow would have to have m' such vertex-disjoint paths between s and $n+1$, each carrying one unit of flow. This completes the proof. ∎

LEMMA EC.2. *In any integral minimum cost flow, for every $u \in N$, there is exactly one unit of flow entering u_{in} and exactly one unit of flow leaving u_{out}.*

Proof. From the proof of Lemma EC.1, any integral feasible flow from s to t in the network is comprised of vertex-disjoint paths between the source vertex s and the dropoff vertex $n+1$. Suppose by way of contradiction, an integral minimum cost flow does not route any flow through v_{in} for some $v \in N$. Let $\mathcal{G}_v$ denote the set of passengers $z \in N$ such that $z < v$ and a unit of flow is routed via (z_{in}, z_{out}). Consider two cases:

1. **Case 1:** $\mathcal{G}_v \neq \emptyset$. Let $u = \max \mathcal{G}_v$, and let $\mathcal{P}_u$ be the path that carries a unit of flow from s to $n+1$ through u_{in} and u_{out}. The first vertex in $\mathcal{P}_u$ after u_{out} is either an entry vertex w_{in}

for some $w \in N$ (with $w > v$), or the dropoff vertex $n+1$. Then, we construct a new flow where $\mathcal{P}_u$ is modified to route its unit of flow from u_{out} first to v_{in} to v_{out} and then to w_{in} or $n+1$, as the case may be. (Note that this new flow is feasible, since $u < v < w < n+1$.) If M and M' denote the costs of the original flow and the new flow, then, we show that $M' < M$, contradicting the optimality of M:

- If the original flow took the route $u_{out} \to w_{in}$, and consequently, the new flow takes the route $u_{out} \to v_{in} \to v_{out} \to w_{in}$, then, $M' = M + S_u S_v - L + S_v S_w - L - (S_u S_w - L) < M$ by our choice of L.
- If the original flow took the route $u_{out} \to n+1$, and consequently, the new flow takes the route $u_{out} \to v_{in} \to v_{out} \to n+1$, then, $M' = M + S_u S_v - L + S_v S_{n+1} - S_u S_{n+1} < M$ by our choice of L.

2. **Case 2:** $\mathcal{G}_v = \emptyset$. Let $w \in N$ be such that a unit of flow is routed from s to w_{in}, $\mathcal{P}_w$ denoting the corresponding path. There may be more than one choice for w_{in} as defined, but all of them satisfy $v < w$, since $\mathcal{G}_v = \emptyset$, so it does not matter which one is picked. As before, we construct a new flow where $\mathcal{P}_w$ is modified to route its unit of flow from s first to v_{in} to v_{out} and then to w_{in}. (Note that this new flow is feasible, since $v < w$.) If M and M' denote the costs of the original flow and the new flow, $M' = M + S_v S_w - L < M$ by our choice of L, contradicting the optimality of M.

This completes the proof. ∎

Lemma EC.3. *The partition of N according to the m' vertex-disjoint paths between s and $n+1$ in an integral minimum cost maximum flow corresponds to the optimal allocation of the n totally ordered passengers among m' uncapacitated vehicles.*

Proof. From Lemma EC.1 and Lemma EC.2, we know that any integral minimum cost maximum flow $\mathcal{F}$ is comprised of m' vertex-disjoint paths between s and $n+1$ that cover all n pickup points between them, by routing a unit of flow along (u_{in}, u_{out}) for all $u \in N$. We adopt a simplified representation of a path by removing the edges from the source vertex s, as well as the edges between u_{in} and u_{out}, the entry and exit vertices corresponding to pickup points $u \in N$. For example, a path $s \to u_{in} \to u_{out} \to v_{in} \to v_{out} \to n+1$ would be contracted to $u \to v \to n+1$. Note that this does not affect the cost computation, since only zero cost edges are removed. For any $u, v \in N$, the cost of any edge (u, v) in the new representation is simply the cost of the edge (u_{out}, v_{in}) in the old representation. Similarly, for any $u \in N$, the cost of any edge $(u, n+1)$ in the new representation is simply the cost of the edge $(u_{out}, n+1)$ in the old representation. Let the set of these m' paths be denoted as $\mathcal{P}_{\mathcal{F}}$. Thus, we have established a one-to-one correspondence between (a) the set of all integral flows $\mathcal{F}$ comprised of m' vertex-disjoint paths $\mathcal{P}_{\mathcal{F}}$ that collectively cover all n pickup locations, and (b) the set of all allocations of n totally ordered passengers (traveling to a common dropoff location $n+1$) to m' uncapacitated vehicles.

For any path $P \in \mathcal{P}_{\mathcal{F}}$, let $|P|$ denote the length of the path, that is, the number of edges in the path. The cost of path P is then given by

$$c(P) = \sum_{\substack{1 \le u < v \le n \\ (u,v) \in P}} \sum (S_u S_v - L) + \sum_{\substack{1 \le u \le n \\ (u,n+1) \in P}} S_u S_{n+1}.$$

Since all paths end with vertex $n+1$, there are $|P| - 1$ terms in the first sum and 1 term in the last sum. Thus, $c(P)$ can be equivalently written as

$$c(P) = \sum_{\substack{1 \le u < v \le n+1 \\ (u,v) \in P}} \sum S_u S_v - (|P| - 1)L.$$

The cost of flow $\mathcal{F}$ is simply the sum of the costs of the paths in $\mathcal{P}_{\mathcal{F}}$, given by

$$c(\mathcal{F}) = \sum_{P \in \mathcal{P}_{\mathcal{F}}} c(P) = \sum\sum_{\substack{1 \le u < v \le n+1 \\ (u,v) \in \bigcup \mathcal{P}_{\mathcal{F}}}} S_u S_v - \sum_{P \in \mathcal{P}_{\mathcal{F}}} (|P| - 1)L.$$

Since $|P|$, the length of path P, also denotes the number of pickup points covered by P, and all the m' paths are vertex-disjoint (except for $n+1$), the summation in the second term is simply $n - m'$, independent of the flow $\mathcal{F}$. Thus,

$$c(\mathcal{F}) = \sum\sum_{\substack{1 \le u < v \le n+1 \\ (u,v) \in \bigcup \mathcal{P}_{\mathcal{F}}}} S_u S_v - (n - m')L = c(\mathcal{A}_{\mathcal{F}}) - (n - m')L, \tag{EC.20}$$

where $c(\mathcal{A}_{\mathcal{F}})$ denotes the cost (total vehicle-miles traveled) of the corresponding allocation of n totally ordered passengers (traveling to a common dropoff location $n+1$) to m' uncapacitated vehicles. From (EC.20), it is clear that the set of integral minimum cost maximum flows $\arg\min_{\mathcal{F}} c(\mathcal{F})$ also corresponds to the set of optimal allocations of n totally ordered passengers among m' uncapacitated vehicles. This completes the proof. ∎

Theorem EC.3. *There exists a polynomial time algorithm to find an optimal allocation of totally ordered passengers to uncapacitated vehicles.*

Proof. Using the one-to-one correspondence established in Lemma EC.3, for each "guess" $1 \le m' \le n$, we find the corresponding optimal allocation by solving a minimum cost maximum flow problem in poly-time, finally choosing a guess with the overall least cost allocation. ∎