Informal and Privatized Transit: Incentives, Efficiency and Coordination

Devansh Jalota
Columbia
dj2757@columbia.edu

Matthew Tsao
Lyft
mattxtsao@gmail.com

March 7, 2026

Abstract

Informal and privatized transit services, such as minibuses and shared auto-rickshaws, are integral to daily travel in large urban metropolises, providing affordable commutes where a formal public transport system is inadequate and other options are unaffordable. Despite the crucial role that these services play in meeting mobility needs, governments often do not account for these services or their underlying incentives when planning transit systems, which can significantly compromise system efficiency.

Against this backdrop, we develop a framework to analyze the incentives underlying informal and privatized transit systems, while proposing mechanisms to guide public transit operation and incentive design when a substantial share of mobility is provided by such profit-driven private operators. We introduce a novel, analytically tractable game-theoretic model of a fully privatized informal transit system with a fixed menu of routes, in which profit-maximizing informal operators (*drivers*) decide where to provide service and cost-minimizing commuters (*riders*) decide whether to use these services. Within this framework, we establish *tight* price of anarchy bounds which demonstrate that decentralized, profit-maximizing driver behavior can lead to bounded yet substantial losses in cumulative driver profit and rider demand served. We further show that these performance losses can be mitigated through targeted interventions, including Stackelberg routing mechanisms in which a modest share of drivers are centrally controlled, reflecting environments where informal operators coexist with public transit, and cross-subsidization schemes that use route-specific tolls or subsidies to incentivize drivers to operate on particular routes. Finally, we reinforce these findings through numerical experiments based on a real-world informal transit system in Nalasopara, India.

1 Introduction

Governments in many cities, particularly in developing nations, often struggle to provide adequate public transport to meet the rising travel demands due to chronic fiscal constraints. Although on-demand mobility services, such as Uber and Lyft, have partially filled this gap, they remain unaffordable for most low and middle-income users in these cities. Thus, commuters face a stark mobility gap: on one side, low-frequency, overcrowded, and geographically limited public transit; on the other, high-cost individualized services that are unaffordable for daily travel. To fill this gap, informal transport (or intermediate public transport) services, such as minibuses, shared auto-rickshaws, matatus, jeepneys, and dollar vans, have become integral to daily travel, providing

affordable commutes where a formal public transport system is absent or inadequate and other options are unaffordable (Cervero 2000, Cervero and Golub 2007). Operated largely by private providers, these services form the backbone of urban mobility in many developing nation cities, often accounting for 50–100% of all trips (Conwell 2025). For instance, in Lagos, Nigeria, minibuses alone serve roughly 62% of daily commutes (Björkegren et al. 2025).

Informal or otherwise privatized fixed-route services are not limited to developing nations, with similar services filling gaps even in mature public transit systems in the developed world. In New York, for example, a range of private operators have emerged, from long-standing jitney services in outer-borough corridors (Goldwyn 2020) to Uber's recent airport shuttle offerings (Uber 2025), which pool riders along fixed routes to offer a lower cost shared alternative to individualized ride-hailing. Together, these examples highlight a broader global pattern: privatized and, often, informal shared transit systems are crucial to meeting the mobility needs of hundreds of millions of daily travelers worldwide that formal public transit leaves unmet.

Despite the critical role that informal and privatized shared transit plays in meeting mobility needs, governments often do not account for these services or their underlying incentives when planning transit systems, which can significantly compromise system efficiency and equity (Klopp and Cavoli 2019). Accounting for these services in mobility planning is essential because, although they resemble public transit operationally, often running shared vehicles along fixed routes, they differ fundamentally in their incentive structures and objectives. Public transit agencies operate under a service mandate aimed at broad geographic coverage while informal operators are private entrepreneurs whose primary goal is to maximize profits. This divergence in incentives fundamentally shapes service patterns: while informal and privatized operators have the potential to complement public transit by extending coverage at low (or no) fiscal cost, they also tend to overserve lucrative corridors and underserve low-income or peripheral neighborhoods with lower returns. These incentive-driven patterns raise a central policy question: *how should the incentives of informal and privatized transit operators be shaped and public transit be designed, i.e., determining where and with what frequency to offer public transit, when a large share of mobility is provided by such profit-driven private actors?* Understanding this interaction is essential for designing public transit systems that complement, rather than conflict with, the informal and privatized systems on which millions of riders rely.

1.1 Our Contributions

This work addresses the above question by developing a framework for analyzing the incentives underlying informal and privatized transit systems and their interaction with public transit. Moreover, we propose mechanisms to guide public transit operation and incentive design when a substantial share of mobility is provided by profit-driven private operators.

Towards this goal, in Section 3, we introduce a novel game-theoretic, two-sided market model of a fully privatized informal transit system (i.e., without public transit), with a fixed set of routes. On these routes, cost-minimizing commuters (*riders*) choose between using informal transit to complete their trip or an outside option, such as walking or staying at home, while profit-maximizing informal or privatized operators (*drivers*) decide which routes to serve. Inspired by the practical operation and global prevalence of fixed-route informal and privatized transit systems, our framework captures two key features that shape incentives and equilibrium outcomes in such settings: (i) competition among drivers for profitable routes, which leads to over-provision on some routes

and under-provision in others, and (ii) rider queuing and waiting delays, a defining feature of informal transit during morning and evening peak periods that shapes riders' willingness to use these services.

To capture driver competition and its effect on their route choice decisions, we use a non-atomic congestion game in which congestion reflects the impact of service over or under-provision on rider and driver payoffs, rather than physical road congestion. To capture rider queuing and waiting delays, we adapt Vickrey's seminal bottleneck model of peak-period congestion (Vickrey 1969) to our informal transit setting. A key departure from the standard bottleneck model is that a route's service capacity is not fixed exogenously, but emerges endogenously from drivers' profit-maximizing route choices. In this sense, the rider-side problem of our framework can be viewed as a *bottleneck model with endogenous service capacities*. By jointly capturing both driver competition and rider delays in a unified framework, our model is sufficiently rich to capture the key operational features of informal and privatized transit systems, yet tractable enough to admit closed-form solutions and isolate the core forces at play. We thus view the *formulation* of our model as a central contribution.

For the above model, in Section 4, we derive price of anarchy (PoA) guarantees under both cumulative driver profit and rider welfare objectives, where rider welfare is measured by the total number of riders served by the informal transit system. Our PoA bounds quantify the worst-case ratio (over problem instances) between the optimal and equilibrium outcomes, thereby capturing the performance loss in terms of cumulative driver profits or rider welfare due to drivers' selfish route choices. In particular, we establish *tight* PoA bounds of 2 for cumulative driver profit and $1 + \frac{p_{\max}}{p_{\min}}$ for rider welfare, where $p_{\min}$ and $p_{\max}$ are the minimum and maximum per-rider profits across routes. To our knowledge, these constitute the first PoA bounds for informal or privatized transit systems. Central to establishing these results is a closed-form characterization of the equilibrium relationship between driver supply and rider demand served by informal transit on each route, whose monotonicity and concavity properties underpin the PoA guarantees. Overall, these guarantees demonstrate that decentralized, self-interested driver decisions can result in *bounded yet substantial inefficiencies* in both cumulative driver profit and rider welfare.

To address the performance losses arising from profit-maximizing driver behavior, we study two mechanisms that a public authority can use to steer informal and privatized transit operators to improved system outcomes: *cross-subsidization* (Section 5) and *Stackelberg routing* (Section 6).

First, we study cross-subsidization, in which a public authority uses route-specific tolls or subsidies to encourage drivers to operate on particular routes, thus shaping the overall service pattern. We show that for any instance of our informal transit system, there exists a budget-balanced (i.e., requiring no net spending) cross-subsidization scheme that aligns individual driver incentives with any desired system objective, and can be computed in polynomial time in the problem parameters for both cumulative driver profit or rider welfare objectives. This result demonstrates that, by appropriately accounting for the incentives of informal or privatized operators, public authorities can eliminate the PoA inefficiencies arising from selfish driver behavior.

Despite its efficacy, cross-subsidization can be difficult to implement in practice, particularly in informal transit systems, where monitoring and enforcement are challenging, and compliance is limited. We thus study Stackelberg routing mechanisms, where a public authority centrally routes a fraction $\alpha \in [0, 1]$ of drivers, while the remaining $1 - \alpha$ fraction consists of informal or privatized drivers who choose profit-maximizing routes in response. Such a mechanism is especially relevant in cities where both public transit and informal or privatized transit operators coexist. Although computing an optimal Stackelberg strategy under both cumulative driver profit and rider welfare

objectives is, in general, NP-hard, we propose two algorithms, Lowest Profit First (LPF) and Linearized Non-Compliant First (L-NCF), with provably strong guarantees. We show that LPF is Stackelberg-optimal for both objectives for $n = 2$ routes, and that L-NCF, which employs a novel linearization technique, achieves near-optimal performance across both metrics for an arbitrary number of routes under practically relevant conditions motivated by our empirical setting.

We complement our theory with numerical experiments in Section 7 based on a real-world informal transit system in Nalasopara, India, which serves nearly 100,000 riders daily. Our results show that, although real-world inefficiencies do not reach the worst-case levels given by our PoA bounds, the corresponding fractions of the optimal cumulative driver profit and rider welfare obtained under operational data from this system remain meaningfully bounded away from 1, implying substantial inefficiencies in practice. We further evaluate our mechanisms in a Stackelberg routing setting, in which case L-NCF and LPF consistently outperform a status-quo *Greedy* baseline, which optimizes the public authority's objective, either cumulative driver profit or rider welfare, without accounting for the strategic behavior of privatized operators. Under the Greedy baseline, improvements in cumulative driver profit (rider welfare) do not materialize until the authority controls 70–80% (40–50%) of drivers. In contrast, our algorithms achieve significant gains in both objectives with as little as 20% centralized control.

Taken together, our work highlights the importance of integrating informal and privatized operators into transit planning and provides practical, actionable mechanisms for aligning their incentives with system objectives. While framed in the context of informal and privatized transit, our model may be broadly applicable to other settings, such as ride-sharing, in which the forces we study, including driver competition for service and rider queuing delays, play a central role.

2 Related Literature

A large and growing literature documents the key role of informal transit in urban mobility and emphasizes the importance of integrating these services with formal public transit (Gadapalli 2016, Kumar et al. 2016). These works highlight defining aspects of informal transit, including institutional arrangements, regulatory challenges, and service patterns. Despite taking many forms across contexts, "informal transport is, stripped to the basics, about profit-seeking operators serving consumer demands" (Cervero 2000). Motivated by this core feature, we study incentive interactions in such systems, focusing on fixed-route informal transit, common in cities worldwide. However, unlike the largely qualitative and case study based informal transit literature rooted in urban planning and political science (Klopp and Cavoli 2019, Klopp 2021), our work develops a formal framework that complements these diagnostic studies by providing prescriptive tools for incentive design and public transit interventions in such systems.

By modeling the incentives of informal and privatized transit systems and the effects of selfish driver behavior on system performance, our work connects to the literature on the efficiency of decentralized and privatized transportation provision. Prior work studies competition among private service providers (Rosaia 2025) and emphasizes search frictions (Brancaccio et al. 2023, Buchholz 2015), where spatial mismatch leads to matching costs or delays. In the context of informal transit, Mbonu and Eaglin (2024) document how territorial segmentation among minibus associations affects service provision, while Mittal et al. (2024) show that informal transit routes often deviate less from shortest paths than formal public transit. Closest to our work, Conwell (2025) develops a model of privatized shared transit capturing profit-maximizing drivers and rider queuing delays,

and studies the efficiency gains from reorganization of privately operated minibuses. However, unlike the discrete stochastic queuing model of Conwell (2025), we adopt a deterministic continuum queuing framework inspired by Vickrey's bottleneck model (Vickrey 1969) and leverage ideas from congestion games, enabling a closed-form equilibrium analysis.

Our work also contributes to the literature on quantifying efficiency losses from decentralized, selfish agent behavior via price of anarchy (PoA) bounds (Paccagnan et al. 2020, Lin et al. 2017, Roughgarden 2005). We extend this line of work, typically focused on a single class of strategic agents (e.g., commuters in routing games), to informal transit systems, where service provision and rider demand are jointly shaped by the strategic actions of both drivers and riders. Related work on informal transit systems considers two-sided incentives but rely primarily on numerical comparisons rather than PoA bounds (Chavis and Daganzo 2013, Sangveraphunsiri et al. 2022), while others focus on equilibrium computation without efficiency comparisons (Fernandez and Marcotte 1992). In contrast, we derive explicit PoA bounds, which, to our knowledge, are the first such results for informal transit systems.

A key component of our framework is Vickrey's bottleneck model (Vickrey 1969), which we adapt to study the rider-side decision of our framework, where we treat waiting time for informal transit akin to queuing delay at a bottleneck (Kraus and Yoshida 2002). Moreover, in line with prior extensions of the bottleneck model, we incorporate outside options (Arnott et al. 1993, Gonzales and Daganzo 2012), allowing riders to forgo transit when queuing delays become large. Despite these similarities, unlike existing bottleneck-based models that focus exclusively on rider behavior under fixed service capacities, we embed bottleneck-style rider behavior within a two-sided informal transit system in which service capacities and outcomes are endogenously determined by the strategic interaction between riders and profit-maximizing drivers.

By studying the allocation of drivers to serve riders in mobility systems, our work relates to the ridesharing and shared rides literature. Queuing theory has been used to model and design prices for ridehailing systems (Banerjee et al. 2015, 2016), typically in stochastic and stationary settings, unlike Vickrey's deterministic and non-stationary bottleneck model. Shared rides services have been studied from the lens of online stochastic optimization (Aouad and Sarıtaç 2022, Ashlagi et al. 2019) and dynamic programming (Yan et al. 2024, Dogan and Jacquillat 0). These works treat rider arrival as exogenous, and do not model the option for riders to strategically time their entry into the market, which we incorporate in our model. These works study exclusively matching pairs of riders together, with the exception of Alonso-Mora et al. (2017), which considers higher-capacity shared rides but abstracts from strategic rider and driver behavior and pricing. Overall, these works study discrete models, as opposed to the fluid model studied in our work, though fluid approximations are often useful even in discrete models for designing algorithms with provable guarantees (Aouad and Sarıtaç 2022, Feng et al. 2024).

Beyond our PoA and modeling contributions, our analysis of settings in which informal or privatized transit coexists with public transit relates to a broader literature on interactions between public transit and other modes, including ride-hailing (Wei et al. 2022, Erhardt et al. 2022, Yang and Geroliminis 2025) and informal or privatized transit (Björkegren et al. 2025) services. While much of this literature is empirical or simulation-based (Neumann et al. 2015, Björkegren et al. 2025, Zwick 2017, Özbilen 2016, Salomon and Silman 1985), a smaller set of modeling-based approaches study competition between public and informal transit, typically using stylized corridor models (Yang et al. 2001, Gronau 2000) or network formulations (Viton 1982). Our work differs by adopting a leader–follower framework in which decentralized informal operators choose profit-maximizing

routes in response to a public authority's routing decisions. In this sense, our work is most closely related to the Stackelberg routing literature (Roughgarden 2001, Krichene et al. 2014), which we discuss in Section 6.

Our work is also related to cross-subsidy mechanisms, the design of public transit systems, and the economic impacts of public transportation infrastructure, which we review in Appendix A.1

3 Model

This section introduces a stylized two-sided market model of informal and privatized transit with a fixed menu of routes. On these routes, *riders* seek to make trips, and service is provided by profit-seeking informal or privatized operators, whom we refer to as *minibus drivers* (or simply *drivers*). Motivated by the practical operations of these systems, our framework captures two key features shaping incentives and equilibrium outcomes: (i) driver route choice and competition for profitable routes, which endogenously determine service capacities across routes, and (ii) rider queuing and waiting delays, which influence their willingness to use informal transit. While real-world informal and privatized transit systems are highly complex, our model abstracts from some operational details to isolate these two core forces (Cervero 2000), while remaining faithful to how these systems operate in practice. These abstractions enable a tractable framework that yields clear insights into the incentives and equilibrium behavior in informal transit systems, and extending the model to incorporate additional operational features is a valuable future research direction. For a discussion of our modeling assumptions beyond that presented in this section, see Appendix A.2.

In the following, we first introduce notation and describe the features of the informal transit environment we study (Section 3.1) and specify rider demand for informal transit as a function of driver supply on each route (Section 3.2). Then, we define driver payoffs and the equilibrium notion we study in Section 3.3. Finally, we introduce the performance metrics of cumulative driver profit and rider welfare and present the price of anarchy notion in Section 3.4.

3.1 Preliminaries: Notation and Features of Informal Transit Environment

We consider an informal transit system during the morning or evening peak period, defined by the interval $[t_1, t_2]$, with a fixed menu of n routes. In this system, a continuum of D *minibus drivers* choose which routes to serve, where each minibus has a fixed capacity F, denoting the mass of riders that can be served in a single trip. Consistent with observed practice during peak periods, a driver serving a route operates at full capacity, picking up riders at the origin, transporting them to their destination, and then returning empty to the origin to continue service (see left of Figure 1).

Each route $i \in [n]$ is characterized by a vector $\Theta_i = (\bar{p}_i, l_i, c_i)$, where $\bar{p}_i$ denotes the per-trip fare paid by riders availing the minibus, l_i is the fixed one-way minibus travel time on a route (so the round-trip travel time is $2l_i$), and c_i is the per-trip operating cost incurred by the driver. As in Vickrey's seminal bottleneck model (Vickrey 1969), we let Λ_i denote the total mass of riders seeking to travel on route i with desired arrival times at the destination uniformly distributed over the peak-period interval $[t_1, t_2]$, and let $\lambda_i = \frac{\Lambda_i}{t_2 - t_1}$ denote the deterministic constant desired arrival rate on that route. In the system, riders on each route choose between traveling by minibus (M) and an outside option (O), such as walking, based on their travel costs defined in Section 3.2 (see right of Figure 1).

A central primitive of our model is the mapping from driver supply on a route to the rider demand served by minibuses, as only a fraction of riders may be served when service capacity is

limited. Let $\mathbf{x} = (x_i)_{i\in[n]}$ denote the driver allocation across routes, where $x_i \geq 0$ is the mass of drivers serving route i, and let $\Omega = \{\mathbf{x} \in \mathbb{R}^n_{\geq 0} : \sum_{i\in[n]} x_i = D\}$ denote the feasible set of driver allocations. All results extend naturally to the case when $\sum_{i\in[n]} x_i \leq D$, and we focus on the equality-constrained case for analytical clarity. Given a driver allocation $\mathbf{x}$, we define $\mathbf{\Lambda}^M(\mathbf{x}) = (\Lambda_i^M(x_i))_{i\in[n]}$ as the vector of rider demands served by the minibuses, where $\Lambda_i^M(x_i)$ is the number of riders served when x_i drivers operate on that route. The remaining $\Lambda_i - \Lambda_i^M(x_i)$ riders take the outside option.

We let $\mathcal{L}$ denote the class of admissible rider demand functions $\mathbf{\Lambda}^M(\cdot)$, induced under particular modeling choices for rider costs and queuing behavior, which we specify in Section 3.2. Then, we have the following description of an instance of an informal transit system.

Definition 1 (The Informal Transit System)**.** An instance of an informal transit system is specified by the tuple $I = (n, F, \overline{\mathbf{p}}, \mathbf{l}, \mathbf{c}, t_1, t_2, D, \mathbf{\Lambda}^M)$, where n is the number of routes, F is the vehicle capacity, and $\overline{\mathbf{p}}, \mathbf{l}, \mathbf{c} \in \mathbb{R}^n$ denote the vectors of per-trip fares, travel times, and driver costs for each of the n routes. $[t_1, t_2]$ is the peak-period interval over which riders' desired destination arrival times are uniformly distributed, D is the total mass of minibus drivers, and $\mathbf{\Lambda}^M : \mathbb{R}^n \to \mathbb{R}^n$ specifies the rider demand for each route as a function of the driver supply on that route. Then, under a rider demand function family $\mathcal{L}$, the set of all instances of the informal transit system is given by:

$$\mathcal{I}_{\mathcal{L}} := \{(n, F, \overline{\mathbf{p}}, \mathbf{l}, \mathbf{c}, t_1, t_2, D, \mathbf{\Lambda}^M) : n \geq 1, \overline{\mathbf{p}}, \mathbf{l}, \mathbf{c} \in \mathbb{R}^n_{>0}, F, t_1, t_2, D \in \mathbb{R}_{>0}, \mathbf{\Lambda}^M \in \mathcal{L}\}.$$

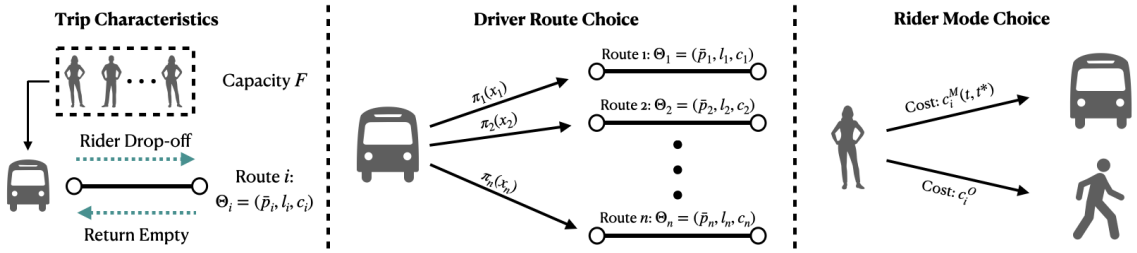


Figure 1: Depiction of the characteristics of the informal transit system on which minibus drivers service riders on a fixed menu of routes. The left panel depicts the trip characteristics, where, on each trip, a minibus transports F riders from the route's origin to its destination and then returns empty to the origin to serve additional riders. The middle panel illustrates driver route choice, with each driver selecting one of n routes to operate on based on profitability. The right panel shows the mode choice decision faced by riders who either travel by minibus or take the outside option, such as walking.

3.2 Minibus Rider Demand Function

This section specifies the class $\mathcal{L}$ of admissible rider demand functions $\mathbf{\Lambda}^M(\cdot)$ that we focus on in this work. Specifically, we consider rider demand functions arising endogenously from equilibrium rider behavior, for each level of driver supply, under an adaptation of Vickrey's bottleneck model of peak-period congestion (Vickrey 1969, Gonzales and Daganzo 2012), in which rider costs reflect queuing delays, a central feature of informal transit systems during morning or evening peak periods. While we focus on this class of rider demand functions, our formulation is sufficiently general as it subsumes an important special case of capacity-constrained demand, which is closely related to formulations commonly studied in adjacent domains such as ride-sharing, where rider queuing delays are abstracted away (Bimpikis et al. 2019), and central to several of our theoretical and algorithmic results (see Propositions 2, 3, and Section 6). That said, extending our analysis to

alternate demand function classes arising from different modeling choices of rider costs and queuing behavior is an important direction for future work.

Demand Function Under Rider Queuing: To specify the rider demand function class we study, we first describe rider behavior under an adaptation of Vickrey's bottleneck model of peak period congestion (Vickrey 1969, Gonzales and Daganzo 2012). Specifically, we consider cost-minimizing riders on each route $i \in [n]$ who choose either an (i) outside option (e.g., walking) with a fixed cost c_i^O, or a (ii) minibus service. Given x_i drivers operating on route i, the cost of using the minibus depends on its service capacity $\mu_i(x_i) = \frac{x_i F}{2l_i}$, where F is the vehicle capacity and $2l_i$ is the round-trip travel time. In the *over-supplied* regime, when the service capacity exceeds the desired rider arrival rate, i.e., $\mu_i(x_i) \geq \lambda_i$, all riders can arrive at the destination at their desired arrival times when using the minibus without queuing or schedule delays. In this case, the cost of using the minibus is a constant given by $c_i^M = \bar{p}_i + \eta_T l_i$, where η_T represents the common value of time across riders. Without loss of generality, we assume that the minibus absent rider queuing and schedule delays is weakly more attractive than the outside option, i.e., $S_i := c_i^O - c_i^M \geq 0$ for all routes i; otherwise, such routes can be removed from consideration, as riders would always prefer the outside option regardless of minibus supply.

In the *under-supplied* regime when $\mu_i(x_i) < \lambda_i$, rider queuing and schedule delays may arise, as is common during morning or evening peak periods in informal transit systems. In this case, riders, beyond choosing their mode (i.e., outside option or minibus), select their arrival time when using the minibus. Accordingly, in addition to the fixed cost c_i^M, the cost of using the minibus consists of (i) queuing (or waiting) delays $w(t)$, representing the time spent waiting for the minibus by a rider whose trip is completed at time t, and (ii) schedule delays $|t^* - t|$, capturing the discrepancy between a rider's desired destination arrival time t^* and actual arrival time t. Let η_W be the penalty for incurring a unit of waiting time delay, and η_E and η_L be the schedule delay penalties for arriving early or late, respectively, with $0 < \eta_E < \eta_W$ and $\eta_L > 0$, as is standard in the literature. Then, normalizing the waiting time penalty $\eta_W = 1$ (with all other parameters scaled accordingly), the minibus travel cost for a user with a desired destination arrival time t^* and actual arrival time t is:

$$c_i^M(t, t^*) = \bar{p}_i + \eta_T l_i + w(t) + \eta_E(t^* - t)_+ + \eta_L(t - t^*)_+. \tag{1}$$

Under the above rider cost structure, for a given driver supply x_i on a route i, a *rider equilibrium* arises when all Λ_i riders minimize travel costs by choosing a mode (minibus or outside option) and, if traveling by minibus, an arrival time, such that no rider has an incentive to deviate. For a fixed driver supply x_i, the resulting equilibrium is akin to that in Vickrey's bottleneck model with an outside option (Gonzales and Daganzo 2012), and as x_i varies, traces out the minibus rider demand function $\Lambda_i^M(\cdot)$ depicted on the left of Figure 2. We provide a closed-form characterization of this demand function in Equation (3) in Proposition 1 (see Section 4.1), along with a detailed discussion of the regimes that arise. Given this characterization, we define the class $\mathcal{L}_V$ as the set of all rider demand functions in Equation (3) for arbitrary choices of strictly positive model primitives $F, \overline{\mathbf{p}}, \mathbf{l}, \mathbf{c}, t_2, t_1, \eta_E, \eta_L, (\Lambda_i)_{i\in[n]}$ and non-negative values of $(S_i)_{i\in[n]}$, where, recall that $S_i = c_i^O - c_i^M$ captures the relative attractiveness of the minibus absent queuing and schedule delays. Note that if $S_i < 0$, the rider demand is zero.

Capacity-Constrained Demand Function: We now consider an important special case of the above rider demand function, corresponding to $S_i = 0$ for a route i. In this case, no queuing or schedule delays arise at equilibrium, as any such delays would strictly increase the cost of the minibus and induce riders to switch to the outside option. Hence, the minibus rider demand function

on a route reduces to Equation (4) (see Section 4.2), exhibiting a piecewise-linear relationship with driver supply as shown on the right of Figure 2. Specifically, the served rider demand on route i increases linearly in the mass of allocated drivers up to a threshold $k_i^* = \frac{2l_i\lambda_i}{F}$ at which $\mu_i(x_i) = \lambda_i$, beyond which the minibus service capacity is sufficient to serve all arriving riders.

The capacity-constrained demand function mirrors standard equilibrium relations where the served demand is limited by the minimum of the demand arrival rate and service capacity, with full demand capture occurring only once the saturation condition $\mu_i(x_i) = \lambda_i$ holds. While both rider demand functions in Figure 2 coincide in the over-supplied regime (i.e., $\mu_i(x_i) \geq \lambda_i$) when the driver supply exceeds k_i^*, beyond which the minibus serves all rider demand, they diverge in the under-supplied regime. In this regime, unlike the $S_i = 0$ case corresponding to the capacity-constrained demand function, when $S_i > 0$, riders are willing to incur queuing or schedule delays for the minibus service, allowing the minibus to capture additional demand even when $\mu_i(x_i) < \lambda_i$. In this case, we show that the rider demand function is concave-quadratic below a threshold $\tilde{k}_i^*$ and this threshold need not coincide with the saturation threshold at which $\mu_i(x_i) = \lambda_i$, i.e., $k_i^* \neq \tilde{k}_i^*$; in particular, there exist regimes in which $\mu_i(x_i) < \lambda_i$, yet all arriving riders are served by the minibus at equilibrium. For a further discussion on these rider demand functions, see Section 4.1.

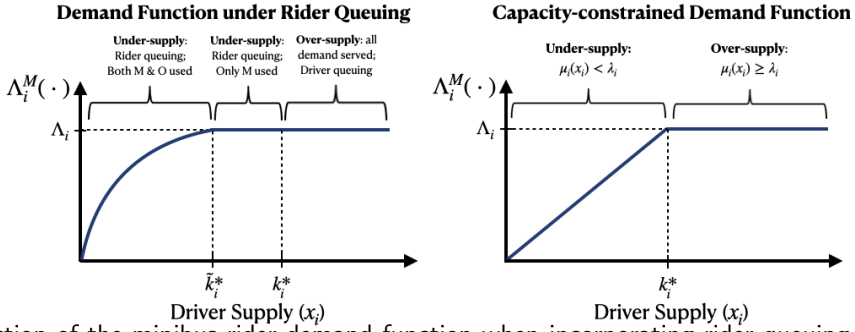


Figure 2: Depiction of the minibus rider demand function when incorporating rider queuing (left) and under a capacity-constrained formulation without rider queuing (right). When rider queuing is incorporated via a Vickrey-style bottleneck model, the minibus rider demand is concave quadratic below a threshold $\tilde{k}_i^* \leq k_i^*$, and remains flat thereafter. In the special case when the cost difference $S_i = 0$, this rider demand function reduces to the capacity-constrained formulation, where the served minibus rider demand increases linearly with driver supply until the capacity-matching threshold k_i^* at which $\mu_i(x_i) = \lambda_i$, and remains flat thereafter.

3.3 Driver Payoffs and Equilibrium Notion

We now describe the supply-side of the model and specify how drivers allocate themselves across routes. Motivated by the profit motive of informal and privatized transit operators, we assume drivers choose routes that maximize their individual profits (see middle of Figure 1). Let $p_i = \bar{p}_i - \frac{c_i}{F} \geq 0$ denote the per-rider profit on route i, defined as the difference between the trip fare and the per-rider operating cost. Then, given x_i drivers serving $\Lambda_i^M(x_i)$ riders on route i, the total route profit is $p_i\Lambda_i^M(x_i)$, which we assume is shared equally among drivers, yielding a per-driver profit $\pi_i(x_i) = \frac{\Lambda_i^M(x_i)p_i}{x_i}$.[1] Drivers select routes to maximize this payoff, taking the rider demand functions and the allocation of other drivers as given. This strategic interaction induces a Wardrop-style (Wardrop 1952) equilibrium driver allocation, in which no driver can profitably deviate, as elucidated below.

[1] $\pi_i(\cdot)$ is well-defined and bounded for $x_i \geq 0$ under the class of minibus rider demand functions we study in Equation (3).

Definition 2 (Equilibrium Driver Allocation)**.** Consider a driver allocation $\mathbf{x} = (x_i)_{i\in[n]}$, with $\Lambda_i^M(\cdot)$ denoting the minibus rider demand function on each route i. Then, $\mathbf{x}$ is an equilibrium driver allocation if for any route i with $x_i > 0$, it holds that $\pi_i(x_i) \geq \pi_j(x_j)$ for all routes $j \in [n]$.

3.4 Price of Anarchy

Since drivers selfishly maximize individual profits, the resulting equilibrium allocation of drivers across routes may not be optimal from a system planner's perspective. In this work, we evaluate the equilibrium outcomes for an informal transit instance $I = (n, F, \overline{\mathbf{p}}, \mathbf{l}, \mathbf{c}, t_1, t_2, D, \mathbf{\Lambda}^M)$ using two metrics of direct relevance to a system planner: (i) *cumulative driver profits*, defined as $P_I(\mathbf{x}) = \sum_{i=1}^n p_i\Lambda_i^M(x_i)$, an aggregate notion of driver welfare, and (ii) *cumulative rider demand served* by the minibus system, given by $R_I(\mathbf{x}) = \sum_{i=1}^n \Lambda_i^M(x_i)$, an aggregate notion of rider welfare.

To quantify the performance loss under these metrics due to the decentralized profit-maximizing route choices of drivers, we adopt the classical notion of the *price of anarchy* (PoA), which quantifies the worst-case ratio (over problem instances) between the equilibrium system objective and optimal system objective obtained under a centralized authority coordinating the driver's actions.

Definition 3 (Price of Anarchy)**.** For a given family of demand functions $\mathcal{L}$, the profit price of anarchy, denoted P-PoA($\mathcal{L}$), is defined to be the worst case profit ratio over informal transit instances (see Definition 1) of the optimum cumulative driver profit to that achieved by an equilibrium driver allocation. The rider welfare price of anarchy, denoted R-PoA($\mathcal{L}$), is defined similarly.

$$\text{P-PoA}(\mathcal{L}) = \sup_{I\in\mathcal{I}_\mathcal{L}} \sup_{\mathbf{x}^*\in\Omega} \sup_{\mathbf{x}^{Eq}\in\Omega^{Eq}} \frac{P_I(\mathbf{x}^*)}{P_I(\mathbf{x}^{Eq})} \qquad \text{R-PoA}(\mathcal{L}) = \sup_{I\in\mathcal{I}_\mathcal{L}} \sup_{\mathbf{x}^*\in\Omega} \sup_{\mathbf{x}^{Eq}\in\Omega^{Eq}} \frac{R_I(\mathbf{x}^*)}{R_I(\mathbf{x}^{Eq})} \tag{2}$$

In the remainder of this work, we focus on the class of demand functions $\mathcal{L}_V$ induced under an adaptation of Vickrey's bottleneck model (see Section 3.2). Hence, for notational simplicity, we will drop the dependency on $\mathcal{L}$ and use P-PoA, R-PoA in place of P-PoA($\mathcal{L}$), R-PoA($\mathcal{L}$) respectively. Similarly, we will use P, R instead of P_I, R_I when the choice of instance I is unambiguous and clear.

4 Price of Anarchy Bounds

We begin the study of our informal transit system through a price of anarchy (PoA) analysis under both cumulative driver profit and rider welfare metrics introduced in Section 3.4. To this end, under the rider behavior model in Section 3.2, we first derive the minibus rider demand as a function of driver supply and analyze its key properties in Section 4.1. Leveraging this characterization, we establish tight PoA bounds for cumulative driver profit and rider welfare in Sections 4.2 and 4.3, respectively. Our results show that decentralized, self-interested driver decisions can lead to bounded yet substantial inefficiencies in both objectives, with comparatively larger losses in rider welfare.

4.1 Derivation of Minibus Demand Function Under Rider Queuing

This section derives the rider demand served by the minibus as a function of driver supply on a route and studies its properties, which underpin our PoA analysis in Sections 4.2 and 4.3.

We begin by presenting the main result of the section, which characterizes, for any driver supply x_i on a route, the total rider demand served by the minibus at equilibrium under the minibus rider costs defined in Equation (1). In presenting this result, we focus on the regime when the cost difference between the outside option and the minibus without rider queuing and schedule delays,

given by $S_i = c_i^O - c_i^M$, is non-negative. Recall that when $S_i < 0$, regardless of the driver supply, the outside option is always preferred to the minibus, i.e., $\Lambda_i^M(x_i) = 0$ for all x_i.

Proposition 1 (Minibus Demand Under Rider Queuing)**.** *Consider an informal transit system with a fixed menu of routes, where riders on each route $i \in [n]$ choose either the minibus and incur a cost as defined in Equation* (1) *or an outside option with a fixed cost c_i^O. Further, suppose that $S_i = c_i^O - c_i^M \geq 0$, and define the driver allocation threshold $\tilde{k}_i^* = \min\left\{\frac{2l_i\lambda_i}{F}, \frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}\right\}$. Then:*

$$\Lambda_i^M(x_i) = \begin{cases} \frac{F}{2l_i}\left(t_2 - t_1 + S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)x_i - \left(\left(\frac{F}{2l_i}\right)^2 S_i\frac{\eta_E+\eta_L}{\eta_E\eta_L}\frac{t_2-t_1}{\Lambda_i}\right)x_i^2, & \text{if } x_i \in [0, \tilde{k}_i^*) \\ \Lambda_i, & \text{if } x_i \geq \tilde{k}_i^* \end{cases} \tag{3}$$

Proof (Sketch). Fix a route i with $S_i \geq 0$ and x_i drivers, resulting in a service rate of $\mu_i(x_i) = \frac{x_iF}{2l_i}$. Then, defining the time horizon over which drivers service the route as $T_i(x_i)$, the total mass of riders served is: $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i)$. Note $T_i(x_i) \geq \Delta = t_2 - t_1$, as riders may be willing to arrive earlier or later than their desired destination arrival times under the costs in Equation (1).

To establish an expression for $T_i(x_i)$, we first characterize the equilibrium of riders' mode choice (minibus versus outside option) and arrival-time decisions on a route i with x_i drivers by adapting the equilibrium characterization in Vickrey's bottleneck model with an outside option from Gonzales and Daganzo (2012) to our setting (see Proposition 5 in Appendix B.1). Combining this equilibrium characterization with the property that rider's desired destination arrival times are uniformly distributed, we derive a closed-form expression for the total service time. In particular, when the system is under-supplied (i.e., $\mu_i(x_i) < \lambda_i$), $T_i(x_i) = \Delta + \min\{S_i, \bar{S}(x_i)\}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\left(1 - \frac{\Delta x_iF}{2l_i\Lambda_i}\right)$, where $\bar{S}(x_i) = \frac{\Lambda_i\eta_E\eta_L}{\mu_i(x_i)(\eta_E+\eta_L)}$ (see Corollary 4 in Appendix B.1). When the system is over-supplied (i.e., $\mu_i(x_i) \geq \lambda_i$), $T_i(x_i) = \Delta$.

Finally, we evaluate the rider demand $\Lambda_i^M(x_i) = \mu_i(x_i)T_i(x_i)$ by substituting the derived service time relation in the different regimes for driver supply, yielding the expression in Equation (3). □

For a complete proof, see Appendix B.1. Proposition 1 implies that the rider demand function is concave quadratic up to a driver supply threshold $\tilde{k}_i^*$, beyond which all rider demand is served. The resulting demand function in Equation (3), shown on the left of Figure 2, highlights a key departure of our setting from standard equilibrium models, in which full demand capture occurs only once the saturation condition $\mu_i(x_i) = \lambda_i$ is met, as with the capacity-constrained demand function on the right of Figure 2. Specifically, Proposition 1 implies that all rider demand is served at a threshold $\tilde{k}_i^*$ which may be strictly smaller than the saturation threshold $k_i^* = \frac{2l_i\lambda_i}{F}$ at which $\mu_i(x_i) = \lambda_i$. Thus, all rider demand can be served even in an under-supplied system ($\mu_i(x_i) < \lambda_i$) if the outside option is sufficiently unattractive, albeit while inducing rider queuing. Once the driver supply exceeds the saturation threshold $k_i^* = \frac{2l_i\lambda_i}{F}$, rider queuing is eliminated at equilibrium.

We now leverage Proposition 1 to establish key continuity and monotonicity properties.

Corollary 1 (Continuity and Monotonicity of Minibus Rider Demand)**.** *The minibus rider demand function $\Lambda_i^M(\cdot)$ given by Equation* (3) *for any route i is continuous and monotonically non-decreasing in the driver allocation x_i on that route.*

Corollary 2 (Continuity and Monotonicity of Per-Driver Profits)**.** *Consider a route i with a demand function $\Lambda_i^M(\cdot)$ in Equation* (3)*. Then, the per-driver profit $\pi_i(x_i) = \frac{\Lambda_i^M(x_i)p_i}{x_i}$ is continuous and non-increasing in the driver allocation x_i on route i.*

For proofs of Corollaries 1 and 2, see Appendices B.2 and B.3, respectively. Corollary 1 implies a natural monotonicity condition that the served rider demand is non-decreasing in the driver supply, implying that the cumulative profit $p_i\Lambda_i^M(x_i)$ on a route is also non-decreasing. However, Corollary 2 shows that per-driver profit decreases in driver supply, a consequence of the concavity of the rider demand function in Equation (3) and, hence, of the profit function $p_i\Lambda_i^M(x_i)$.

4.2 Price of Anarchy for Cumulative Driver Profit

Building on the rider demand function characterization in Proposition 1 and the monotonicity properties in Corollaries 1 and 2, we now analyze the PoA in our informal transit system with respect to cumulative driver profits. We show that the profit PoA is at most 2 and that this bound is tight: for any $\epsilon>0$, there exists an instance $I \in \mathcal{I}_{\mathcal{L}_V}$ for which the profit ratio is at least $2-\epsilon$.

Theorem 1 (PoA Upper Bound for Cumulative Driver Profit)**.** *Consider an informal transit system with a class $\mathcal{L}_V$ of rider demand functions $(\Lambda_i^M(\cdot))_{i\in[n]}$ in Equation (3). Then, P-PoA ≤ 2.*

Proposition 2 (Tightness of P-PoA Bound)**.** *For any $\epsilon > 0$, there exists an informal transit instance $I \in \mathcal{I}_{\mathcal{L}_V}$ such that the profit ratio exceeds $2-\epsilon$.*

These results show that in the worst case, selfish driver behavior can reduce cumulative driver profits to half of the centralized optimum, implying a substantial efficiency loss. Yet, unlike settings with unbounded PoAs (Kannan et al. 2013), the resulting loss in cumulative driver profits remains provably bounded, a consequence of the concavity of the demand function in Equation (3). We now prove these results.

Proof of Theorem 1. Fix an informal transit instance I with rider demand functions given by Equation (3) for all routes $i \in [n]$. Since our proof applies to all feasible instances I, we drop the dependency on I in our notation. For this instance, let $\mathbf{x}^*$ be the cumulative profit-maximizing allocation, i.e., $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\Omega} P(\mathbf{x}) = \sum_{i=1}^n p_i\Lambda_i^M(x_i)$, and let $\mathbf{x}^{Eq}$ be any equilibrium allocation. Moreover, let Q be the set of routes for which the the cost difference $S_i \geq 0$. Note for any route $i \in [n]\backslash Q$, the minibus serves no rider demand and generates no profit regardless of the allocation.

At any equilibrium allocation $\mathbf{x}^{Eq}$, all routes i with $x_i^{Eq} > 0$ have the same per-driver profit, denoted by $\pi^{Eq} = \frac{\Lambda_i^M(x_i^{Eq})p_i}{x_i^{Eq}}$. Moreover, by Definition 2, $\pi^{Eq} \geq \pi_j(x_j^{Eq})$ for all $j \in [n]$. Thus, the total profit under $\mathbf{x}^{Eq}$ can be re-expressed as $P(\mathbf{x}^{Eq}) = \sum_{i\in Q} p_i\Lambda_i^M(x_i^{Eq}) \overset{(a)}{=} \sum_{i\in Q} \pi_i(x_i^{Eq})x_i^{Eq} \overset{(b)}{=} \sum_{i\in Q} \pi^{Eq}x_i^{Eq}$, where (a) follows from the definition of $\pi_i(\cdot)$ and (b) follows as $\pi^{Eq} = \pi_i(x_i^{Eq})$ for all routes i with $x_i^{Eq} > 0$. Analogously, letting $\pi_i^* = \pi_i(x_i^*)$ be the per-driver profit on each route $i \in Q$ at the profit-maximizing allocation $\mathbf{x}^*$, the optimal cumulative profit is $P(\mathbf{x}^*) = \sum_{i\in Q} \pi_i^* x_i^*$.

Defining the sets $L_1 = \{i \in Q : x_i^{Eq} \geq x_i^*\}$ and $L_2 = \{i \in Q : x_i^{Eq} < x_i^*\}$, we have:

$$P(\mathbf{x}^*) = \sum_{i\in Q} \pi_i^* x_i^* = \sum_{i\in L_1} \pi_i^* x_i^* + \sum_{i\in L_2} \pi_i^* x_i^* \overset{(a)}{\leq} \sum_{i\in L_1} \pi^{Eq} x_i^{Eq} + \sum_{i\in L_2} \pi_i^* x_i^* \overset{(b)}{\leq} P(\mathbf{x}^{Eq}) + \sum_{i\in L_2} \pi_i^* x_i^*,$$
$$\overset{(c)}{\leq} P(\mathbf{x}^{Eq}) + \sum_{i\in L_2} \pi^{Eq} x_i^* \overset{(d)}{\leq} P(\mathbf{x}^{Eq}) + P(\mathbf{x}^{Eq}) = 2P(\mathbf{x}^{Eq})$$

where (a) follows from Corollary 1, as $x_i^{Eq} \geq x_i^*$ for all $i \in L_1$, (b) follows as $\sum_{i\in L_1} \pi^{Eq}x_i^{Eq} \leq \sum_{i\in Q} \pi^{Eq}x_i^{Eq}$ as $L_1 \subseteq Q$, (c) follows as $\pi^{Eq} \geq \pi_i^*$ for all $i \in L_2$ by Corollary 2, and (d) follows as $\sum_{i\in L_2} \pi^{Eq}x_i^* \leq \sum_{i\in Q} \pi^{Eq}x_i^* = \pi^{Eq}\sum_{i\in Q} x_i^{Eq}$ as $L_2 \subseteq Q$ and $\sum_{i\in Q} x_i^* = \sum_{i\in Q} x_i^{Eq} = D$. The above analysis holds for any equilibrium allocation $\mathbf{x}^{Eq}$ and instance I, thus establishing our claim. □

Proof of Proposition 2. Consider an instance with $n = 2$ routes and a driver supply $D = 1$. Moreover, suppose that $S_i = 0$ on both routes. Then, the minibus rider demand function on both routes reduces to the capacity-constrained demand function shown on the right of Figure 2:

$$\Lambda_i^M(x_i) = \begin{cases} \frac{F(t_2-t_1)}{2l_i} x_i, & \text{if } x_i \in [0, \tilde{k}_i^*] \\ \Lambda_i, & \text{if } x_i > \tilde{k}_i^*, \end{cases} \tag{4}$$

where $\tilde{k}_i^* = k_i^* = \frac{2l_i\lambda_i}{F}$. Given Equation (4), consider an instance where $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$. Then, the cumulative profit-maximizing allocation is $\mathbf{x}^* = (\tilde{k}_1^*, \tilde{k}_2^*)$, resulting in a total profit $P(\mathbf{x}^*) = p_1\Lambda_1 + p_2\Lambda_2$. Next, suppose that the per-driver profit on the two routes satisfies $\pi_1(1) = \pi_2(0)$, i.e., the system supports an equilibrium $\mathbf{x}^{Eq} = (1, 0)$. Then, the following equalities must hold: $p_1\Lambda_1 = \pi_1(1) = \pi_2(0) = p_2\frac{F(t_2-t_1)}{2l_2} = p_2\Lambda_2\frac{F}{2l_2\lambda_2} = p_2\Lambda_2\frac{1}{\tilde{k}_2^*}$. Using this relation, the profit ratio between the optimal and equilibrium allocation is given by: $\frac{P(\mathbf{x}^*)}{P(\mathbf{x}^{Eq})} = \frac{p_1\Lambda_1+p_2\Lambda_2}{p_1\Lambda_1} = 1+\frac{p_2\Lambda_2}{p_1\Lambda_1} = 1+\tilde{k}_2^*$. Finally, taking the limit as $\tilde{k}_2^* \to 1$ and $\tilde{k}_1^* \to 0$ while satisfying $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$, the above analysis implies that the profit ratio approaches two, establishing our claim. □

These proofs rely on key properties of our informal transit system established in Section 4.1. The profit PoA upper bound in Theorem 1 follows from the monotonicity of the minibus rider demand and per-driver profits (Corollaries 1 and 2) along with the concavity of the rider demand function in Equation (3). To establish the tightness of this bound, Proposition 2 constructs an instance in which the cost difference $S_i = 0$ for all routes i. In this regime, the rider demand function reduces to a piecewise-linear form depicted on the right of Figure 2, implying that the capacity-constrained demand function corresponds to the worst-case profit PoA instance.

4.3 Price of Anarchy for Rider Welfare

This section extends the PoA analysis from the previous section to the rider welfare metric. Letting $p_{\max}$ and $p_{\min}$ denote the maximum and minimum per-rider minibus profit across the n routes, we show that the PoA with respect to cumulative rider welfare is bounded above by $1+\frac{p_{\max}}{p_{\min}}$. Moreover, we show that this bound is tight akin to Proposition 2. These results are formalized below.

Theorem 2 (PoA Upper Bound for Rider Welfare)**.** *Consider an informal transit system with a class $\mathcal{L}_V$ of rider demand functions $(\Lambda_i^M(\cdot))_{i\in[n]}$ in Equation (3). Then, R-PoA $\leq 1 + \frac{p_{\max}}{p_{\min}}$.*

Proposition 3 (Tightness of R-PoA Bound)**.** *For any $\epsilon > 0$, there exists an informal transit instance $I \in \mathcal{I}_{\mathcal{L}_V}$ such that rider welfare ratio exceeds $1 + \frac{p_{\max}}{p_{\min}} - \epsilon$.*

These results highlight that decentralized, selfish driver decisions can substantially reduce the cumulative rider demand served to as little as a $1/(1 + \frac{p_{\max}}{p_{\min}})$ fraction of the demand that can be served under a centralized optimum that coordinates driver actions. Since $\frac{p_{\max}}{p_{\min}} \geq 1$, in the worst case, selfish driver behavior is more detrimental to rider welfare than to cumulative driver profits (see Section 4.2). That said, in practical settings, the ratio $\frac{p_{\max}}{p_{\min}}$ is typically small, as suggested by our numerical experiments (see Section 7); for instance, using data from Nalasopara, India, this ratio equals three, resulting in a R-PoA of four. While the proof of Proposition 3 follows analogously to the proof of Proposition 2 (see Appendix B.5), we emphasize that Theorem 2 does not follow directly from the profit PoA bound in Theorem 1. In particular, a naive application of the profit PoA bound would yield a rider welfare PoA of $2\frac{p_{\max}}{p_{\min}}$. However, establishing the sharper bound in Theorem 2 requires additional arguments and a different set of inequalities (see Appendix B.4).

5 Mechanism I: Cross-subsidization

To mitigate the efficiency losses arising from selfish driver behavior, we study mechanisms through which a public authority can steer informal and privatized transit operators toward improved system outcomes in terms of cumulative driver profit and rider welfare. This section studies *cross-subsidization*, in which the public authority sets route-specific tolls and subsidies to influence driver route choices and shape service patterns. The key idea is that tolls can deter excessive driver entry on highly lucrative routes, with the resulting revenue used to subsidize service on less profitable routes. Such mechanisms have practical antecedents in adjacent domains such as ride-hailing (Uber Blog 2022).

In this section, we show that for any instance of our informal transit system, there exists a budget-balanced (i.e., zero net expenditure) cross-subsidization scheme, which can be derived in closed form, that aligns individual driver incentives with any desired system objective, and can be computed in polynomial time for both cumulative driver profit and rider welfare objectives. These results demonstrate that by appropriately accounting for driver incentives, a public authority can eliminate the PoA inefficiencies identified in Section 4 without incurring any net fiscal costs.

We first introduce cross-subsidization and define the notion of a *budget-balanced* cross-subsidy scheme. In our setting with cross-subsidies, a public authority assigns each route $i \in [n]$ a transfer $\tau_i \in \mathbb{R}$, which influences driver payoffs. Negative transfers represent tolls paid by drivers to operate on a route, while positive transfers represent subsidies. Then, under a driver allocation x_i on route i, drivers earn transfer-adjusted profits $\tilde{\pi}(x_i) = \pi_i(x_i) + \tau_i$. With these modified payoffs, the equilibrium driver allocation under cross-subsidies can be defined analogously to Definition 2, with the two notions coinciding when the transfer vector $\boldsymbol{\tau} = (\tau_i)_{i\in[n]} = \mathbf{0}$. Crucially, cross-subsidies affect only driver payoffs and have no direct impact on rider payoffs or the rider demand function, with riders being influenced only through induced changes in driver allocations across routes. We focus, in particular, on *budget-balanced* cross-subsidy schemes, under which the induced equilibrium driver allocation $\mathbf{x}$ satisfies $\sum_{i=1}^{n} \tau_i x_i = 0$, so that the collected revenues finance driver subsidies.

We now characterize and establish the existence of a budget-balanced cross-subsidy scheme that induces any target driver allocation $\mathbf{x}^*$, e.g., those maximizing cumulative driver profits or rider welfare, as an equilibrium. Moreover, the corresponding cross-subsidy transfer vector yields per-driver profits that are a weighted combination of their profits at $\mathbf{x}^*$ without cross-subsidies.

Theorem 3 (Optimal Budget-Balanced Cross-Subsidization Scheme)**.** *For any informal transit instance with a class $\mathcal{L}_V$ of rider demand functions in Equation* (3)*, let $\mathbf{x}^* \in \Omega$ be a target driver allocation with $x_j^* > 0$ for at least one route j. Then, there exists a budget-balanced cross-subsidization scheme $\boldsymbol{\tau} = (\tau_i)_{i\in[n]}$, which induces $\mathbf{x}^*$ as an equilibrium driver allocation. Moreover, $\boldsymbol{\tau}$ can be derived in closed-form and induces equilibrium per-driver profits $\tilde{\pi}^{Eq} = \frac{\sum_{i\in[n]} \pi_i^* x_i^*}{\sum_{i\in[n]} x_i^*}$, where $\pi_i^* = \pi_i(x_i^*)$.*

Proof (Sketch). For each route i, define the per-driver profit at $\mathbf{x}^*$ (without cross-subsidies) as $\pi_i^* = \pi_i(x_i^*) = \frac{\Lambda_i^M(x_i^*)p_i}{x_i^*}$. To induce $\mathbf{x}^*$ as an equilibrium under a cross-subsidy scheme $\boldsymbol{\tau}$, for any route i with $x_i^* > 0$, its transfer-adjusted per-driver profits are at least that of any other route. Thus, we construct $\boldsymbol{\tau}$ that satisfies this equilibrium condition with equality for all routes, i.e., $\pi_i^* + \tau_i = \pi_j^* + \tau_j$ for all i, j. Combining these relations with the budget-balance condition $\sum_{i\in[n]} \tau_i x_i^* = 0$ yields a system of n (unique) equations with n unknowns (see Equation (7)), corresponding to the entries of $\boldsymbol{\tau}$. This linear system admits a unique solution if there is some

route j with $x_j^* > 0$, and solving it yields an expression for $\boldsymbol{\tau}$, under which the per-driver profits are $\tilde{\pi}^{Eq}$. □

For a complete proof, see Appendix B.6. Theorem 3 shows that cross-subsidies can eliminate PoA inefficiencies arising from selfish driver behavior in informal transit systems by aligning driver incentives with any target driver allocation. In this sense, cross-subsidies play a role akin to marginal-cost pricing in congestion games, which restores efficiency under selfish routing (Roughgarden 2005); however, unlike marginal-cost pricing, our cross-subsidy scheme is budget-balanced.

The constructive proof of Theorem 3 implies a natural algorithm to implement any desired driver allocation as an equilibrium. First, a public authority computes a target allocation $\mathbf{x}^*$ that maximizes a chosen objective, such as cumulative driver profit or rider welfare. It then computes the per-driver profits $\pi_i^* = \pi_i(x_i^*) = \frac{\Lambda_i^M(x_i^*)p_i}{x_i^*}$ using the rider demand function in Equation (3) for each route i. Finally, it sets the transfer vector $\boldsymbol{\tau}$ that induces $\mathbf{x}^*$ as an equilibrium by solving the linear system of equations derived in Theorem 3's proof (see Equation (7)). This procedure implies that the efficacy of cross-subsidization is limited only by the public authority's ability to compute the allocation $\mathbf{x}^*$. As a corollary, the optimal allocations for both cumulative driver profit and rider welfare objectives can be implemented in polynomial time as equilibria via a budget-balanced cross-subsidy scheme.

Corollary 3. *For any informal transit instance with a class $\mathcal{L}_V$ of demand functions in Equation* (3)*, let $\mathbf{x}^*$ be the cumulative driver profit or rider welfare-maximizing allocation. Then, there exists a budget-balanced cross-subsidy scheme, computable in polynomial time, that induces $\mathbf{x}^*$ as an equilibrium.*

The proof of this result follows by combining Theorem 3 with the fact that the rider demand functions in Equation (3) are concave quadratic and the feasible set Ω of driver allocations is convex, enabling the cumulative driver profit and rider welfare maximization problems to be solved in polynomial time. Beyond the above algorithmic and computational results, we provide additional properties of the budget-balanced cross-subsidy scheme derived in Theorem 3 in Appendix C.2, and discuss multiplicity and uniqueness of induced equilibrium outcomes in Appendices C.3 and C.4.

6 Mechanism II: Stackelberg Routing

While cross-subsidization can mitigate inefficiencies arising from selfish driver behavior, it is often difficult to implement in practice, particularly in informal transit systems, where monitoring and enforcement are challenging, and compliance is limited. To that end, this section studies another operationally feasible policy lever that does not rely on monetary transfers to drivers and instead involves reallocating the public authority's own fleet of drivers across routes (i.e., determining where and with what frequency to offer public transit) to influence the routing decisions of informal or privatized transit drivers. Such a mechanism is especially relevant in settings where both public transit and informal or privatized transit operators coexist, as in many cities worldwide.

We model such mixed public–private transit systems through a Stackelberg routing framework, introduced in Section 6.1. While computing an optimal Stackelberg strategy under both cumulative driver profit and rider welfare objectives is NP-hard in general, we propose two algorithms, *Lowest Profit First* (LPF) and *Linearized Non-Compliant First* (L-NCF), with provably strong guarantees. We show that LPF is Stackelberg-optimal for both objectives for $n = 2$ routes (Section 6.2), and that L-NCF, our main algorithmic contribution, achieves near-optimal performance across both

metrics for an arbitrary number of routes under practically relevant conditions (Section 6.3). We view L-NCF as being of independent interest, offering a new perspective on classical Stackelberg routing algorithms and suggesting that similar guarantees may hold in related settings.

6.1 Stackelberg Routing Setup and Hardness

We introduce a Stackelberg routing framework capturing the interaction between centralized public transit and decentralized informal operators, and show that computing optimal Stackelberg strategies under cumulative driver profit and rider welfare objectives is, in general, NP-hard.

Building on the informal transit system introduced in Section 3, we consider a setting in which a public authority centrally controls an $\alpha \in [0,1]$ fraction of drivers, representing public transit provision, while the remaining drivers operate as privatized transit providers. In line with the Stackelberg routing literature (Roughgarden 2005), we assume that that centrally controlled and privately operated drivers are identical in all operational characteristics (e.g., per-trip costs, revenues, etc.) specified in Section 3, and differ only in their incentives.[2] In particular, the centrally controlled α fraction of drivers are assigned to routes by a public authority, akin to determining public transit service locations and frequencies, while privatized drivers independently choose profit-maximizing routes in response, giving rise to a Stackelberg routing game. For any such centralized driver allocation, this interaction induces an equilibrium response of the decentralized operators, as defined below.

Definition 4 (Induced Driver Equilibrium and Allocation)**.** Consider the set of feasible allocations for centrally controlled drivers, $\Omega_\alpha = \{\mathbf{x} \in \mathbb{R}^n_{\geq 0} : \sum_{i=1}^n x_i = \alpha D\}$, and fix an allocation $\mathbf{x}^C \in \Omega_\alpha$. Moreover, for each route $i \in [n]$, define the adjusted per-driver profit $\hat{\pi}_i(x_i) = \pi_i(x_i^C + x_i)$. Then, an allocation $\mathbf{x}^P \in \Omega_{1-\alpha}$ is said to be an *equilibrium induced by* $\mathbf{x}^C$ if it constitutes an equilibrium driver allocation for the original informal transit instance with $(1-\alpha)D$ drivers and per-driver profit functions $\hat{\pi}_i(\cdot)$. The resulting equilibrium driver allocation induced by $\mathbf{x}^C$ is given by $\mathbf{x} = \mathbf{x}^C + \mathbf{x}^P$.

For a given informal transit instance and centrally controlled fraction α, let $\Omega(\mathbf{x}^C)$ denote the set of equilibrium driver allocations induced by the centrally controlled allocation $\mathbf{x}^C$. Then, we define a centrally controlled allocation $\mathbf{x}^{C,*}$ to be *Stackelberg optimal* for the cumulative driver profit objective if $\mathbf{x}^{C,*} \in \arg\max_{\mathbf{x}^C \in \Omega_\alpha} \min_{\mathbf{x} \in \Omega(\mathbf{x}^C)} P(\mathbf{x})$, with an analogous definition for the rider welfare objective. In words, a Stackelberg optimal allocation corresponds to determining the service locations and frequencies of a public transit service to maximize a chosen system objective, anticipating the equilibrium response of decentralized informal or privatized operators.

Importantly, we define Stackelberg optimality with respect to system-level equilibrium outcomes rather than the cumulative profits or rider demand served by the centrally controlled drivers alone. This system-level perspective provides a natural lens for evaluating transit policies, as it captures the performance of the transit system as a whole, including centrally controlled and decentralized drivers, on both rider welfare and cumulative driver profit (a measure of driver welfare) objectives, which are particularly policy-relevant when privatized operators constitute a large share of service provision. Accordingly, our analysis and proposed LPF and L-NCF algorithms optimize system-wide outcomes, and our numerical results in Section 7 show substantial losses when a public authority adopts a *Greedy* strategy, which optimizes either objective for the centrally controlled

[2]This assumption isolates the role of incentives, the primary focus of our study; however, alternative models capturing additional heterogeneity between public and informal transit operators are an interesting direction for future work.

drivers alone, reflecting the status quo in which privatized operators' incentives are ignored in transit planning.

We conclude this section by showing that computing the optimal Stackelberg strategy is, in general, NP-hard, for both the cumulative driver profit and rider welfare objectives.

Proposition 4 (NP-Hardness of Stackelberg Routing)**.** *Consider a mixed public and informal transit system with a class $\mathcal{L}_V$ of demand functions $(\Lambda_i^M(\cdot))_{i\in[n]}$ in Equation* (3)*. Then, computing the optimal Stackelberg strategy is NP-hard for both cumulative driver profit and rider welfare objectives.*

The proof of this result follows ideas similar to the hardness result in Roughgarden (2001); thus, we defer its details to Appendix B.7. The key distinction in our setting is that, unlike the linear costs in the scheduling model of Roughgarden (2001), per-driver profits on a route in our model have a piecewise functional form: linear up to a threshold $\tilde{k}_i^*$, and nonlinear beyond that point, taking the form $\pi_i(x_i) = \frac{\Lambda_i p_i}{x_i}$ for all $x_i \geq \tilde{k}_i^*$ (see Appendix B.3). Accordingly, the challenge in our reduction involves constructing instances of our setting in which the thresholds $\tilde{k}_i^*$ are sufficiently large so that Stackelberg equilibrium driver allocations lie in the linear portion of the per-driver profit function across routes.

6.2 Lowest Profit First

Given the hardness of computing optimal Stackelberg strategies, we turn to designing efficient algorithms that closely approximate the optimal Stackelberg outcomes. Unlike the status-quo *Greedy* baseline that does not account for the incentives of privatized operators (see Section 7), our algorithms explicitly account for privatized operators' incentives by allocating centrally controlled drivers on less lucrative routes while leaving the more profitable ones to decentralized, profit-maximizing drivers, thereby de-congesting high-profit corridors while improving overall coverage.

This section presents our first algorithm, *Lowest Profit First* (LPF), which adapts the classical lowest latency first strategy of Roughgarden (2001) to our setting. Although LPF is well defined for any number of routes, our main contribution is a characterization of its performance in the two-route case ($n = 2$), a setting not analyzed in Roughgarden (2001). We focus on the $n = 2$ route setting, as extending the analysis to general n would closely parallel techniques in Roughgarden (2001) and, while yielding different bounds in our setting, would provide limited additional insight. For completeness, we therefore complement our theoretical results for $n = 2$ routes with an empirical evaluation of LPF for $n > 2$ routes in Section 7.

We study LPF under both cumulative driver profit and rider welfare objectives, denoted LPF-P and LPF-R, respectively. Both LPF variants are identical other than their computation of a target allocation $\mathbf{x}^* \in \Omega$. In LPF-P, $\mathbf{x}^*$ maximizes cumulative driver profits, while in LPF-R, it maximizes rider welfare, assuming all drivers are centrally controlled. Given $\mathbf{x}^*$, routes are ordered in descending order of per-driver profits such that $\pi_1(x_1^*) \geq \pi_2(x_2^*) \geq \ldots \geq \pi_n(x_n^*)$. Then, the Stackelberg allocation $\mathbf{x}^C \in \Omega_\alpha$ is computed by allocating the α fraction of centrally-controlled drivers to routes in the ascending order of per-driver profits, sequentially saturating routes at $x_i^C = x_i^*$ starting with the lowest-profit route n until the centrally controlled supply is exhausted.

We show that the LPF strategies compute the optimal Stackelberg allocation for both objectives for $n = 2$ routes. This result contrasts the hardness result in Proposition 4, which relies on a reduction from the $\frac{2}{3} - \frac{1}{3}$ partition problem (see Appendix B.7) and thus requires constructing instances with $n > 2$ routes. For the two-route case, we also derive bounds on the cumulative

driver profit and rider welfare ratios, i.e., the ratio of the optimal cumulative driver profit or rider welfare (under full centralized control) to that achieved under LPF as $\alpha \in [0,1]$ is varied.

Theorem 4 (Optimality of LPF for $n = 2$ Routes)**.** *Consider a mixed public and informal transit system with $n = 2$ routes and a class $\mathcal{L}_V$ of demand functions in Equation (3). Then, for any $\alpha \in [0,1]$:*

1. *LPF-P (LPF-R) computes the profit-maximizing (rider welfare-maximizing) Stackelberg strategy.*
2. *Let $\mathbf{x}^*$ be the profit-maximizing allocation (assuming full centralized control) and $\tilde{\mathbf{x}}_\alpha$ be the induced LPF-P allocation for any $\alpha \in [0,1]$. Then, the profit ratio $\frac{P(\mathbf{x}^*)}{P(\tilde{\mathbf{x}}_\alpha)} \leq 2-\alpha$. Moreover, for all α and any $\epsilon > 0$, there exists an instance such that LPF-P's profit ratio is at least $\frac{2}{1+\alpha} - \epsilon$.*
3. *Letting $\mathbf{x}^*$ be the rider welfare-maximizing allocation and $\tilde{\mathbf{x}}_\alpha$ be the LPF-R allocation given α, then the rider welfare ratio $\frac{R(\mathbf{x}^*)}{R(\tilde{\mathbf{x}})} \leq 1 + \frac{p_{\max}}{p_{\min}}(1-\alpha)$, where $p_{\max}$ and $p_{\min}$ are the maximum and minimum per-rider minibus profits across routes. Moreover, for all α and any $\epsilon > 0$, there exists an instance such that LPF-R's rider welfare ratio is at least $\frac{p_{\max}+p_{\min}}{\alpha p_{\max}+p_{\min}} - \epsilon$.*

For a proof of Theorem 4, see Appendix B.8. The Stackelberg optimality of LPF under both objectives does not extend beyond the two-route setting. When $n = 2$, exactly one route is less lucrative (in terms of per-driver profits) than the most lucrative route under the optimal allocation, making the Stackelberg optimal action unambiguous. When $n > 2$, multiple less lucrative routes exist, introducing non-trivial trade-offs in how centrally controlled drivers should be allocated.

Beyond the Stackelberg optimality of LPF, Theorem 4 also provides bounds on the profit and welfare ratios of LPF-P and LPF-R, respectively, which smoothly interpolate between the fully centralized and fully decentralized PoA bounds in the two-route setting. When $\alpha = 1$, corresponding to full centralization, LPF achieves the optimal allocation under both objectives; when $\alpha = 0$, the upper and lower bounds on the profit and rider welfare ratios coincide with the fully decentralized PoA bounds in Section 4. For intermediate $\alpha \in (0,1)$, these bounds exhibit a natural monotonic, at least linear, improvement, with the worst-case driver profit and rider welfare ratios of both LPF strategies converging to one as α increases, reflecting the gains from increasing centralization.

6.3 Linearized Non-Compliant First

This section studies the multi-route setting for the Stackelberg routing problem under the cumulative driver profit objective. Given Proposition 4, we do not expect to find an optimal algorithm that runs in polynomial time. Instead, we establish an approximation ratio for a Linearized variant of the Non-Compliant First (NCF) algorithm. Krichene et al. (2014) proposed NCF and showed that it is Stackelberg optimal for congestion routing problems whose objectives satisfy a specific structure, including an initial region where the total latency on an edge is linear in the number of users traversing the edge. Motivated by this, we approximate our problem by modifying our objective to one with structural similarities to those studied in Krichene et al. (2014), and bound the loss under this approximation.

Linearized Non-Compliant First Algorithm (L-NCF): In this algorithm, NCF is applied to a piecewise linear approximation to the profit function $P(\cdot)$. L-NCF constructs $\widetilde{P}(\cdot)$ with $\widetilde{P}(\mathbf{x}) := \sum_i p_i \widetilde{\Lambda}_i^M(x_i)$, where $\widetilde{\Lambda}_i^M(x_i) := \Lambda_i \min\left(\frac{x_i}{k_i^*}, 1\right)$ is a piecewise linear under-approximation of $\Lambda_i^M(\cdot)$

which matches $\Lambda_i^M(\cdot)$ on the set $\{0\} \cup [k_i^*, \infty)$, similar to the capacity constrained demand function depicted on the right side of Figure 2. The algorithm computes its allocation $\mathbf{y}$ via $\mathbf{y} \in \arg\max_{\mathbf{z} \geq 0, \|\mathbf{z}\|_1 = \alpha D} \widetilde{P}(\mathbf{x}^0 + \mathbf{z})$, where $\mathbf{x}^0$ is the equilibrium allocation that the $(1-\alpha)D$ privatized drivers would choose under the profit function $\widetilde{P}(\cdot)$ in the absence of centrally controlled drivers. Finally, the centrally controlled drivers are allocated according to $\mathbf{y}$ and the privatized drivers choose an equilibrium response $\mathbf{x}^1$ according to the true profit function $P(\cdot)$. Unlike the NCF algorithm in Krichene et al. (2014), where the objective is similar to the linearized $\widetilde{P}(\cdot)$, the non-linearity of the profit function $P(\cdot)$ in our setting necessitates a final reallocation step of the privatized drivers, making the analysis substantially more involved.

Theorem 5 (Approximation Ratio for L-NCF)**.** *Consider a mixed public and informal transit system with a class $\mathcal{L}_V$ of demand functions in Equation* (3)*. Let $\mathbf{y}$ be the allocation of the centrally controlled drivers computed by L-NCF and $\mathbf{x}^1$ be the corresponding equilibrium response taken by privatized drivers. If $\mathbf{y}^*$ is a profit-maximizing Stackelberg strategy and $\mathbf{x}^*$ its equilibrium response, then $P(\mathbf{x}^1 + \mathbf{y}) \geq \frac{1}{1+\gamma} P(\mathbf{x}^* + \mathbf{y}^*)$, where $\gamma := \max_i \frac{S_i}{t_2 - t_1} \frac{\eta_E + \eta_L}{\eta_E \eta_L}$.*

Here γ is a measure of the non-linearity of the objective $P(\cdot)$, which is small when $S_i \ll t_2 - t_1$, as observed in our data for our numerical experiments in Section 7. Indeed, when $\gamma = 0$, which occurs when $S_i = 0$ as with the capacity-constrained demand shown on the right of Figure 2, L-NCF is optimal. See Section F for a proof of Theorem 5, which proceeds in three main steps. First, we upper bound the profit achievable by any allocation under the true profit function $P(\cdot)$ by $1 + \gamma$ times what that allocation would achieve under the approximate profit function $\widetilde{P}(\cdot)$. Second, we show that any allocation gets at least as much profit under $P(\cdot)$ as under $\widetilde{P}(\cdot)$. Third, the optimality of L-NCF for objectives with the form of $\widetilde{P}$ completes the proof.

Similar ideas can be applied to the rider welfare objective $R(\cdot)$, though the approximation factor achieved is $\frac{1}{1+\gamma} \frac{1}{1+\gamma \frac{p_{\min}}{p_{\max}}} \frac{\max_i l_i}{\min_i l_i}$ which is weaker due to misaligned objectives of the centrally controlled (rider welfare) and the privatized drivers (profit). See Appendix G for more details.

7 Numerical Experiments

This section presents numerical experiments based on a real-world case study of an informal transit system in Nalasopara, India, where shared auto-rickshaws serve nearly 100,000 riders daily. Our empirical results demonstrate that, although real-world inefficiencies in cumulative driver profit and rider welfare do not reach the worst-case levels given by our PoA bounds, the profit and welfare ratios under operational data from this system remain meaningfully bounded away from 1. We further extend this analysis to a Stackelberg routing setting, which demonstrates the efficacy of our LPF and L-NCF algorithms. Taken together, our numerical results underscore the importance of incorporating the incentives of informal and privatized operators into public transit planning. In the following, we describe the methodology to calibrate our model parameters based on data from Nalasopara's informal transit system in Section 7.1 and present our results in Section 7.2.

7.1 Experimental Setup and Model Calibration

In Nalasopara, shared auto-rickshaws serve riders on $n = 18$ routes (see Appendix D.1) connecting residential neighborhoods to a railway station. For each route i, we obtained trip times (l_i) and fares ($\bar{p}_i$) from a local NGO, and estimated the per-trip operating costs (c_i) using expenditure statistics, which indicate that auto expenditures constitute roughly 80% of drivers' daily earnings.

Accordingly, we set $c_i = 0.8F\bar{p}_i$, yielding a per-rider profit $p_i = \bar{p}_i - \frac{c_i}{F} = 0.2\bar{p}_i$, where $F = 4$ is the fixed capacity of the auto-rickshaws. We study this system during the evening peak period from 5 PM to midnight, when demand in Nalasopara is highest.[3] During this period, we estimate the total rider demand Λ_i for each route i using population data from the most recent Indian Census. For details of the demand calibration procedure, see Appendix D.2. The resulting estimates imply approximately 100,000 daily travelers across the eighteen routes, consistent with current estimates of local train ridership in Nalasopara after accounting for population growth since the last census (Mumbai Live Team 2025).

Finally, we calibrate the rider cost function parameters. We set the value of time to $\eta_T =$ Rs. 2.5 per minute, corresponding to an hourly wage of Rs. 150, reflecting average worker earnings in India in 2026 (ERI Economic Research Institute 2026). For schedule delay penalties, we follow the estimates in Small (1982) and set the earliness parameter to $\eta_E = 0.61$ and the lateness parameter $\eta_L = 2.4$. Finally, we model the cost of the outside option c_i^O as the time required to walk from the origin to the destination stop on route i.

7.2 Results

We first quantify inefficiencies from decentralized, selfish driver behavior in this real-world informal transit setting through driver profit and rider welfare ratios, i.e., the ratio of the optimal cumulative driver profit or rider welfare to that achieved under an equilibrium allocation under the calibrated model parameters. We then examine a Stackelberg routing setting in which a public authority controls a fraction of drivers, and compare our LPF and L-NCF algorithms to a status-quo *Greedy* baseline that optimizes the public authority's objective, either cumulative driver profit or rider welfare, for the centrally controlled drivers while ignoring the incentives of privatized operators.

Profit and Welfare Ratios: Figure 3 depicts the variation in the profit ratio, rider welfare ratio, and equilibrium per-driver profits as the number of drivers D in Nalasopara is varied from 100 to 2,000. We find that the profit ratio ranges between 1.1-1.2 and the rider welfare ratio ranges between 1.1-1.32 when $D \leq 1300$, a range consistent with practice.[4] As the number of drivers exceeds this threshold, both ratios converge to one as profit-maximizing and rider welfare maximizing allocations already absorb all rider demand; hence, additional drivers increase the corresponding objective under the equilibrium allocation without improving the optimal benchmarks.

While the empirically observed profit and rider welfare ratios do not reach the worst-case levels given by our PoA bounds, which equal 2 for profit and $1 + \frac{p_{\max}}{p_{\min}} = 4$ for rider welfare, where $\frac{p_{\max}}{p_{\min}} = 3$ in our data, they imply substantial inefficiencies in the practically relevant regime when $D \leq 1300$. Specifically, our observed profit and rider welfare ratios correspond to a 9–17% reduction in total profits and a 9–25% reduction in rider demand served. These losses translate into roughly 9,000–25,000 fewer riders served and average income losses of Rs. 22.5–42.5 per driver per day (about \$0.27–\$0.51), a substantial amount for daily-wage workers whose net income is around Rs. 250 (\$3) daily. These results highlight the significant value of cross-subsidization in mitigating the inefficiencies of selfish driver behavior in such high-stakes informal transit settings (see Section 5).

[3] Unlike typical evening peaks between 3-8 PM, Nalasopara exhibits a late-night demand surge, driven by long rail commutes as riders return home from major employment hubs and substantial nighttime commercial activity near the station.

[4] The number of drivers is not observed in the data and is likely to lie between 750-1250, as this corresponds to equilibrium per-driver profits of around Rs. 250 ($\approx$ \$3) in Figure 3 (right), consistent with average daily driver profits in Nalasopara.

Finally, note from the right of Figure 3 that the equilibrium profit per driver is monotonically decreasing in the number of drivers. This pattern is consistent with Corollary 2 and reflects the increasing competition for riders as driver supply grows. We further illustrate how the equilibrium allocation of drivers across routes evolves with the total driver supply D in Figure 6 in Appendix D.4.

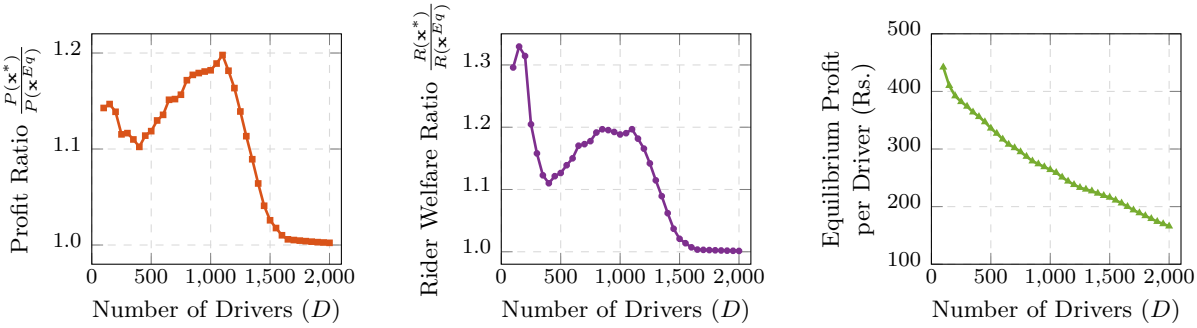


Figure 3: Depiction of the profit ratio (left), rider welfare ratio (center), and equilibrium profit per driver (right) as the number of drivers in the system is varied.

Stackelberg Routing: While the results in Figure 3 assume all drivers are profit maximizing, we now examine the Stackelberg routing setting where a fraction α of drivers is centrally controlled. Figure 4 compares the profit and rider welfare ratios of our LPF and N-LCF algorithms against a status-quo Greedy baseline as α ranges from 0 (fully privatized) to 1 (fully centralized), with intermediate values representing the coexistence of formal public transit and informal operators.

Across all three approaches, profit and rider welfare ratios generally decrease in α, which is natural as increased centralized control progressively aligns equilibrium outcomes with optimal allocations. However, when a public authority does not account for privatized operators' incentives, as with the status-quo Greedy baseline, no gains in cumulative driver profit (rider welfare) are realized until at least 70-80% (40-50%) of drivers are centrally controlled. In contrast, both LPF and L-NCF, which explicitly account for these incentives, deliver substantial improvements on both metrics at much lower levels of centralized control. For example, with only 30% centrally controlled drivers, L-NCF reduces the profit ratio from 1.18 to 1.1 and the welfare ratio from 1.19 to 1.09.

Moreover, LPF and L-NCF exhibit comparable performance across both metrics, with L-NCF slightly outperforming LPF on rider welfare for most values of α. Since L-NCF closely matches the performance of LPF, an adaptation of a benchmark Stackelberg routing algorithm (see Section 6.2), it highlights the efficacy of the linearization approximation of the rider demand function for our studied setting and points to the broader potential of such techniques in Stackelberg routing.

Overall, our numerical results highlight the significant losses associated with ignoring privatized operators and the sizable benefits of incorporating their incentives in public transit planning.

8 Conclusion and Future Work

This work developed a framework for analyzing the incentives in informal and privatized transit systems, and proposed mechanisms to guide public transit operation and incentive design when a substantial share of mobility is provided by such profit-driven private operators. We showed, through PoA bounds, that selfish driver behavior can result in bounded yet substantial losses in cumulative driver profit and rider welfare. However, these losses can be mitigated through targeted interventions, including cross-subsidization schemes and Stackelberg routing mechanisms. We further reinforced these findings through numerical experiments. Overall, our work highlights

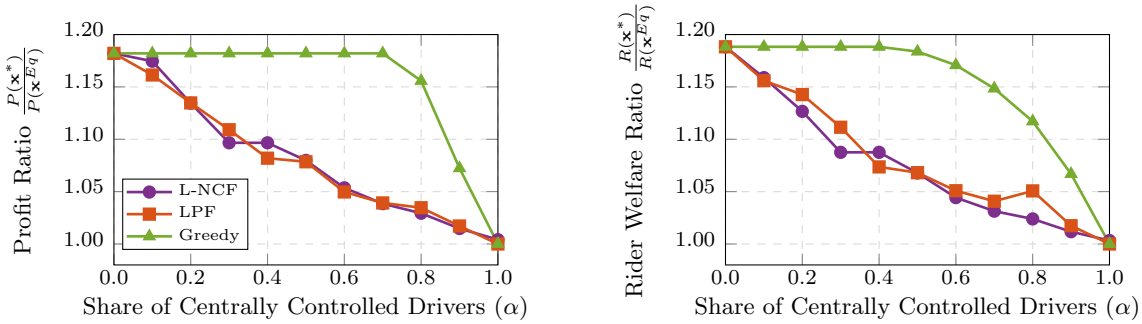


Figure 4: Profit ratio (left) and rider welfare ratio (right) as the share of centrally controlled drivers α is varied, with the remaining drivers operating as informal or privatized transit operators. We depict three Stackelberg routing algorithms: *L-NCF* and *LPF*, which account for the incentives of informal and privatized transit, and a *Greedy* baseline that does not, reflecting the status-quo and current practice among public transit agencies.

the importance of integrating informal and privatized operators into transit planning and provides practical mechanisms for aligning their incentives to achieve desired system outcomes. In doing so, we introduce a new, relatively understudied, application domain to the operations research and computer science communities.

Several directions merit future study. First, our model can be extended to incorporate heterogeneity in driver route preferences and *physical* road congestion, where travel times depend on the number of drivers on a route. For the rider-side problem in our framework, it would be valuable to explore standard relaxations of Vickrey's bottleneck model, such as heterogeneous values of time or non-uniform desired arrival time distributions. Finally, our Stackelberg routing mechanisms can be generalized to settings where centrally controlled drivers differ from privatized operators along additional dimensions beyond incentives, such as operating on a different set of routes, and to settings with driver coalitions, e.g., a ride-sharing platform operating autonomous shuttles.

Acknowledgments

This work was supported by the Data Science Institute Postdoctoral Fellowship at Columbia University. We also thank Dr. Jacqueline Klopp for several insightful discussions and the team at the Society for the Promotion of Area Resource Centers (SPARC), including Alex Mohan Kandathil, Anreerudha Paul, Shreya Baoni, and Kshitija Akre for their support in providing the data on Nalasopara's informal transit system. This data collection was enabled through the Mumbai Living Lab as part of PRISM (Partnership for Research on Informal and Shared Mobility).

References

Robert Cervero. *Informal transport in the developing world*. UN-HABITAT, 2000.

Robert Cervero and Aaron Golub. Informal transport: A global perspective. *Transport Policy*, 14(6):445–457, 2007. ISSN 0967-070X. doi: 10.1016/j.tranpol.2007.04.011. URL `https://www.sciencedirect.com/science/article/pii/S0967070X07000546`.

Lucas Conwell. Privatized provision of public transit. SSRN Scholarly Paper No. 4489658, revised April 26, 2025, 2025. URL `https://ssrn.com/abstract=4489658`.

Daniel Björkegren, Alice Duhaut, Geetika Nagpal, and Nick Tsivanidis. Public and private transit: Evidence from lagos. Technical report, National Bureau of Economic Research, 2025.

Eric Goldwyn. Anatomy of a new dollar van route: Informal transport and planning in new york city. *Journal of Transport Geography*, 88:102309, 2020. ISSN 0966-6923. URL https://www.sciencedirect.com/science/article/pii/S0966692317306282.

Uber. Jfk airport uber shuttle: our new low-cost shared-ride option. https://www.uber.com/global/en/r/airports/jfk/shuttle/, 2025. [Accessed: 2025-12-05].

Jacqueline M. Klopp and Clemence Cavoli. Mapping minibuses in maputo and nairobi: engaging paratransit in transportation planning in african cities. *Transport Reviews*, 39(5):657–676, 2019. ISSN 0144-1647. URL https://www.sciencedirect.com/science/article/pii/S0144164722001465.

William S. Vickrey. Congestion theory and transport investment. *The American Economic Review*, 59(2): 251–260, 1969. ISSN 00028282. URL http://www.jstor.org/stable/1823678.

Ravi Gadapalli. Role of intermediate public transport in indian cities. *Economic and Political Weekly*, 51 (9):46–49, 2016. ISSN 00129976, 23498846. URL http://www.jstor.org/stable/44004442.

Megha Kumar, Seema Singh, Akshima T. Ghate, Sarbojit Pal, and Sangeetha Ann Wilson. Informal public transport modes in india: A case study of five city regions. *IATSS Research*, 39(2):102–109, 2016. ISSN 0386-1112. doi: https://doi.org/10.1016/j.iatssr.2016.01.001. URL https://www.sciencedirect.com/science/article/pii/S0386111216000029.

Jacqueline M. Klopp. Chapter seven - from "para-transit" to transit? power, politics and popular transport. In Rafael H.M. Pereira and Geneviève Boisjoly, editors, *Social Issues in Transport Planning*, volume 8 of *Advances in Transport Policy and Planning*, pages 191–209. Academic Press, 2021. doi: https://doi.org/10.1016/bs.atpp.2021.07.002. URL https://www.sciencedirect.com/science/article/pii/S2543000921000317.

Nicola Rosaia. Competing platforms and transport equilibrium. *Econometrica*, 93(6):2235–2271, 2025. doi: https://doi.org/10.3982/ECTA21773. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA21773.

Giulia Brancaccio, Myrto Kalouptsidi, Theodore Papageorgiou, and Nicola Rosaia. Search frictions and efficiency in decentralized transport markets. *The Quarterly Journal of Economics*, 138(4):2451–2503, 2023.

Nicholas Buchholz. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. *Working paper, Tech. Rep.*, 2015.

Oluchi Mbonu and F Christopher Eaglin. Market segmentation and coordination costs: Evidence from johannesburg's minibus networks. Technical report, Working Paper, 2024.

Kush Mohan Mittal, Marc Timme, and Malte Schröder. Efficient self-organization of informal public transport networks. *Nature Communications*, 15(1):4910, 2024.

Dario Paccagnan, Rahul Chandan, and Jason R. Marden. Utility design for distributed resource allocation—part i: Characterizing and optimizing the exact price of anarchy. *IEEE Transactions on Automatic Control*, 65(11):4616–4631, 2020. doi: 10.1109/TAC.2019.2961995.

Weixuan Lin, John Z. F. Pang, Eilyan Bitar, and Adam Wierman. Networked cournot competition in platform markets: Access control and efficiency loss. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 4606–4611, 2017. doi: 10.1109/CDC.2017.8264340.

Tim Roughgarden. *Selfish routing and the price of anarchy*. MIT press, 2005.

Celeste Chavis and Carlos F. Daganzo. Analyzing the structure of informal transit: The evening commute problem. *Research in Transportation Economics*, 39(1):277–284, 2013. ISSN 0739-8859. doi: https://doi.org/10.1016/j.retrec.2012.06.025. URL https://www.sciencedirect.com/science/article/pii/S0739885912000972. THREDBO 12: Recent developments in the reform of land passenger transport.

Tawit Sangveraphunsiri, Michael J. Cassidy, and Carlos F. Daganzo. Jitney-lite: a flexible-route feeder service for developing countries. *Transportation Research Part B: Methodological*, 156:1–13, 2022. ISSN 0191-2615. doi: https://doi.org/10.1016/j.trb.2021.12.015. URL https://www.sciencedirect.com/science/article/pii/S0191261521002393.

Enrique Fernandez and Patrice Marcotte. Operators-users equilibrium model in a partially regulated transit system. *Transportation Science*, 26(2):93–105, 1992. doi: 10.1287/trsc.26.2.93. URL https://doi.org/10.1287/trsc.26.2.93.

Marvin Kraus and Yuichiro Yoshida. The commuter's time-of-use decision and optimal pricing and service in urban mass transit. *Journal of Urban Economics*, 51(1):170–195, 2002. ISSN 0094-1190. doi: https://doi.org/10.1006/juec.2001.2242. URL https://www.sciencedirect.com/science/article/pii/S0094119001922429.

Richard Arnott, André de Palma, and Robin Lindsey. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 83(1):161–179, 1993. ISSN 00028282. URL http://www.jstor.org/stable/2117502.

Eric J. Gonzales and Carlos F. Daganzo. Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. *Transportation Research Part B: Methodological*, 46(10):1519–1534, 2012. ISSN 0191-2615. doi: 10.1016/j.trb.2012.07.009. URL https://www.sciencedirect.com/science/article/pii/S0191261512001026.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, page 639, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334105. doi: 10.1145/2764468.2764527. URL https://doi.org/10.1145/2764468.2764527.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Dynamic pricing in ridesharing platforms. *SIGecom Exch.*, 15(1):65–70, September 2016. doi: 10.1145/2994501.2994505. URL https://doi.org/10.1145/2994501.2994505.

Ali Aouad and Ömer Sarıtaç. Dynamic stochastic matching under limited time. *Operations Research*, 70(4):2349–2383, 2022. doi: 10.1287/opre.2022.2293. URL https://doi.org/10.1287/opre.2022.2293.

Itai Ashlagi, Maximilien Burq, Chinmoy Dutta, Patrick Jaillet, Amin Saberi, and Chris Sholley. Edge weighted online windowed matching. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, page 729–742, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367929. doi: 10.1145/3328526.3329573. URL https://doi.org/10.1145/3328526.3329573.

Chiwei Yan, Julia Yan, and Yifan Shen. Pricing shared rides. In *Proceedings of the 25th ACM Conference on Economics and Computation*, EC '24, page 1, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400707049. doi: 10.1145/3670865.3673446. URL https://doi.org/10.1145/3670865.3673446.

Mustafa Dogan and Alexandre Jacquillat. On-demand service sharing via collective dynamic pricing. *Manufacturing & Service Operations Management*, 0(0):null, 0. doi: 10.1287/msom.2024.1301. URL https://doi.org/10.1287/msom.2024.1301.

Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467, 2017. doi: 10.1073/pnas.1611675114. URL https://www.pnas.org/doi/abs/10.1073/pnas.1611675114.

Yiding Feng, Rad Niazadeh, and Amin Saberi. Two-stage stochastic matching and pricing with applications to ride hailing. *Operations Research*, 72(4):1574–1594, 2024. doi: 10.1287/opre.2022.2398. URL https://doi.org/10.1287/opre.2022.2398.

Keji Wei, Vikrant Vaze, and Alexandre Jacquillat. Transit planning optimization under ride-hailing competition and traffic congestion. *Transportation Science*, 56(3):725–749, 2022. doi: 10.1287/trsc.2021.1068. URL https://doi.org/10.1287/trsc.2021.1068.

Gregory D Erhardt, Richard Alexander Mucci, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. Do transportation network companies increase or decrease transit ridership? empirical evidence from san francisco. *Transportation*, 49(2):313–342, 2022.

Zhenyu Yang and Nikolas Geroliminis. Regulating competition between transit and ride-hailing with transit priority zones. *Transportation Research Part C: Emerging Technologies*, 173:105016, 2025.

Andreas Neumann, Daniel Röder, and Johan W. Joubert. Toward a simulation of minibuses in south africa. *Journal of Transport and Land Use*, 8(1):137–154, 2015. ISSN 19387849. URL http://www.jstor.org/stable/26202706.

F Zwick. *Analysis and simulation of Santiago de Chile's colectivo system*. PhD thesis, Bachelor's thesis. TU Berlin, 2017.

Başar Özbilen. Integration of dolmuş as a paratransit mode to the existing public transport network: Ankara example. Master's thesis, Middle East Technical University, 2016.

Ilan Salomon and Lionel Silman. Scheduled bus and sherut taxi operation in israel. *Transportation Research Part B: Methodological*, 19(3):259–264, 1985. ISSN 0191-2615. doi: https://doi.org/10.1016/0191-2615(85)90009-8. URL https://www.sciencedirect.com/science/article/pii/0191261585900098.

Hai Yang, Hoi Yan Kong, and Qiang Meng. Value-of-time distributions and competitive bus services. *Transportation Research Part E: Logistics and Transportation Review*, 37(6):411–424, 2001. ISSN 1366-5545. doi: https://doi.org/10.1016/S1366-5545(01)00009-6. URL https://www.sciencedirect.com/science/article/pii/S1366554501000096.

Reuben Gronau. Optimum diversity in the public transport market. *Journal of Transport Economics and Policy*, 34(1):21–41, 2000. ISSN 00225258. URL http://www.jstor.org/stable/20053827.

Philip A. Viton. Privately-provided urban transport services: Entry deterrence and welfare. *Journal of Transport Economics and Policy*, 16(1):85–94, 1982. ISSN 00225258. URL http://www.jstor.org/stable/20052635.

Tim Roughgarden. Stackelberg scheduling strategies. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, STOC '01, page 104–113, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 1581133499. doi: 10.1145/380752.380783. URL https://doi.org/10.1145/380752.380783.

Walid Krichene, Jack D. Reilly, Saurabh Amin, and Alexandre M. Bayen. Stackelberg routing on parallel networks with horizontal queues. *IEEE Transactions on Automatic Control*, 59(3):714–727, 2014.

Kostas Bimpikis, Ozan Candogan, and Daniela Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.

John Glen Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3):325–362, 1952. doi: 10.1680/ipeds.1952.11259. URL https://doi.org/10.1680/ipeds.1952.11259.

Rajgopal Kannan, Costas Busch, and Paul Spirakis. The price of anarchy is unbounded for congestion games with superpolynomial latency costs. *arXiv preprint arXiv:1308.4101*, 2013.

Uber Blog. Introducing boost+, 2022. URL https://www.uber.com/en-US/blog/introducing-boost/. Accessed: 2026-02-04.

Mumbai Live Team. Mumbai local news: Nalasopara station gets elevated deck aimed to reduce congestion. https://www.mumbailive.com/en/infrastructure/mumbais-nalasopara-station-gets-elevated-deck-in-75-crore-overhaul-to-improve-travel-89404, 2025. URL https://www.mumbailive.com/en/infrastructure/mumbais-nalasopara-station-gets-elevated-deck-in-75-crore-overhaul-to-improve-travel-89404. Accessed: 2026-01-21.

ERI Economic Research Institute. General worker salary in India. https://www.erieri.com/salary/job/general-worker/india, 2026. Accessed: 2026-01-20.

Kenneth A. Small. The scheduling of consumer activities: Work trips. *American Economic Review*, 72(3): 467–479, 1982.

Zheng Zhu, Jintao Ke, and Hai Wang. A mean-field markov decision process model for spatial-temporal subsidies in ride-sourcing markets. *Transportation Research Part B: Methodological*, 150:540–565, 2021. ISSN 0191-2615. doi: https://doi.org/10.1016/j.trb.2021.06.014. URL https://www.sciencedirect.com/science/article/pii/S0191261521001259.

Yi-Neng Wang, Xi Lin, Fang He, Zhengtian Xu, and Zuo-Jun Max Shen. Efficiency and interventions of strategic driver relocation for ride-hailing platforms. *Available at SSRN 4454177*, 2023.

Devansh Jalota, Kiril Solovey, Karthik Gopalakrishnan, Stephen Zoepf, Hamsa Balakrishnan, and Marco Pavone. When efficiency meets equity in congestion pricing and revenue refunding schemes. *IEEE Transactions on Control of Network Systems*, 11(2):1127–1138, 2024. doi: 10.1109/TCNS.2023.3333413.

Anita Schöbel. Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510, 2012.

Gabriel Kreindler, Arya Gaduh, Tilman Graff, Rema Hanna, and Benjamin A Olken. Optimal public transportation networks: Evidence from the world's largest bus rapid transit system in jakarta. Technical report, National Bureau of Economic Research, 2023.

Siddhartha Banerjee, Chamsi Hssaine, Qi Luo, and Samitha Samaranayake. Plan your system and price for free: Fast algorithms for multimodal transit operations. *Transportation Science*, 59(1):13–27, 2025. doi: 10.1287/trsc.2022.0452. URL https://doi.org/10.1287/trsc.2022.0452.

Noémie Périvier, Chamsi Hssaine, Samitha Samaranayake, and Siddhartha Banerjee. Real-time approximate routing for smart transit systems. *SIGMETRICS Perform. Eval. Rev.*, 49(1):73–74, June 2022. ISSN 0163-5999. doi: 10.1145/3543516.3460096. URL https://doi.org/10.1145/3543516.3460096.

Pascal Van Hentenryck, Connor Riley, Anthony Trasatti, Hongzhao Guan, Tejas Santanam, Jorge A Huertas, Kevin Dalmeijer, Kari Watkins, Juwon Drake, and Samson Baskin. Marta reach: Piloting an on-demand multimodal transit system in atlanta. *arXiv preprint arXiv:2308.02681*, 2023.

Sophie Pavia, David Rogers, Amutheezan Sivagnanam, Michael Wilbur, Danushka Edirimanna, Youngseo Kim, Philip Pugliese, Samitha Samaranayake, Aron Laszka, Ayan Mukhopadhyay, and Abhishek Dubey. Deploying mobility-on-demand for all by optimizing paratransit services. In Kate Larson, editor, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 7430–7437. International Joint Conferences on Artificial Intelligence Organization, 8 2024. doi: 10.24963/ijcai.2024/822. URL https://doi.org/10.24963/ijcai.2024/822. AI for Good.

Kayla Cummings, Alexandre Jacquillat, and Vikrant Vaze. Activated benders decomposition for day-ahead paratransit itinerary planning. *INFORMS Journal on Computing*, 0(0):null, 2025. doi: 10.1287/ijoc.2023.0311. URL https://doi.org/10.1287/ijoc.2023.0311.

Yizhen Gu, Chang Jiang, Junfu Zhang, and Ben Zou. Subways and road congestion. *American Economic Journal: Applied Economics*, 13(2):83–115, April 2021. doi: 10.1257/app.20190024. URL https://www.aeaweb.org/articles?id=10.1257/app.20190024.

Nick Tsivanidis. Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. *Unpublished manuscript*, 18, 2022.

Roger Behrens, Saksith Chalermpong, and Daniel Oviedo. Informal paratransit in the global south. In *The Routledge Handbook of Public Transport*, pages 236–251. Routledge, 2021.

A Additional Discussion

A.1 Additional Related Work

This section surveys additional works in addition to those covered in Section 2 that are related to this work.

Beyond Stackelberg routing, we study cross-subsidy mechanisms that utilize route-specific tolls or subsidies to shape service patterns of informal or privatized transit drivers, akin to targeted driver subsidies in ride-sharing (Zhu et al. 2021, Wang et al. 2023). In contrast to these works, we demonstrate that a *budget-balanced* cross-subsidy mechanism, which requires no net government expenditure, can align operator incentives in the informal transit setting we study, echoing revenue-neutral congestion pricing with revenue-refunding schemes developed for traffic routing contexts (Jalota et al. 2024).

Our work also contributes to the literature on designing public transit systems, including network and line planning (Schöbel 2012, Kreindler et al. 2023), multi-modal integration (Banerjee et al. 2025, Périvier et al. 2022), and para-transit and micro-transit design (Van Hentenryck et al. 2023, Pavia et al. 2024, Cummings et al. 2025). However, unlike these works, which study optimal network design in centrally coordinated transit systems, we adopt a complementary approach by studying incentives and equilibrium outcomes in decentralized, privatized transit systems with a predefined menu of routes.

Finally, our work is related to studies on the economic impacts of public transportation infrastructure on outcomes such as congestion and welfare, which have mainly focused on formal transit systems, including subways (Gu et al. 2021) and BRT lines (Tsivanidis 2022). In contrast, our analysis centers on informal and privatized transit, highlighting how their incentives shape outcomes in terms of cumulative driver profits and rider welfare in settings where formal public transit is limited or inadequate.

A.2 Additional Discussion of Modeling Assumptions

This section provides additional discussion on our modeling assumptions. First, in modeling a minibus trip (left of Figure 1), we assume that after dropping off riders, drivers return empty to the origin to continue service on the same route, reflecting the operation of informal transit systems during morning or evening peak periods when demand is highly unidirectional (Björkegren et al. 2025). Moreover, while we assume a round-trip travel time of $2l_i$, our results extend naturally when the two legs of the trip have asymmetric travel times. Finally, consistent with observed practice, where over 96% of minibuses often operate at full capacity (Björkegren et al. 2025), we assume that all minibuses operate at capacity F.

Next, we abstract from physical road congestion and assume fixed travel times l_i independent of the number of drivers operating on a route, consistent with prior work (Björkegren et al. 2025, Conwell 2025). Rather than physical road congestion, our congestion-game captures *competition among drivers for riders*, reflecting how service over or under-provision affects rider and driver payoffs in informal transit systems.

Further, in line with the operational realities and institutional norms of informal transit systems, we assume that drivers commit to a single route to provide service for the entire horizon (e.g., the duration of the morning or evening peak) and do not switch routes between trips. We further show in Appendix C.1 that gains from route switching are limited at equilibria in our framework.

Finally, our assumption of a fixed menu of routes and fares reflects the operation of many informal and privatized transit systems, where routes and fares are often regulated (Behrens et al. 2021) and remain stable over short to medium horizons (e.g., several months), as in the Nalasopara system studied in Section 7. Given routes and fares as fixed primitives, our analysis focuses on the incentive and equilibrium effects in informal transit systems, where changes to route structures or fares are often infeasible or of limited relevance due to regulatory constraints. In this sense, our work serves as a natural starting point for studying higher-level planning decisions such as route or fare design, particularly when new informal transit systems are planned or old ones substantially redesigned.

Overall, while real-world informal and privatized transit systems are highly complex, our model abstracts from some operational details to isolate the core forces at play, most notably, driver competition for profitable routes and rider queuing delays, while remaining faithful to how these systems operate in practice. These abstractions enable a tractable framework that yields clear insights into the incentives and equilibrium behavior in informal transit systems, and extending the model to incorporate additional operational features is a valuable future research direction.

B Proofs

B.1 Proof of Proposition 1

B.1.1 Proof Sketch

Fix a route i with $S_i \geq 0$ and x_i drivers, resulting in a service rate of $\mu_i(x_i) = \frac{x_i F}{2l_i}$. Then, defining the time horizon over which drivers service the route as $T_i(x_i)$, the total mass of riders served is: $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i)$. Note $T_i(x_i) \geq \Delta = t_2 - t_1$, as riders may be willing to arrive earlier or later than their desired destination arrival times under the costs in Equation (1).

To establish an expression for $T_i(x_i)$, we first characterize the equilibrium of riders' mode choice (minibus versus outside option) and arrival-time decisions on a route i with x_i drivers by adapting the equilibrium characterization in Vickrey's bottleneck model with an outside option from Gonzales and Daganzo (2012) to our setting (see Proposition 5 in Appendix B.1). Combining this equilibrium characterization with the property that rider's desired destination arrival times are uniformly distributed, we derive a closed-form expression for the total service time. In particular, when the system is under-supplied (i.e., $\mu_i(x_i) < \lambda_i$), $T_i(x_i) = \Delta + \min\{S_i, \bar{S}(x_i)\}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\left(1 - \frac{\Delta x_i F}{2l_i\Lambda_i}\right)$, where $\bar{S}(x_i) = \frac{\Lambda_i\eta_E\eta_L}{\mu_i(x_i)(\eta_E+\eta_L)}$ (see Corollary 4 in Appendix B.1). When the system is over-supplied (i.e., $\mu_i(x_i) \geq \lambda_i$), $T_i(x_i) = \Delta$.

Finally, we evaluate the rider demand $\Lambda_i^M(x_i) = \mu_i(x_i)T_i(x_i)$ by substituting the derived service time relation. When the system is over-supplied (i.e., $\mu_i(x_i) \geq \lambda_i$) or $S_i \geq \bar{S}(x_i)$, we get $\Lambda_i^M(x_i) = \Lambda_i$ implying all rider demand is served when $x_i \geq \tilde{k}_i^*$. For $x_i < \tilde{k}_i^*$, substituting the service-time expression into $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i) = \mu_i(x_i)T_i(x_i)$ yields the concave quadratic form.

B.1.2 Complete Proof

Fix a route i with $S_i \geq 0$ and x_i minibus drivers operating on that route, resulting in a service rate of $\mu_i(x_i) = \frac{x_i F}{2l_i}$. Then, defining the total time horizon over which the drivers service the route as $T_i(x_i)$, the total mass of riders served by the minibus is given by: $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i)$.

Note that $T_i(x_i) \geq \Delta = t_2 - t_1$, as riders may be willing to arrive earlier or later than their desired arrival times at the destination under the cost functions in Equation (1). Thus, the key to proving our result is to establish a relation for the total time $T_i(x_i)$ over which riders are serviced on a route i, given an allocation of x_i minibus drivers on that route.

Note that if the system is over-supplied with $\mu_i(x_i) \geq \lambda_i$, since $S_i \geq 0$, there will be no rider queuing or waiting delays and all riders can arrive at their destinations at their desired times by the minibus; hence, $T_i(x_i) = \Delta$ in this regime when $x_i \geq \frac{2l_i\lambda_i}{F}$.

Thus, consider the under-supplied regime when $\mu_i(x_i) < \lambda_i$. In this regime, to derive a relation for the total service time, we first characterize riders' equilibrium mode and arrival time decisions on a given route i with x_i drivers operating on that route. We do so by adapting the associated equilibrium characterization in Vickrey's bottleneck model with an outside option from Gonzales and Daganzo (2012) to our informal transit setting. In particular, defining the threshold $\bar{S}(x_i) = \frac{\Lambda_i \eta_E \eta_L}{\mu_i(x_i)(\eta_E+\eta_L)}$ given an allocation of x_i drivers to route i, we characterize riders' equilibrium mode and arrival time decisions in two cases: (i) $S_i \geq \bar{S}(x_i)$, and (ii) $S_i \in [0, \bar{S}(x_i))$.

Case (i): In this setting, the outside option is never cost-effective relative to traveling by the minibus, even for the user experiencing the highest travel cost with a wait time of $\bar{S}(x_i)$. The resulting equilibrium waiting time profile of drivers reduces to the classical Vickrey bottleneck model without an outside option and is depicted on the left of Figure 5, with riders being serviced over the interval $[t'_A, t'_B]$. In this case, since all riders are serviced by the minibus at equilibrium, the total service time $T_i(x_i) = t'_B - t'_A = \frac{\Lambda_i}{\mu_i(x_i)}$.

Case (ii): In this regime, only a fraction of the total rider demand is served by the minibus. Adapting Proposition 1 from Gonzales and Daganzo (2012) to our informal transit setting, we obtain the following equilibrium characterization in the regime when $S_i \in [0, \bar{S}(x_i))$.

Proposition 5 (Rider Equilibrium Characterization (Gonzales and Daganzo 2012))**.** *Suppose x_i minibus drivers operate on route i with $\mu_i(x_i) < \lambda_i$, where users choose between the minibus and an outside option where the cost difference between the outside option and the minibus without rider queuing delays $S_i \in [0, \bar{S}(x_i))$. Then, assuming riders are serviced in the order of their desired destination arrival times, there exists a unique rider equilibrium satisfying (see right of Figure 5):*

- *The number of minibus riders that arrive at their destination earlier than their desired time is given by $N_E = \frac{\mu_i(x_i)S_i}{\eta_E}$ and they travel at the beginning of the rush between periods $[t_A, t_B]$.*
- *The number of minibus riders that arrive at their destination later than their desired time is given by $N_L = \frac{\mu_i(x_i)S_i}{\eta_L}$ and they travel at the end of the rush between periods $[t_C, t_D]$.*
- *The number of users that arrive exactly on time by the minibus and those that use the outside option are strictly decreasing functions of the cost difference S_i and they travel in the middle of the rush between $[t_B, t_C]$.*

While Gonzales and Daganzo (2012) characterize the above equilibrium for a broad class of desired destination arrival time distributions for riders, they do not provide a closed-form characterization of the total service time $T_i(x_i) = t_D - t_A$. In our setting, by focusing on uniformly distributed desired destination arrival time distributions, which is consistent with the classical bottleneck model of Vickrey (1969) while also significantly more general than the single departure time formulations most commonly studied in the literature, we obtain a closed-form characterization of the service times in our setting.

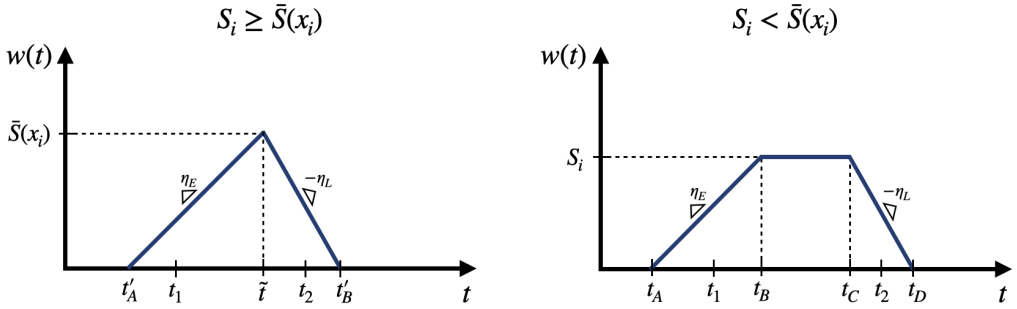


Figure 5: Depiction of the equilibrium rider waiting time profiles in the regime when the cost difference between the outside option and that of using the minibus without queuing delays satisfies $S_i \geq \bar{S}(x_i)$ (left) and $S_i < \bar{S}(x_i)$ (right).

Corollary 4 (Equilibrium Service Time)**.** *Suppose x_i minibus drivers operate on route i with $\mu_i(x_i) < \lambda_i$. Then, in the regime when the cost difference between the outside option and the minibus without rider queuing delays is non-negative, i.e., $S_i \geq 0$, the total service time $T_i(x_i) = \Delta + \min\{S_i, \bar{S}(x_i)\} \left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right) \left(1 - \frac{\Delta x_i F}{2l_i\Lambda_i}\right)$.*

Proof. First consider the regime when $S_i \geq \bar{S}(x_i)$. Then, substituting the expression for $\bar{S}(x_i)$ in the expression for the total service time in the statement of the corollary and simplifying, we obtain that $T_i(x_i) = \frac{\Lambda_i}{\mu_i(x_i)}$, consistent with the expression in the analysis of case (i) above.

Next, consider the regime when $S_i < \bar{S}(x_i)$. In this case, from the right of Figure 5, note that the service time $T_i(x_i) = t_D - t_A = (t_D - t_C) + (t_C - t_B) + (t_B - t_A)$. Next, from Proposition 5, since all riders are served by the minibus in the period $[t_A, t_B]$, it follows that $N_E = \frac{\mu_i(x_i)S_i}{\eta_E} = \mu_i(x_i)(t_B - t_A)$, which implies $t_B - t_A = \frac{S_i}{\eta_E}$. Analogously, it follows that $t_D - t_C = \frac{S_i}{\eta_L}$.

Finally, to derive the relation for $t_C - t_B$, note that the total fraction of the demand Λ_i that is served early or late during the periods $[t_A, t_B]$ and $[t_C, t_D]$ is $\frac{S_i}{\bar{S}(x_i)}$ by the linearity of the equilibrium waiting time function in Figure 5. Hence, it follows that $t_C - t_B = \left(1 - \frac{S_i}{\bar{S}(x_i)}\right)\Delta$. Then:

$$\begin{aligned}
T_i(x_i) &= t_D - t_A = (t_D - t_C) + (t_C - t_B) + (t_B - t_A), \\
&= \frac{S_i}{\eta_L} + \frac{S_i}{\eta_E} + (t_2 - t_1)(1 - \frac{S_i}{\bar{S}(x_i)}), \\
&= \Delta + S_i \left(\frac{1}{\eta_L} + \frac{1}{\eta_E} - \frac{\Delta}{\bar{S}(x_i)}\right), \\
&= \Delta + S_i \left(\frac{\eta_E + \eta_L}{\eta_E\eta_L} - \frac{t_2 - t_1}{\frac{\Lambda_i\eta_E\eta_L}{\mu_i(x_i)(\eta_E+\eta_L)}}\right), \\
&= \Delta + S_i \left(\frac{\eta_E + \eta_L}{\eta_E\eta_L}\right)\left(1 - \frac{\Delta x_i F}{2l_i\Lambda_i}\right),
\end{aligned}$$

where we substitute $\mu_i(x_i) = \frac{x_i F}{2l_i}$ to obtain the final equality. This establishes our desired relation on the total service time. □

Having established the relations for the total service time in both the under and over-supplied

regimes, we now evaluate the minibus rider demand $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i)$ by substituting the derived expressions for the service time.

First consider the over-supplied regime when $\mu_i(x_i) \geq \lambda_i$, corresponding to the setting when $x_i \geq \frac{2l_i\lambda_i}{F}$. In this case, recalling that $T_i(x_i) = \Delta$, we have $\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i) = \lambda_i\Delta = \Lambda_i$.

Next, in the under-supplied case, suppose that it further holds that $S_i \geq \bar{S}(x_i)$, corresponding to the setting when $x_i \geq \frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}$. Then, substituting the expression for $\bar{S}(x_i)$ in the total service time expression in Corollary 4, it follows that $T_i(x_i) = \frac{\Lambda_i}{\mu_i(x_i)}$. Hence, in this regime, it again follows that $\Lambda_i^M(x_i) = \Lambda_i$. Together, the above relations imply that for any driver allocation $x_i \geq \min\{\frac{2l_i\lambda_i}{F}, \frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}\} = \tilde{k}_i^*$, all the rider demand is served by the minibus, i.e., $\Lambda_i^M(x_i) = \Lambda_i$ for all $x_i \geq \tilde{k}_i^*$.

Finally, consider the regime when $x_i < \tilde{k}_i^*$. In this case, we leverage the relation for the total service time derived in Corollary 4 to derive the minibus rider demand as a function of the driver allocation on that route. In this regime, note that:

$$\begin{aligned}\Lambda_i^M(x_i) = \min\{\mu_i(x_i), \lambda_i\}T_i(x_i) &\overset{(a)}{=} \mu_i(x_i)T_i(x_i),\\ &\overset{(b)}{=} \frac{x_iF}{2l_i}\left(\Delta + S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\left(1 - \frac{\Delta x_iF}{2l_i\Lambda_i}\right)\right),\\ &= \frac{F}{2l_i}\left(\Delta + S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)x_i - \left(\left(\frac{F}{2l_i}\right)^2 S_i\frac{\eta_E+\eta_L}{\eta_E\eta_L}\frac{\Delta}{\Lambda_i}\right)x_i^2\end{aligned}$$

where (a) follows as $\mu_i(x_i) < \lambda_i$ in the regime when $x_i < \tilde{k}_i^*$ and (b) follows by substituting the relation for $\mu_i(x_i)$ and that for $T_i(x_i)$ derived in Corollary 4. This establishes our desired relation for the minibus rider demand in the regime when $x_i < \tilde{k}_i^*$, which proves our result.

B.2 Proof of Corollary 1

To establish continuity of the minibus rider demand function, it suffices to show that $\Lambda_i^M(\tilde{k}_i^*) = \Lambda_i$. This follows directly by substituting the two candidate expressions for $\tilde{k}_i^*$, i.e., $\frac{2l_i\lambda_i}{F}$ and $\frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}$, into the quadratic expression defining $\Lambda_i^M(\cdot)$.

Next, to establish monotonicity, it suffices to show that $\Lambda_i^M(\cdot)$ is monotonically non-decreasing in the range $x_i \in [0, \tilde{k}_i^*]$. To see this, we consider the regime when the derivative of the minibus rider demand function is non-negative:

$$\frac{d\Lambda_i^M(x_i)}{dx_i} = \frac{F}{2l_i}\left(t_2 - t_1 + S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right) - 2\left(\left(\frac{F}{2l_i}\right)^2 S_i\frac{\eta_E+\eta_L}{\eta_E\eta_L}\frac{t_2-t_1}{\Lambda_i}\right) \geq 0.$$

Rearranging the above inequality and simplifying, we obtain the following condition on the driver allocation x_i for the non-negativity of the derivative of the minibus demand function $\Lambda_i^M(\cdot)$:

$$x_i \leq \frac{2l_i\lambda_i}{F} + \frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}.$$

Since the expression on the right hand side of the above inequality is at least $\tilde{k}_i^* = \min\left\{\frac{2l_i\lambda_i}{F}, \frac{2l_i\Lambda_i\eta_E\eta_L}{FS_i(\eta_E+\eta_L)}\right\}$, it follows that $\Lambda_i^M(\cdot)$ is monotonically non-decreasing in the range $x_i \in [0, \tilde{k}_i^*]$.

B.3 Proof of Corollary 2

In the regime when $x_i > \tilde{k}_i^*$, the per-driver profit $\pi_i(x_i) = \frac{\Lambda_i p_i}{x_i}$. On the other hand, when $x_i \leq \tilde{k}_i^*$, substituting the relation for the demand function $\Lambda_i^M(\cdot)$, the per-driver profit is given by:

$$\pi_i(x_i) = \frac{\Lambda_i^M(x_i)p_i}{x_i} = p_i \frac{F}{2l_i}\left(t_2 - t_1 + S_i\left(\frac{\eta_E + \eta_L}{\eta_E \eta_L}\right)\right) - p_i\left(\left(\frac{F}{2l_i}\right)^2 S_i \frac{\eta_E + \eta_L}{\eta_E \eta_L}\frac{t_2 - t_1}{\Lambda_i}\right)x_i. \quad (5)$$

From these relations and the continuity of the rider demand function $\Lambda_i^M(\cdot)$, the per-driver profit function is continuous and monotonically decreasing in the driver allocation x_i.

B.4 Proof of Theorem 2

Fix an informal transit instance I with associated rider demand functions belonging to the class $\mathcal{L}_V$, given by Equation (3) for all routes $i \in [n]$. Since our proof applies to all feasible instances I, we drop the dependency on I in our notation for the remainder of this proof.

For this instance, let $\mathbf{x}^*$ be the cumulative rider welfare allocation, i.e., $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\Omega} R(\mathbf{x}) = \sum_{i=1}^n \Lambda_i^M(x_i)$, and let $\mathbf{x}^{Eq}$ be any equilibrium driver allocation. Moreover, as in the proof of Theorem 1, it suffices to restrict attention to the set of routes Q for which the difference $S_i \geq 0$.

Then, defining the two sets: $L_1 = \{i \in Q : x_i^{Eq} \geq x_i^*\}$ and $L_2 = \{i \in Q : x_i^{Eq} < x_i^*\}$, we have:

$$\begin{aligned}
R(\mathbf{x}^*) &= \sum_{i\in Q}\Lambda_i^M(x_i^*) = \sum_{i\in L_1}\Lambda_i^M(x_i^*) + \sum_{i\in L_2}\Lambda_i^M(x_i^*) \overset{(a)}{\leq} \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \sum_{i\in L_2}\Lambda_i^M(x_i^*),\\
&\overset{(b)}{\leq} \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{1}{p_{\min}}\sum_{i\in L_2} p_i\Lambda_i^M(x_i^*) \overset{(c)}{=} \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{1}{p_{\min}}\sum_{i\in L_2}\pi_i^* x_i^*,\\
&\overset{(d)}{\leq} \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{1}{p_{\min}}\pi^{Eq}\sum_{i\in L_2} x_i^* \overset{(e)}{\leq} \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{1}{p_{\min}}\pi^{Eq}\sum_{i\in Q} x_i^{Eq},\\
&= \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{1}{p_{\min}}\sum_{i\in Q} p_i\Lambda_i^M(x_i^{Eq}) \leq \sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) + \frac{p_{\max}}{p_{\min}}\sum_{i\in Q}\Lambda_i^M(x_i^{Eq}),\\
&\overset{(f)}{\leq} \left(1 + \frac{p_{\max}}{p_{\min}}\right) R(\mathbf{x}^{Eq}),
\end{aligned}$$

where (a) follows by the monotonicity of the minibus rider demand in the driver allocation as established in Corollary 1, (b) follows as $p_{\min} \leq p_i$ for all routes i, and (c) follows by the definition of $\pi_i^* = \frac{p_i\Lambda_i^M(x_i^*)}{x_i^*}$. Moreover, (d) follows as $\pi^{Eq} \geq \pi_i^*$ for all $i \in L_2$ by the monotonicity of the per-driver profit function in Corollary 2, (e) follows as $\sum_{i\in L_2} x_i^* \leq \sum_{i\in Q} x_i^* = \sum_{i\in Q} x_i^{Eq} = D$, as the same number of drivers are allocated under the optimal and equilibrium allocations, and (f) follows as $\sum_{i\in L_1}\Lambda_i^M(x_i^{Eq}) \leq \sum_{i\in Q}\Lambda_i^M(x_i^{Eq}) = R(\mathbf{x}^{Eq})$, as $L_1 \subseteq Q$. The above analysis holds for any equilibrium allocation $\mathbf{x}^{Eq}$ and instance of our informal transit system, thus establishing our claim that the R-PoA is at most $1 + \frac{p_{\max}}{p_{\min}}$.

B.5 Proof of Proposition 3

We proceed in the vein of Proposition 2, and consider an informal transit instance with $n = 2$ routes, a normalized demand $D = 1$, and the cost difference $S_i = 0$ for both routes. Consequently, we

obtain piece-wise linear rider demand functions for both routes, as in Equation (4) with $\tilde{k}_i^* = \frac{2l_i\lambda_i}{F}$ for $i \in \{1, 2\}$. Moreover, consider the setting where $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$. In this case, the rider welfare maximizing allocation is $\mathbf{x}^* = (\tilde{k}_1^*, \tilde{k}_2^*)$, resulting in a total maximum achievable rider demand served of $R(\mathbf{x}^*) = \Lambda_1^M(\tilde{k}_1^*) + \Lambda_2^M(\tilde{k}_2^*) = \Lambda_1 + \Lambda_2$.

Next, suppose that profit per driver on the two routes satisfies $\pi_1(1) = \pi_2(0)$, i.e., the informal transit system supports an equilibrium driver allocation $\mathbf{x}^{Eq} = (1, 0)$. For this condition to hold, note that the following equalities must be satisfied:

$$p_1\Lambda_1 = \pi_1(1) = \pi_2(0) = p_2\frac{F(t_2 - t_1)}{2l_2} = p_2\Lambda_2\frac{F}{2l_2\lambda_2} = p_2\Lambda_2\frac{1}{\tilde{k}_2^*}$$

Using this relation, minibus rider demand ratio between the optimal and above-defined equilibrium allocation is given by:

$$\frac{R(\mathbf{x}^*)}{R(\mathbf{x}^{Eq})} = \frac{\Lambda_1 + \Lambda_2}{\Lambda_1} = 1 + \frac{\Lambda_2}{\Lambda_1} = 1 + \tilde{k}_2^*\frac{p_1}{p_2}.$$

Finally, taking the limit as $\tilde{k}_2^* \to 1$ and $\tilde{k}_1^* \to 0$ while satisfying $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$, the above analysis implies that the minibus rider demand ratio approaches $1 + \frac{p_{\max}}{p_{\min}}$, thus establishing our claim.

B.6 Proof of Theorem 3

For brevity of notation, we define the per-driver profit on route i (without cross-subsidies) at the target allocation $\mathbf{x}^*$ as $\pi_i^* = \pi_i(x_i^*) = \frac{\Lambda_i^M(x_i^*)p_i}{x_i^*}$, where the rider demand function is given by Equation (3). Then, to induce $\mathbf{x}^*$ as an equilibrium driver allocation under a cross-subsidy scheme defined by $\boldsymbol{\tau}$, the following equilibrium condition must hold: for any route i with $x_i^* > 0$, the per-driver profits under cross-subsidy transfers is the at least that of any other route j, i.e.,

$$\tilde{\pi}_i(x_i^*) \geq \tilde{\pi}_j(x_i^*) \quad \implies \quad \pi_i^* + \tau_i \geq \pi_j^* + \tau_j, \tag{6}$$

where the inequality is met with an equality for all routes j with $x_j^* > 0$.

In the remainder of this proof, we construct a transfer vector $\boldsymbol{\tau}$ that satisfies the above equilibrium condition with equality for all routes, i.e., $\pi_i^* + \tau_i = \pi_j^* + \tau_j$ for all routes i, j, while satisfying budget balance, i.e., $\sum_{i\in[n]} \tau_i x_i^* = 0$. Thus, we seek to satisfy the following n (unique) equations with n unknowns, corresponding to the entries of the transfer vector $\boldsymbol{\tau}$:

$$\begin{aligned}
\pi_1^* + \tau_1 &= \pi_2^* + \tau_2, \\
\pi_1^* + \tau_1 &= \pi_3^* + \tau_3, \\
&\vdots \\
\pi_1^* + \tau_1 &= \pi_n^* + \tau_n, \\
\sum_{i=1}^{n} \tau_i x_i^* &= 0.
\end{aligned}$$

Rearranging the above equations, our goal is to find a vector $\boldsymbol{\tau}$ that solves the following system:

$$\begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ & & & \vdots & \\ 1 & 0 & 0 & \dots & -1 \\ x_1^* & x_2^* & x_3^* & \dots & x_n^* \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \vdots \\ \tau_n \end{bmatrix} = \begin{bmatrix} \pi_2^* - \pi_1^* \\ \pi_3^* - \pi_1^* \\ \vdots \\ \pi_n^* - \pi_1^* \\ 0 \end{bmatrix} \tag{7}$$

Note that the above matrix is full-rank and thus admits a unique solution if there is any route j such that $x_j^* > 0$. Then, note from the budget-balance constraint that:

$$\begin{aligned} 0 = \sum_{i=1}^{n} \tau_i x_i^* &= \sum_{i \in [n]: x_i^* > 0} \tau_i x_i^* = \tau_j x_j^* + \sum_{i \in [n]: x_i^* > 0, i \neq j} \tau_i x_i^* \overset{(a)}{=} \tau_j x_j^* + \sum_{i \in [n]: x_i^* > 0, i \neq j} (\pi_j^* + \tau_j - \pi_i^*) x_i^* \\ &= \sum_{i \in [n]: x_i^* > 0, i \neq j} (\pi_j^* - \pi_i^*) x_i^* + \tau_j \sum_{i \in [n]: x_i^* > 0} x_i^*, \end{aligned}$$

where (a) follows from the equation $\pi_i^* + \tau_i = \pi_j^* + \tau_j$ for all pairs of routes i, j. Rearranging the above expression, we obtain $\tau_j = \frac{\sum_{i \in [n]: x_i^* > 0, i \neq j} (\pi_i^* - \pi_j^*) x_i^*}{\sum_{i \in [n]: x_i^* > 0} x_i^*}$. Next, for the remaining $i' \neq j$, we have:

$$\tau_{i'} = \pi_j^* - \pi_{i'}^* + \tau_j = \pi_j^* - \pi_{i'}^* + \frac{\sum_{i \in [n]: x_i^* > 0, i \neq j} (\pi_i^* - \pi_j^*) x_i^*}{\sum_{i \in [n]: x_i^* > 0} x_i^*}.$$

The resulting per-driver profit under these tolls, which is fixed across routes is given by:

$$\tilde{\pi}^{Eq} = \tilde{\pi}_j(x_j^*) = \pi_j^* + \tau_j = \pi_j^* + \frac{\sum_{i \in [n]: x_i^* > 0, i \neq j} (\pi_i^* - \pi_j^*) x_i^*}{\sum_{i \in [n]: x_i^* > 0} x_i^*} = \frac{\sum_{i \in [n]: x_i^* > 0} \pi_i^* x_i^*}{\sum_{i \in [n]: x_i^* > 0} x_i^*} = \frac{\sum_{i \in [n]} \pi_i^* x_i^*}{\sum_{i \in [n]} x_i^*},$$

which establishes our claim.

B.7 Proof of Proposition 4

Following Roughgarden (2001), we reduce from an instance of $\frac{2}{3} - \frac{1}{3}$ partition, which consists of a sequence of n positive integers $a_1, \dots, a_n$ with $\sum_{i \in [n]} a_i = A$, and involves the task of deciding there there is a subset Q of these integers satisfying $\sum_{i \in Q} a_i = \frac{A}{3}$.

Construction of Stackelberg Instance: We now construct a Stackelberg routing instance for our studied mixed public and informal transit system with $n+1$ routes, a total of $D = 2A$ drivers, with $\alpha = 0.25$, corresponding to $\tilde{\alpha} = \alpha D = \frac{A}{2}$ centrally controlled drivers. In our constructed instance, we let the per-rider profit $p_i = p = \frac{4A^2}{\min_i a_i}$ be the same for all routes, in which case the cumulative driver profit maximization objective reduces to the rider welfare maximization objective. Thus, our reduction will apply to both objectives, and we use the term profit to refer to the objective of our Stackelberg routing instance for the remainder of this proof.

Furthermore, in our instance, we let $l_i = l$ and $S_i = S$ be the same across all routes $i \in [n]$, and let $\frac{F(t_2 - t_1)}{2l} = \frac{\min_i a_i}{2A}$ and $\frac{FS(\eta_E + \eta_L)}{2l\eta_E\eta_L} = \frac{\min_i a_i}{2A}$. Moreover, let $\Lambda_i = a_i \min_i a_i$ for all $i \in [n]$. Then, it follows for all routes $i \in [n]$ that:

$$\tilde{k}_i^* = \min \left\{ \frac{2l\Lambda_i}{F(t_2 - t_1)}, \frac{2l\Lambda_i \eta_E \eta_L}{FS(\eta_E + \eta_L)} \right\} = 2Aa_i \geq 2A,$$

which implies that $\tilde{k}_i^* \geq 2A = D$ for all routes $i \in [n]$, i.e., any feasible driver allocation always lies in the quadratic portion of the rider demand function for each of the routes.

Next, we obtain the following relation for the profit function for the first n routes when $x_i \leq \tilde{k}_i^*$:

$$p\Lambda_i^M(x_i) = p\frac{F}{2l}\left(t_2 - t_1 + S\left(\frac{\eta_E + \eta_L}{\eta_E \eta_L}\right)\right)x_i - p\left(\left(\frac{F}{2l}\right)^2 S\frac{\eta_E + \eta_L}{\eta_E \eta_L}\frac{t_2 - t_1}{\Lambda_i}\right)x_i^2 = -\frac{x_i^2}{a_i} + 4Ax_i, \forall i \in [n].$$

From the above relation, in our instance, we have $p\Lambda_i^M(x_i) = -\frac{x_i^2}{a_i} + 4Ax_i$ for all $i \in [n]$ when $x_i \leq \tilde{k}_i^*$. Moreover, for the $(n+1)$'th route, we set $p_{n+1}\Lambda_{n+1}^M(x_{n+1}) = (4A+4)x_{n+1} - \frac{3x_{n+1}^2}{A}$ and let $\tilde{k}_{n+1}^* \geq 2A$, which can be straightforwardly set as the problem parameters here need not depend on a_i, as was required for the first n routes. We now claim that we have a "yes" instance of the $\frac{2}{3} - \frac{1}{3}$ partition problem if and only if there is a Stackelberg strategy in this defined mixed public and informal transit instance with profit at least $A(8A - \frac{3}{4})$.

($\implies$): Suppose we have a "yes" instance of partition, i.e., there is some subset Q with $\sum_{i \in Q} a_i = \frac{2}{3}A$. Then, consider a centrally controlled driver strategy $\mathbf{x}^C$ with $x_i^C = \frac{3}{4}a_i$ for $i \in Q$ and $x_i^C = 0$ otherwise. Since $\sum_{i \in Q} x_i^C = \frac{3}{4}\sum_{i \in Q} a_i = \frac{1}{2}A$, this is a feasible Stackelberg strategy. Note that the resulting equilibrium is such that $x_i^P = 0$ for all $i \in Q$, resulting in a per-driver profit of $4A - \frac{3}{4}$ on those routes, $x_i^P = \frac{a_i}{4}$ for all $i \in [n] \backslash Q$, resulting in a per-driver profit of $4A - \frac{1}{4}$ on those routes, and $x_{n+1}^P = \frac{17}{12}A$, resulting in a per-driver profit of $4A + 4 - \frac{3 \times 17A}{12A} = 4A - \frac{1}{4}$ on route $n+1$. Note here that $\sum_{i=1}^{n+1} x_i^P = \sum_{i \in [n] \backslash Q} \frac{a_i}{4} + \frac{17A}{12} = \frac{1}{4}\frac{A}{3} + \frac{17A}{12} = \frac{3A}{2}$, i.e., the resulting solution is indeed feasible. Moreover, the total profit corresponding to this solution is given by: $\frac{A}{2}(4A - \frac{3}{4}) + \frac{3A}{2}(4A - \frac{1}{4}) = 8A^2 - \frac{3A}{4} = A(8A - \frac{3}{4})$, establishing the forward direction of our claim.

($\impliedby$): Now, suppose we have a "no" instance of partition. Consider a Stackelberg strategy $\mathbf{x}^C$ with a resulting induced equilibrium for the privatized drivers $\mathbf{x}^P$. We seek to show that the total profit of this outcome is strictly smaller than $A(8A - \frac{3}{4})$. To do so, first define Q as the set of routes with $x_i^P = 0$. We first note that route $\{n+1\} \notin Q$, as even if all centrally controlled drivers are allocated on this route, some of the privatized drivers will use that route at equilibrium. Moreover, all routes with $x_i^P > 0$ have the same profits per driver at equilibrium. Further, all routes i, j with $x_i^P = x_j^P = 0$ have the same profits per driver under the optimal Stackelberg assignment (see Roughgarden (2001) for more details), which follows as the marginal profit of route i $(4A - \frac{2x_i}{a_i})$ is at least the marginal profit of route j $(4A - \frac{2x_j}{a_j})$ if and only if the profit per driver on route i $(4A - \frac{x_i}{a_i})$ is at least the profit per driver of route j $(4A - \frac{x_j}{a_j})$. Thus, all routes have one of two possible profits per driver.

Next, note that if $Q = \emptyset$, then we have an assignment with $x_i = \frac{a_i}{2}$ for all $i \in [n]$ and $x_{n+1} = \frac{3A}{2}$ must be the equilibrium with a profit per driver on each route given by $4A - \frac{1}{2}$, and a total profit of $2A(4A - \frac{1}{2}) = A(8A - 1) < A(8A - \frac{3}{4})$. Thus, we now consider the setting when Q is non-empty. In this case, define $\beta \in (0, 1]$ such that $\sum_{i \in Q} a_i = \beta A$. Moreover, let $\eta \in (0, \frac{1}{2}]$ be such that $\sum_{i \in Q} x_i^C = \eta A$. We now upper bound the profit of the resulting induced assignment as a function of β and $\tilde{\alpha}$.

First, since all routes with $x_i^P = 0$ have equal profits per driver, it follows that $x_i = x_i^C = \frac{\tilde{\alpha}}{\beta}a_i$ for those routes with profits per driver given by $4A - \frac{\tilde{\alpha}}{\beta}$. Moreover, since all routes with $x_i^P > 0$ also have the same profits per driver at equilibrium, it follows that for $i \in [n] \backslash Q$: $x_i = a_i\left(\frac{2 - 3\tilde{\alpha}}{4 - 3\beta}\right)$, with

a profit per driver of $4A - \frac{2-3\tilde{\alpha}}{4-3\beta}$. Moreover, we have for route $n+1$ that: $x_{n+1} = A\left(\frac{4}{3} + \frac{2-3\tilde{\alpha}}{12-9\beta}\right)$, resulting in a total profit per driver of $4A + 4 - \frac{3A\left(\frac{4}{3}+\frac{2-3\tilde{\alpha}}{12-9\beta}\right)}{A} = 4A - \frac{2-3\tilde{\alpha}}{4-3\beta}$.

Consequently, the total system profit is given by:

$$\Pi(\tilde{\alpha},\beta) = \tilde{\alpha}A\left(4A - \frac{\tilde{\alpha}}{\beta}\right) + (2-\tilde{\alpha})A\left(4A - \frac{2-3\tilde{\alpha}}{4-3\beta}\right) = A\left[8A - \frac{\tilde{\alpha}^2}{\beta} - \frac{(2-\tilde{\alpha}a)(2-3\tilde{\alpha})}{4-3\beta}\right],$$
$$= A\left[8A - \frac{(4-3\beta)\tilde{\alpha}^2 + \beta(2-\tilde{\alpha})(2-3\tilde{\alpha})}{\beta(4-3\beta)}\right]$$

When we fix β, the above equation has a unique maximizer at $\tilde{\alpha} = \beta$ when $\beta \leq \frac{1}{2}$ and $\tilde{\alpha} = \frac{1}{2}$ when $\beta \geq \frac{1}{2}$. In the setting when $\beta \leq \frac{1}{2}$, we have that the profit is given by: $A(8A - \frac{4-4\beta}{4-3\beta})$. Differentiating this expression with respect to β this expression has a unique maximizer at $\beta = \frac{1}{2}$, giving a total profit of $A(8A - \frac{4}{5}) < A(8A - \frac{3}{4})$. When $\beta \geq \frac{1}{2}$, we have the following expression for the profit when $\tilde{\alpha} = \frac{1}{2}$, resulting in a profit of $A(8A - \frac{1}{\beta(4-3\beta)})$. This profit function is maximized at $\beta = \frac{2}{3}$ at which we get a profit of $A(8A - \frac{3}{4})$. However, since we have a no-instance of partition, we know that $\beta \neq \frac{2}{3}$; hence the profit is strictly smaller than $A(8A - \frac{3}{4})$. We have thus considered all cases, completing our proof.

B.8 Proof of Theorem 4

In this section, we prove Theorem 4. We first establish the Stackelberg optimality of LPF-P and prove the upper bound on the profit ratio of LPF-P in Appendix B.8.1. Then, we analyze the lower bounds for the profit ratio of LPF-P in Appendix B.8.2. Finally, we extend the analysis of the profit-maximization setting to the rider welfare maximization setting in Appendix B.8.3.

B.8.1 Stackelberg Optimality and Profit Ratio Bound of LPF-P

In the following, we prove the Stackelberg optimality result and upper bound on the profit ratio of LPF-P for the profit-maximization objective and defer the details of the lower bound, which follow ideas similar to the proofs of Propositions 2 and 3, and the results for the rider welfare maximization objective to Appendix B.8.

Without loss of generality, order the two routes under the cumulative driver profit-maximizing allocation $\mathbf{x}^*$ such that $\pi_1(x_1^*) \geq \pi_2(x_2^*)$. Moreover, we normalize the total demand to $D = 1$, such that $x_1^* + x_2^* = 1$, and establish our claim by considering two cases: (i) $\alpha \geq x_2^*$, and (ii) $\alpha < x_2^*$.

Case (i): In this case, the LPF-P strategy allocates $x_2^C = x_2^*$ to the second route and $x_1^C = \alpha - x_2^*$ to the first route. Since $\pi_1(x_1^*) \geq \pi_2(x_2^*)$, at equilibrium, all non-centrally controlled drivers use route one, resulting in an outcome where the driver allocation induced by LPF-P satisfies $\mathbf{x}^{LPF} = \mathbf{x}^*$. Thus, in this case, the LPF-P strategy attains the profit-maximizing outcome, and, hence, is also the optimal Stackelberg strategy that maximizes cumulative driver profits.

Case (ii): In this case, all α centrally controlled drivers are allocated to route two under the LPF-P strategy, i.e., $x_2^C = \alpha$ and $x_1^C = 0$. Given this allocation of the centrally controlled drivers, we now analyze two sub-cases: (a) $\pi_1(1-\alpha) \geq \pi_2(\alpha)$ and (b) $\pi_1(1-\alpha) < \pi_2(\alpha)$.

Case (iia): In this case, all the $1-\alpha$ non-centrally controlled drivers will use the first route at equilibrium; hence, the total profit corresponding to the LPF-P strategy is: $P_{LPF} = (1-\alpha)\pi_1(1-\alpha) + \alpha\pi_2(\alpha) = p_1\Lambda_1^M(1-\alpha) + p_2\Lambda_2^M(\alpha)$.

Next, note that at any Stackelberg equilibrium, at least $1-\alpha$ drivers will use route one (as if $x_1 < 1-\alpha$, then the non-centrally controlled users on route two have a profitable deviation). Thus, the optimal Stackelberg strategy can be characterized by the following optimization problem:

$$\max_{x_1} p_1\Lambda_1^M(x_1) + p_2\Lambda_2^M(x_2) \quad \text{s.t.} \quad x_1 \geq 1-\alpha, \quad x_2 \geq 0, \quad x_1 + x_2 = 1.$$

Letting μ be the dual variable corresponding to the constraint $x_1 + x_2 = 1$, the optimal solution $\hat{\mathbf{x}}$ of the above problem satisfies the following inequalities:

$$p_1(\Lambda_1^M)'(\hat{x}_1) \leq \mu, \quad p_2(\Lambda_2^M)'(\hat{x}_2) \leq \mu,$$

where the inequalities are met with equality if $\hat{x}_1 > 1-\alpha$ and $\hat{x}_2 > 0$, respectively.

Next, without loss of generality, we focus on optimal allocations satisfying $x_2^* \leq \tilde{k}_2^*$, as all demand for route two is served for any $x_2 \geq \tilde{k}_2^*$. Then, since $x_2^* > \alpha \geq \hat{x}_2 = 0$, at the profit maximizing outcome, it holds that $p_1(\Lambda_1^M)'(x_1^*) \leq p_2(\Lambda_2^M)'(x_2^*)$, with equality if $x_1^* > 0$. Then, we have:

$$\mu \geq p_2(\Lambda_2^M)'(\hat{x}_2) > p_2(\Lambda_2^M)'(x_2^*) \geq p_1(\Lambda_1^M)'(x_1^*) \geq p_1(\Lambda_1^M)'(\hat{x}_1),$$

where the strict inequality follows due to the the nature of the minibus rider demand function in Equation (3) in the range $x_2 \in [0, \tilde{k}_2^*]$. The above inequalities imply that $p_1(\Lambda_1^M)'(\hat{x}_1) < \mu$, which implies that $\hat{x}_1 = 1-\alpha$ and $\hat{x}_2 = \alpha$. This establishes our claim that the LPF-P strategy is the optimal Stackelberg strategy under the conditions of case (iia).

Case (iib): In this case, there is some $\bar{x}_1 \in [x_1^*, 1-\alpha]$ for which $\pi_1(\bar{x}_1) = \pi_2(1-\bar{x}_1)$. Then, the total profit under the LPF-P strategy is $\pi_1(\bar{x}_1)$. Now suppose that at the optimal Stackelberg strategy, we have an induced driver allocation $\hat{\mathbf{x}}$. Then, it must be that $\hat{x}_1 \geq \bar{x}_1$, as if $\hat{x}_1 < \bar{x}_1$, then non-centrally controlled users on route two have a profitable deviation. Then, formulating the same optimization problem as in case (iia) (where we replace $1-\alpha$ with $\bar{x}_1$) and using the same analytical arguments, we can again show that the LPF-P strategy is the optimal Stackelberg strategy even in case (iib), which establishes our claim regarding the optimality of LPF-P for $n = 2$ routes.

Next, letting $\tilde{\mathbf{x}}$ be the allocation induced by LPF-P, we obtain the following:

$$\begin{aligned}
P(\mathbf{x}^*) &= \pi_1(x_1^*)x_1^* + \pi_2(x_2^*)x_2^* \overset{(a)}{\leq} \pi_1(\tilde{x}_1)\tilde{x}_1 + \pi_2(x_2^*)x_2^* = \pi_1(\tilde{x}_1)\tilde{x}_1 + \pi_2(x_2^*)\tilde{x}_2 + \pi_2(x_2^*)(x_2^* - \tilde{x}_2), \\
&\overset{(b)}{\leq} \pi_1(\tilde{x}_1)\tilde{x}_1 + \pi_2(\tilde{x}_2)\tilde{x}_2 + \pi_2(x_2^*)(x_2^* - \tilde{x}_2) = P(\tilde{\mathbf{x}}) + \pi_2(x_2^*)(x_2^* - \tilde{x}_2), \\
&\overset{(c)}{\leq} P(\tilde{\mathbf{x}}) + \pi_2(\tilde{x}_2)(x_2^* - \tilde{x}_2) \overset{(d)}{\leq} P(\tilde{\mathbf{x}}) + P(\tilde{\mathbf{x}})(x_2^* - \tilde{x}_2) = P(\tilde{\mathbf{x}})\left(1 + (x_2^* - \tilde{x}_2)\right), \\
&\overset{(e)}{\leq} P(\tilde{\mathbf{x}})\left(1 + x_2^* - \alpha\right),
\end{aligned}$$

where (a) follows by the monotonicity of the profit function as $\tilde{x}_1 \geq x_1^*$, (b) follows as $\pi_2(\tilde{x}_2) \geq \pi_2(x_2^*)$ by the monotonicity of the profit per driver and the fact that $\tilde{x}_2 \leq x_2^*$, (c) also follows as $\pi_2(x_2^*) \leq \pi_2(\tilde{x}_2)$, and (d) follows as a convex combination of a set of numbers is greater than its minimum. For (d), note that since $\pi_1(x_1^*) \geq \pi_2(x_2^*)$, it must be that $x_1^P > 0$ (i.e., at least some of the non-centrally controlled drivers take route one); hence, $\pi_1(\tilde{x}_1) \geq \pi_2(\tilde{x}_2)$. Finally, (e) follows as $\tilde{x}_2 \geq \alpha$, as the LPF-P strategy allocates the centralized drivers to the least profit routes first.

The above analysis implies that the profit ratio between the LPF-P and the optimal allocation satisfies $\frac{P(\mathbf{x}^*)}{P(\tilde{\mathbf{x}})} \leq 1 + (x_2^* - \alpha)_+ \leq 2 - \alpha$, where $\alpha \in [0,1]$ is the fraction of the driver population that is centrally controlled in the setting when the number of routes $n = 2$.

B.8.2 Analysis of Lower Bound on Profit Ratio of LPF-P

To prove the result for the profit-maximization objective, we proceed in the vein of Proposition 2, and consider a setting with $n = 2$ routes, a normalized demand $D = 1$, where the cost difference $S_i = 0$ for both routes. We note that the following proof naturally generalizes to the setting when $S_i = \epsilon$ for any arbitrarily small $\epsilon > 0$. Consequently, we obtain piece-wise linear rider demand functions for both routes, as in Equation (4) with $\tilde{k}_i^* = \frac{2l_i\lambda_i}{F}$ for both routes.

Moreover, consider the setting where $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$. In this case, the cumulative driver profit maximizing allocation is $\mathbf{x}^* = (\tilde{k}_1^*, \tilde{k}_2^*)$, resulting in a total maximum achievable profit of $P(\mathbf{x}^*) = p_1\Lambda_1^M(\tilde{k}_1^*) + p_2\Lambda_2^M(\tilde{k}_2^*) = p_1\Lambda_1 + p_2\Lambda_2$.

Next, suppose that profit per driver on the two routes satisfies $\pi_1(1) = \pi_2(0)$, i.e., the informal transit system supports an equilibrium driver allocation $\mathbf{x}^{Eq} = (1, 0)$. For this condition to hold, note that the following equalities must be satisfied:

$$p_1\Lambda_1 = \pi_1(1) = \pi_2(0) = p_2\frac{F(t_2 - t_1)}{2l_2} = p_2\Lambda_2\frac{F}{2l_2\lambda_2} = p_2\Lambda_2\frac{1}{\tilde{k}_2^*}$$

Using this relation, profit ratio between the optimal and above-defined equilibrium allocation is given by:

$$\frac{P(\mathbf{x}^*)}{P(\tilde{\mathbf{x}})} = \frac{p_1\Lambda_1 + p_2\Lambda_2}{p_1\Lambda_1 + \min\left\{\frac{\alpha}{\tilde{k}_2^*}, 1\right\}p_2\Lambda_2} = \frac{p_2\Lambda_2\left(1 + \frac{1}{\tilde{k}_2^*}\right)}{p_2\Lambda_2\left(\frac{1}{\tilde{k}_2^*} + \min\left\{\frac{\alpha}{\tilde{k}_2^*}, 1\right\}\right)} = \frac{\tilde{k}_2^* + 1}{\min\{\alpha, \tilde{k}_2^*\} + 1}.$$

Finally, taking the limit as $\tilde{k}_2^* \to 1$ and $\tilde{k}_1^* \to 0$ while satisfying $\tilde{k}_1^* + \tilde{k}_2^* = D = 1$, the above analysis implies that the profit ratio approaches $\frac{2}{1+\alpha}$, thus establishing our claim.

B.8.3 Analysis of Rider Welfare Maximization

Without loss of generality, order the two routes under the cumulative rider welfare-maximizing solution $\mathbf{x}^*$ such that the per-driver profits satisfy $\pi_1(x_1^*) \geq \pi_2(x_2^*)$. Moreover, we normalize the total demand to $D = 1$, such that $x_1^* + x_2^* = 1$. Then, the proof of the optimality of LPF-R for rider welfare maximization follows almost entirely analogously to proof of the corresponding result for driver profit maximization with the profits $p_i\Lambda_i^M(\cdot)$ replaced with the rider demands $\Lambda_i^M(\cdot)$. We omit the details for brevity and make a note here on equilibrium multiplicity. The only regime when multiple equilibria arise is if $S_i = 0$ for both routes and the following condition holds: $\frac{p_1\Lambda_1}{k_1^*} = \frac{p_2\Lambda_2}{k_2^*}$. In this case, for any $\mathbf{x}^*$ such that $x_1^* + x_2^* \leq k_1^* + k_2^*$, it holds that $\pi_1(x_1^*) = \frac{p_1\Lambda_1}{k_1^*} = \frac{p_2\Lambda_2}{k_2^*} = \pi_2(x_2^*)$. Then, in this special case, to break ties, we order the routes such that LPF-R allocates to the route yielding a higher rider welfare first, i.e., for which $\frac{\Lambda_1}{k_1^*} \geq \frac{\Lambda_2}{k_2^*}$. Under this tie-breaking, LPF-R still yields the Stackelberg-optimal allocation, as the privatized drivers will take either of the routes in response, and this allocation guarantees the maximum fraction of drivers on the route with the higher rider welfare per unit of driver allocated.

We now derive a price of anarchy guarantee for the induced Stackelberg optimal allocation $\tilde{\mathbf{x}}$ under the LPF strategy, where note that $\tilde{x}_1 \geq x_1^*$ and $\tilde{x}_2 \leq x_2^*$. Analogous to the proof of profit maximization setting, note that if $\alpha \geq x_2^*$, then LPF-R attains the optimal rider welfare, achieving

a rider welfare ratio of one. Thus, consider the case when $\alpha < x_2^*$. In this case, we have:

$$\begin{aligned} R(\mathbf{x}^*) &= \Lambda_1^M(x_1^*) + \Lambda_2^M(x_2^*) \overset{(a)}{\leq} \Lambda_1^M(\tilde{x}_1) + \Lambda_2^M(x_2^*), \\ &= \Lambda_1^M(\tilde{x}_1) + \Lambda_2^M(\tilde{x}_2) + (\Lambda_2^M(x_2^*) - \Lambda_2^M(\tilde{x}_2)), \\ &= R(\tilde{\mathbf{x}}) + (\Lambda_2^M(x_2^*) - \Lambda_2^M(\tilde{x}_2)), \\ &\overset{(b)}{=} R(\tilde{\mathbf{x}}) + \frac{\pi_2(x_2^*)x_2^*}{p_2} - \frac{\pi_2(\tilde{x}_2)\tilde{x}_2}{p_2}, \\ &\overset{(c)}{\leq} R(\tilde{\mathbf{x}}) + \frac{\pi_2(\tilde{x}_2)}{p_2}(x_2^* - \tilde{x}_2), \end{aligned}$$

where (a) follows by the monotonicity of the minibus rider demand function, (b) follows as by the definition of the per-driver profit function $\pi_i(x_i) = \frac{p_i \Lambda_i^M(x_i)}{x_i}$, and (c) follows by the monotonicity of the per-driver profit function.

Next, note that

$$\frac{\pi_2(\tilde{x}_2)}{p_2} \leq \frac{\pi_1(\tilde{x}_1)}{p_2} = \frac{\pi_1(\tilde{x}_1)}{p_1}\frac{p_1}{p_2} \leq \frac{\pi_1(\tilde{x}_1)}{p_1}\frac{p_{\max}}{p_{\min}},$$

where $p_{\max}$ and $p_{\min}$ are the maximum and minimum per-rider profits across routes. From the above relation, it follows that:

$$R(\tilde{\mathbf{x}})\frac{p_{\max}}{p_{\min}} = \Lambda_1^M(\tilde{x}_1)\frac{p_{\max}}{p_{\min}} + \Lambda_2^M(\tilde{x}_2)\frac{p_{\max}}{p_{\min}} = \frac{\pi_1(\tilde{x}_1)\tilde{x}_1}{p_1}\frac{p_{\max}}{p_{\min}} + \frac{\pi_2(\tilde{x}_2)\tilde{x}_2}{p_2}\frac{p_{\max}}{p_{\min}} \geq \frac{\pi_2(\tilde{x}_2)}{p_2}.$$

Using the above inequality, we obtain:

$$\begin{aligned} R(\mathbf{x}^*) &\leq R(\tilde{\mathbf{x}}) + \frac{\pi_2(\tilde{x}_2)}{p_2}(x_2^* - \tilde{x}_2) \leq R(\tilde{\mathbf{x}})\left(1 + \frac{p_{\max}}{p_{\min}}(x_2^* - \tilde{x}_2)\right), \\ &\overset{(a)}{\leq} R(\tilde{\mathbf{x}})\left(1 + \frac{p_{\max}}{p_{\min}}(x_2^* - \alpha)\right) \leq R(\tilde{\mathbf{x}})\left(1 + \frac{p_{\max}}{p_{\min}}(1 - \alpha)\right), \end{aligned}$$

where (a) follows as $\tilde{x}_2 \geq \alpha$, as the LPF-R strategy allocates the centrally controlled drivers to the least profit routes first.

The above analysis implies that the rider-welfare ratio satisfies $\frac{R(\mathbf{x}^*)}{R(\tilde{\mathbf{x}})} \leq 1 + \frac{p_{\max}}{p_{\min}}(x_2^* - \alpha)_+ \leq 1 + \frac{p_{\max}}{p_{\min}}(1 - \alpha)$, where $\alpha \in [0, 1]$ is the fraction of the driver population that is centrally controlled in the setting when the number of routes $n = 2$.

Finally, the lower bound on the welfare ratio of LPF-R follows an almost entirely analogous line of reasoning to the lower bound on the profit ratio of LPF-P established in Appendix B.8.2; hence, we omit the details for brevity.

C Additional Discussions and Theoretical Results

C.1 Limited Gains from Route Switching

We show that, once all riders on a route have been served, the potential gains to a driver from switching routes are bounded and unlikely to outweigh the associated switching costs, such as the travel time required to move between geographically separated routes.

To this end, we first note from Corollary 4 that the total service time T_i on any route under our studied framework are bounded above by $\Delta + S_i \left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)$. Letting $\bar{S} = \max_i S_i$, it follows for any two routes i, i' that the difference in the service times satisfies $|T_i - T_{i'}| \leq \bar{S}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)$.

We now show that there exists a bounded switching cost under which drivers would not change routes if $T_i < T_{i'}$. In this case, we can define the driver profit as the sum of three terms: (i) Driver profit on route i, (ii) Driver profit on route i' during the period $T_{i'} - T_i$, and (iii) the negative of the cost to switch routes. Thus, for the drivers to not switch routes, we just require the switching cost to be high enough to cancel out the Driver profit on route i' during the period $T_{i'} - T_i$.

Now, the maximum driver profit on any route i' is bounded above by

$$p_{i'}(\Lambda_{i'}^M)'(0) = \frac{p_{i'}F}{2l_{i'}}\left(\Delta + S_{i'}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right) \leq \frac{p_{\max}F}{2\underline{l}}\left(\Delta + \bar{S}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right),$$

as the rider demand function Λ_i^M has a decreasing slope. Here, we take $p_{\max} = \max_i p_i$ and let $\underline{l} = \min_i l_i > 0$. Thus, for a bounded switching cost that is at least $\left(\Delta + \bar{S}\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)$, drivers will not seek to change routes even if $T_i < T_{i'}$.

C.2 Additional Properties of Budget-Balanced Cross-Subsidization Scheme

Beyond the algorithmic and computational properties highlighted in Section 5, we emphasize the generality of Theorem 3 by highlighting several other properties of the optimal budget-balanced cross-subsidization scheme.

First, if $\mathbf{x}^*$ is a solution to a system optimization problem beyond cumulative driver profit or rider welfare maximization that can only be solved approximately, up to an approximation factor β, then implementing this allocation via cross-subsidization results in an equilibrium that is likewise β-optimal. Thus, cross-subsidization preserves approximation guarantees when translating system-optimal driver allocations into equilibrium outcomes.

Next, since Theorem 3 provides a method to implement any feasible target driver allocation $\mathbf{x}^*$ with at least one strictly positive entry as an equilibrium, it applies not only to allocations optimizing different system objectives but also to allocations satisfying additional constraints, such as upper or lower bounds on driver supply across routes.

Furthermore, while the proof of Theorem 3 analyzed the setting when the equilibrium condition in Equation (6) holds with equality for all routes, it can be directly extended to settings when this equilibrium condition holds with a strict inequality for routes where the target allocation satisfies $x_j^* = 0$.

Finally, while our equilibrium condition assumes that drivers choose routes based solely on relative per-driver profits, selecting the route that maximizes earnings without imposing requirements on the absolute level of profits, our framework can be readily generalized to incorporate individual rationality constraints when drivers have an outside option with reservation wage W. In particular, if a target allocation $\mathbf{x}^*$ satisfies individual rationality for drivers, so that $\pi_j^* \geq W$ for all routes j with $x_j^* > 0$, then this condition is preserved under cross-subsidization:

$$\tilde{\pi}^{Eq} = \frac{\sum_{i\in[n]} \pi_i^* x_i^*}{\sum_{i\in[n]} x_i^*} \geq \frac{W\sum_{i\in[n]} x_i^*}{\sum_{i\in[n]} x_i^*} = W.$$

Thus, drivers are never made worse off under cross-subsidization relative to their reservation wage.

C.3 Uniqueness and Multiplicity of Equilibria under Cross-Subsidization

An equilibrium driver allocation induced by a cross-subsidy scheme may, in general, be non-unique. This not influence the validity of Theorem 3 and Corollary 3, as our cross-subsidization scheme guarantees that the target allocation is an equilibrium, irrespective of the existence of other equilibria. That said, in the regime when the cost difference $S_i > 0$ for all routes i, a condition that commonly holds in informal transit systems, where the cost of using the minibus without rider queuing and schedule delays is substantially lower than that of the outside option, the equilibrium driver allocation is guaranteed to be unique (see Appendix C.4). More generally, even when multiple equilibria arise, cumulative driver profits are identical across all equilibrium allocations, although rider welfare may vary across equilibrium allocations (see Appendix C.4). Thus, potential equilibrium multiplicity does not undermine the effectiveness of cross-subsidization in aligning driver incentives with the cumulative driver profit metric. While equilibrium multiplicity under rider welfare may lead to the realization of an equilibrium allocation with a lower rider welfare than the optimal, this issue is unlikely to arise in the empirically relevant regimes described above.

C.4 Uniqueness of Cumulative Driver Profits and Equilibrium Driver Allocation

In this section, we show that the cumulative driver profits are the same at any equilibrium allocation for the rider demand functions specified in Equation (3). Moreover, if the quadratic coefficient in Equation (3) is strictly positive (i.e., if the free-flow cost difference $S_i > 0$), we show that the resulting equilibrium driver allocation is unique.

To see this, consider two equilibrium driver allocations $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, where $\mathbf{x}^{(1)} \neq \mathbf{x}^{(2)}$. Then, since $\sum_{i\in[n]} x_i^{(1)} = D = \sum_{i\in[n]} x_i^{(2)}$, it follows that there exist non-empty sets $L_1 = \{i : x_i^{(1)} > x_i^{(2)}\}$ and $L_2 = \{i : x_i^{(1)} < x_i^{(2)}\}$. Moreover, define $L_3 = \{i : x_i^{(1)} = x_i^{(2)}\}$ and let $\pi^{(1)}$ be the equilibrium per-driver profit under $\mathbf{x}^{(1)}$, where $\pi_i(x_i^{(1)}) = \pi^{(1)}$ for all routes i with $x_i^{(1)} > 0$. Analogously define $\pi^{(2)}$. Then, for some routes $i_1 \in L_1$ and $i_2 \in L_2$, we obtain the following relation for the per-driver profit:

$$\pi_{i_1}(x_{i_1}^{(2)}) \overset{(a)}{\geq} \pi_{i_1}(x_{i_1}^{(1)}) = \pi^{(1)} \overset{(b)}{\geq} \pi_{i_2}(x_{i_2}^{(1)}) \overset{(c)}{\geq} \pi_{i_2}(x_{i_2}^{(2)}) = \pi^{(2)} \overset{(d)}{\geq} \pi_{i_1}(x_{i_1}^{(2)}), \tag{8}$$

where (a) and (c) follow by the monotonicity of the per-driver profit established in Corollary 2, and (b) and (d) follow since $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are equilibrium driver allocations. Since the left and right most terms in the above sequence of inequalities are the same, it follows that each of the above inequalities is met with an equality and, in particular, $\pi^{(1)} = \pi^{(2)}$.

Given this, we now compare the cumulative profits under the two equilibrium driver allocations. In particular, we have:

$$P(\mathbf{x}^{(1)}) = \sum_{i=1}^{n} \pi_i(x_i^{(1)}) x_i^{(1)} = \pi^{(1)} \sum_{i=1}^{n} x_i^{(1)} \overset{(a)}{=} \pi^{(2)} \sum_{i=1}^{n} x_i^{(2)} = \sum_{i=1}^{n} \pi_i(x_i^{(2)}) x_i^{(2)} = P(\mathbf{x}^{(2)}),$$

where (a) follows as $\pi^{(1)} = \pi^{(2)}$ and the same number of drivers are allocated under both allocations. Thus, we have that the cumulative profits under any two equilibrium allocations are the same.

Next, we show that the equilibrium allocation is unique if the quadratic coefficient in Equation (3) is strictly positive. To see this, first note that under this condition, the per-driver profits is

strictly monotonically decreasing in the driver allocation. Then, the only way for the inequalities in Equation (8) to be held with equality is if $x_{i_1}^{(1)} = x_{i_1}^{(2)}$ for all $i \in L_1$ and $x_{i_2}^{(1)} = x_{i_2}^{(2)}$ for all $i \in L_2$, a contradiction. Hence, it must follow that $\mathbf{x}^{(1)} = \mathbf{x}^{(2)}$, which establishes our claim.

D Additional Details on Numerical Experiments

D.1 Details on Route Selection

Our original dataset on the informal transit system in Nalasopara contains information on twenty-one routes. However, some of these routes serve destinations in the same geographic area, for which only a single destination coordinate is available. In these cases, we aggregate routes by constructing weighted averages of route characteristics, including trip time, distance, and fare. In addition, geographic coordinates are unavailable for one route, which are required for demand estimation. Thus, for our experiments, we focus on a final set of eighteen routes.

D.2 Rider Demand Calibration Procedure

We calibrate the demand Λ_i for each route i using population statistics from the 2011 Indian Census, the most recent census conducted in India, which reports data at the level of administrative subdivisions referred to as *wards*. To this end, we first assign each route i to a ward based on the geographic coordinates of its destination stop. Since we study the evening commute period in our experiments, we take the Nalasopara railway station to be the common origin of all routes with the destination being the respective residential neighborhoods, reflecting that the majority of trips during the study period are home-bound.

Next, we estimate the population that can feasibly access each route by defining a pedestrian catchment area around its destination stop based on walking accessibility. Specifically, for routes with total length of at most 5 km, we assume users are willing to walk five minutes to access the route, while for longer routes we assume a walking radius of ten minutes. Using a walking speed of 1.3 m/s, we obtain circular catchment areas around each destination stop.

However, since these circular catchment areas may overlap across destination stops, we then generate disjoint catchments by assigning users to their nearest stop, ensuring that each user is associated with at most one route. To do so, we use Monte Carlo nearest-neighbor sampling. Specifically, for each circular catchment area, we repeatedly draw a large number of points uniformly at random, interpreting each draw as a representative residential location, and determine the fraction of those points that are geographically closest (in Euclidean distance) to that catchment area's destination stop. The fraction of points assigned to a given destination stop determines the share of the circular catchment area that can access that stop. Assuming that population density within the circular catchment area is equal to the ward-level population density of the ward to which the destination stop belongs, we compute the total population that can access each route by multiplying the total population within the circular catchment area by this fraction.

Finally, we assume that 40% of the population commute daily by train, consistent with standard mode share estimates in the Mumbai Metropolitan Region. Multiplying this commuting share with the above estimated population that can access each route yields the total mass of users Λ_i seeking to make trips on that route during our study period.

D.3 Implementation Details to Compute Equilibrium Driver Allocation

We now describe a procedure to compute an approximate equilibrium driver allocation, where we leverage the fact that, in equilibrium, all routes with a strictly positive mass of drivers have the same per-driver profit. Specifically, we search over a discretized set of possible equilibrium profit levels within a bounded interval $[\underline{\pi}, \bar{\pi}]$, where $\underline{\pi}, \bar{\pi}$ denote lower and upper bounds on the equilibrium per-driver profits. For each candidate per-driver profit level on the discretized grid, we compute the corresponding allocation of drivers across routes that would result in per-driver profits at that level and sum these allocations across the routes to obtain the total number of drivers willing to operate at that per-driver profit level. Then, we obtain an approximate equilibrium per-driver profit as the value on the discretized grid for which the resulting aggregate driver allocation is closest to the actual mass of drivers D, with the corresponding route-level allocation of drivers across routes taken as the approximate equilibrium driver allocation.

D.4 Equilibrium Driver Allocation Across Routes

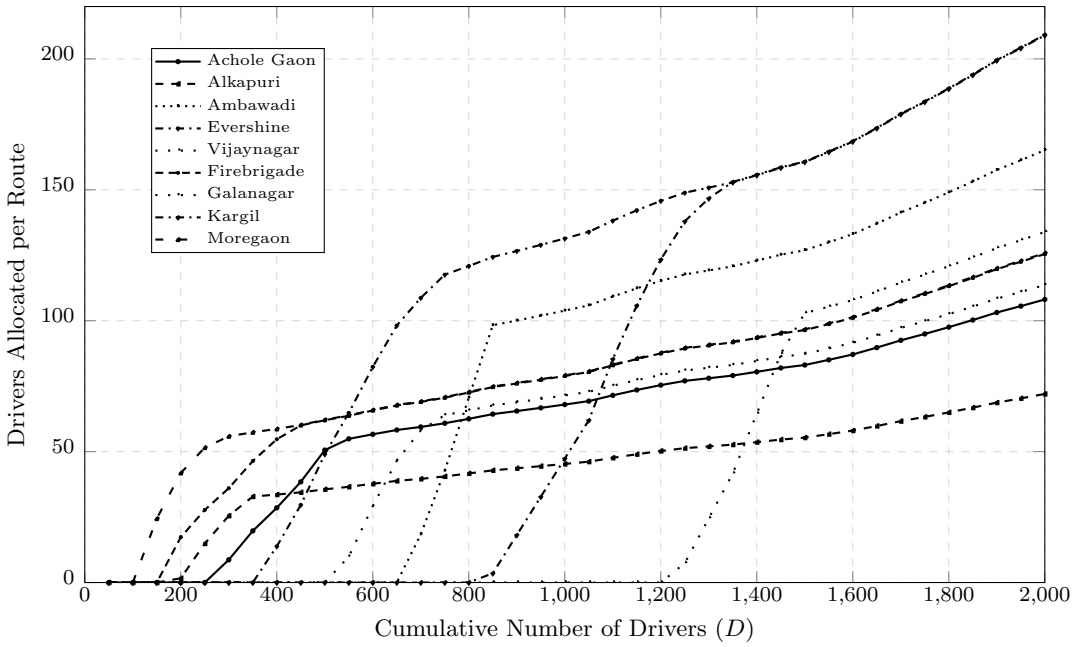


Figure 6: Depiction of the equilibrium number of drivers allocated to nine of the eighteen routes in our dataset from Nalasopara as the cumulative number of drivers D is varied.

E L-NCF Optimality for $S \equiv 0$

In this section, we prove that L-NCF is optimal for Stackelberg Routing whenever $S_i = 0$ for all routes, for both the profit P and demand R objectives. When $S_i = 0$, we have $\Lambda_i^M(x) = \widetilde{\Lambda}_i^M(x)$ for all x. In other words, when $S \equiv 0$, the approximation L-NCF makes to the objective function

is actually not an approximation (it is exact). This section will use the notational setup from Section F.1.

Consider the Stackelberg game where a leader determines the allocation for coalition drivers $\underline{y}$. Then, the non-coalition drivers choose their route allocations x according to the profit function $\widetilde{P}$. x is a user equilibrium in the following sense: all routes that have positive allocations of non-coalition drivers have the same profit per driver. Mathematically, this means that for all i, j with $x_i, x_j > 0$, we have

$$\frac{p_i \widetilde{\Lambda}_i^M(x_i + y_i)}{x_i + y_i} = \frac{p_j \widetilde{\Lambda}_j^M(x_j + y_j)}{x_j + y_j}.$$

Let x^0 denote a user equilibrium that arises when $y = 0$, and let π^0 be the corresponding equilibrium profit per driver. Furthermore, define the following indices based on x^0:

$$\begin{aligned} E_T &:= \left\{ i : x_i^0 > k_i^* \right\} \\ E_P &:= \left\{ i : x_i^0 \in (0, k_i^*] \right\} \\ E_N &:= \left\{ i : x_i^0 = 0 \right\}. \end{aligned}$$

Note that x^0 is not unique if $|E_P| > 1$. More generally, the equilibrium is not unique if and only if $E_{\bar{P}}(\pi^0) := \{i : p_i \widetilde{\zeta}_i = \pi^0\}$ has cardinality at least two. To address this, we introduce the following tie-breaking rule. Define a rank function $\text{Rank} : [n] \to [n]$ where for $i, j \in [n]$, $\widetilde{\zeta}_i < \widetilde{\zeta}_j \implies \text{Rank}(i) > \text{Rank}(j)$. If there are ties, i.e., if there exist $i \neq j$ for which $\widetilde{\zeta}_i = \widetilde{\zeta}_j$, then break the tie arbitrarily. When the equilibrium is not unique, we will pick the one which maximizes $\sum_{i \in E_{\bar{P}}(\pi^0)} \text{Rank}(i) x_i$. Namely, the equilibrium that prioritizes allocating to the highest ranked routes in $E_{\bar{P}}$.

In this section, we will show that the L-NCF algorithm that chooses y according to the following optimization problem

$$\max_{y \in \mathbb{R}_+^n : \|y\|_1 = \alpha m} \widetilde{W}(x^0 + y)$$

achieves the maximum demand served in this Stackelberg game when $\widetilde{W}$ is chosen to be $\widetilde{R}$, and achieves the maximum profit in this Stackelberg game when $\widetilde{W}$ is chosen to be $\widetilde{P}$.

First, we will show that for any Stackelberg equilibrium (x, y), it must be the case that $x_i + y_i \geq x_i^0$ for all $i \in E_T$. We argue by contradiction. Let π be the equilibrium profit per driver for (x, y). Suppose there exists $i \in S$ for which $x_i + y_i < x_i^0$. Note that on the interval (k_i^*, ∞), the profit per driver on route i is given by $\frac{p_i \widetilde{\Lambda}_i^M(z)}{z} = \frac{p_i \Lambda_i}{z}$ and is thus strictly decreasing. Since the profit per driver on route i is always non-increasing, this means that $\pi = \frac{p_i \widetilde{\Lambda}_i^M(x_i + y_i)}{(x_i + y_i)} > \pi^0$. Since $\sum_j x_j = \sum_j x_j^0$ and $x \neq x^0$ (since in particular we know $x_i < x_i^0$), there must exist some index i' for which $x_{i'}^0 < x_{i'}$. But this leads to a contradiction:

$$\pi^0 \overset{(a)}{\geq} \frac{p_{i'} \widetilde{\Lambda}_{i'}^M(x_{i'}^0)}{(x_{i'}^0)} \overset{(b)}{\geq} \frac{p_{i'} \widetilde{\Lambda}_{i'}^M(x_{i'})}{(x_{i'})} \overset{(c)}{\geq} \frac{p_{i'} \widetilde{\Lambda}_{i'}^M(x_{i'} + y_{i'})}{(x_{i'} + y_{i'})} = \pi > \pi^0.$$

Here, (a) is by definition of x^0 being a user equilibrium. (b) and (c) are due to the fact that the profit per driver for any route is a non-increasing function of the number of drivers allocated to the route. Hence it must be the case that $x_i + y_i \geq x_i^0$ for all $i \in E_T$.

Second, we will show that for any Stackelberg equilibrium (x, y), where π is the corresponding equilibrium profit per driver and x is the unique user equilibrium that maximizes $\sum_{i \in E_{\bar{P}}(\pi)} \text{Rank}(i) x_i$ (as per the tiebreaking rule), it must be the case that $x_i + y_i \geq x_i^0$ for all $i \in E_P$. If $\pi < \pi^0$, then $x_i + y_i > k_i^*$ for all $i \in E_P$, and as a consequence, $x_i + y_i \geq x_i^0$. In the case that $\pi = \pi^0$, we will argue by contradiction. Suppose $\{j \in E_P : x_j + y_j < x_j^0\}$ is non-empty. Then there exists an index in this set with the highest rank, which we will denote i. First, note that $\pi = \pi^0$ has the following implications.

- $x_i + y_i = x_i^0$ for all $i \in E_T$. This is because profit per driver is a non-decreasing function of the number of drivers allocated to the route. Furthermore, it is strictly decreasing in the interval (k_i^*, ∞). Since $x_i^0 \in (k_i^*, \infty)$, x_i^0 is the only driver allocation to this route that will result in π^0 profit per driver on this route.
- $x_i + y_i \leq k_i^*$ for all $i \in E_P$. This is because the profit per driver is strictly decreasing on the interval (k_i^*, ∞) and constant on $[0, k_i^*]$.

Note that $x_i + y_i < x_i^0 < k_i^*$ implies that:

- $x_j^0 = k_i^*$ for all $j \in E_P$ where $\text{Rank}(j) > \text{Rank}(i)$. Indeed, x^0 would only allocate to route i if it has finished allocating to all higher ranked routes in E_P.
- $x_j = 0$ for all $j \in E_P$ for which $\text{Rank}(j) < \text{Rank}(i)$. Indeed, drivers would be allocated to i before they would be allocated to any such j.

From these observations, we see that $x_i \leq x_i^0$ for all indices in $E_T \cup \{j \in E_P \setminus \{i\}\}$. However, these observations lead to the following contradiction:

$$\begin{aligned}
\sum_j x_j^0 &= \sum_{j \in T} x_j^0 + \sum_{j \in P \setminus \{i\}} x_j^0 + x_i^0 \\
&\geq \sum_{j \in T} x_j + \sum_{j \in P \setminus \{i\}} x_j + x_i^0 \\
&> \sum_{j \in T} x_j + \sum_{j \in P \setminus \{i\}} x_j + x_i \\
&= \sum_j x_j.
\end{aligned}$$

Indeed, this contradicts the fact that the total amount of non-coalition drivers should be the same in every equilibrium (in particular it is $(1 - \alpha)m$). Hence by contradiction it must be the case that $x_i + y_i \geq x_i^0$ for all $i \in E_P$.

Third, $x_i^0 = 0$ for all $i \in E_N$. Thus we have shown that we have shown that every Stackelberg equilibrium (x, y) must satisfy $x_i + y_i \geq x_i^0$ for all i. Therefore the solution to the following

optimization problem is an upper bound on the maximum demand served by any Stackelberg equilibrium:

$$\max_{z:z\succeq x^0,\|z\|_1=m} \widetilde{W}(z).$$

Fourth, the L-NCF algorithm chooses y according to the following optimization problem

$$\max_{0\preceq y\preceq(k^*-x^0)_+,\|y\|_1\leq\alpha m} \widetilde{W}(x^0+y).$$

Let $y^{\text{L-NCF}}$ be any solution to the L-NCF optimization program. It is straightforward to show that these two programs have the same optimal value, and that the profit per driver for all routes in $T\cup P$ (i.e., the support of x^0) is still π^0, meaning that $(x^0, y^{\text{L-NCF}})$ is a Stackelberg equilibrium whose demand served is an upper bound on the demand served of any Stackelberg equilibrium. This proves that L-NCF is an optimal algorithm for this problem, for both profit and welfare objectives.

F NCF Performance Guarantees (Profit Objective)

F.1 Setup and Notation

Recall that

$$\Lambda_i^M(x_i) = \frac{F}{2l_i}\left(t_2-t_1+S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)x_i - \left(\left(\frac{F}{2l_i}\right)^2 S_i\frac{\eta_E+\eta_L}{\eta_E\eta_L}\frac{t_2-t_1}{\Lambda_i}\right)x_i^2. \tag{9}$$

To this end, let us define the following coefficients:

$$\zeta_{i,1} := \frac{F}{2l_i}\left(t_2-t_1+S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)$$
$$\zeta_{i,2} := \left(\frac{F}{2l_i}\right)^2 S_i\frac{\eta_E+\eta_L}{\eta_E\eta_L}\frac{t_2-t_1}{\Lambda_i}$$

so that $\Lambda_i^M(x) = \zeta_{i,1}x - \zeta_{i,2}x^2$ whenever $x\leq k_i^* = \frac{2l_i\Lambda_i}{F(t_2-t_1)}$, and otherwise we have $\Lambda_i^M(x)=\Lambda_i$.

It is clear that $\zeta_{i,2}\geq 0$ for all i, meaning that λ_i^M is a concave function of x. Therefore, the function

$$\widetilde{\Lambda}_i^M(x) := \Lambda_i\min\left(\frac{x}{k_i^*},1\right)$$

is a pointwise lower bound of Λ_i^M. Indeed, for any $x\geq k_i^*$, we have $\Lambda_i^M(x)=\widetilde{\Lambda}_i^M(x)=\Lambda_i$. For the other case, whenever $x<k_i^*$, define $\tau = x/k_i^*\in[0,1]$. By concavity we have:

$$\begin{aligned}\Lambda_i^M(x) &= \Lambda_i^M\left(k_i^*\tau+0(1-\tau)\right)\\ &\geq \tau\Lambda_i^M\left(k_i^*\right)+(1-\tau)\Lambda_i^M\left(0\right)\\ &= \Lambda_i\frac{x}{k_i^*} = \widetilde{\Lambda}_i^M(x).\end{aligned}$$

Also using concavity, we have

$$\begin{aligned}\Lambda_i^M(x) &\leq \zeta_{i,1}\min\left(x, k_i^*\right) = \zeta_{i,1}k_i^*\min\left(\frac{x}{k_i^*}, 1\right) = \frac{\zeta_{i,1}k_i^*}{\Lambda_i}\widetilde{\Lambda}_i^M(x)\\ &\overset{(a)}{=} \frac{F}{2l_i}\left(t_2 - t_1 + S_i\left(\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\right)\frac{2l_i\Lambda_i}{F(t_2-t_1)}\frac{1}{\Lambda_i}\widetilde{\Lambda}_i^M(x)\\ &= \left(1 + \frac{S_i}{t_2-t_1}\frac{\eta_E+\eta_L}{\eta_E\eta_L}\right)\widetilde{\Lambda}_i^M(x)\\ &\overset{(b)}{\leq} (1+\gamma)\widetilde{\Lambda}_i^M(x)\end{aligned}$$

where (a) is obtained by substituting the expressions for $\zeta_{i,1}, k_i^*$, and (b) is obtained by defining $\gamma := \max_i \frac{S_i}{t_2-t_1}\frac{\eta_E+\eta_L}{\eta_E\eta_L}$.

Hence, we have

$$\widetilde{\Lambda}_i^M(x) \leq \Lambda_i^M(x) \leq (1+\gamma)\widetilde{\Lambda}_i^M(x). \tag{10}$$

We denote the system profit as $P(x) := \sum_{i=1}^n p_i\Lambda_i^M(x_i)$, and use $\widetilde{P}(x) := \sum_{i=1}^n p_i\widetilde{\Lambda}_i^M(x_i)$ to denote an approximate profit function which will be used by the L-NCF algorithm.

For notational convenience, for each i, we use $\widetilde{\zeta}_{i,1} := \frac{\Lambda_i}{k_i^*}$ to denote the maximum profit per driver for route i under the $\widetilde{P}$ objective function. Note that $\widetilde{\zeta}_{i,1} \leq \zeta_{i,1}$.

It will be helpful for the analysis to express the profits (for both $\Pi, \widetilde{\Pi}$ profit functions) in terms of their coalition and non-coalition components:

$$\begin{aligned}P(x+y) &= P_x(x+y) + P_y(x+y)\\ &\text{where } P_x(x+y) := \sum_i p_i\frac{x_i}{(x_i+y_i)}\lambda_i^M(x_i+y_i),\\ &\qquad P_y(x+y) := \sum_i p_i\frac{y_i}{(x_i+y_i)}\lambda_i^M(x_i+y_i).\\ \widetilde{P}(x+y) &= \widetilde{P}_x(x+y) + \widetilde{P}_y(x+y)\\ &\text{where } \widetilde{P}_x(x+y) := \sum_i p_i\frac{x_i}{(x_i+y_i)}\widetilde{\lambda}_i^M(x_i+y_i),\\ &\qquad \widetilde{P}_y(x+y) := \sum_i p_i\frac{y_i}{(x_i+y_i)}\widetilde{\lambda}_i^M(x_i+y_i).\end{aligned}$$

F.2 The L-NCF algorithm and its intermediate steps

There are a total of m drivers, α of which are coordinated by a driver union. Here is how the L-NCF algorithm works:

1. The coalition simulates how the $(1-\alpha)m$ non-compliant drivers would allocate themselves, x^0, assuming that the rider demand functions are $\left\{\widetilde{\Lambda}_i^M\right\}_i$. Namely, x^0 is a driver user equilibrium.

All routes with a positive allocation of non-compliant drivers have the same profit per driver. Call this value π^0.

2. Upon observing x^0, the coalition picks an allocation y of its drivers to maximize system profit: $y \in \underset{y \succeq 0, \|y\|_1 = \alpha m}{\operatorname{argmax}} \widetilde{P}(x^0 + y)$.

3. Given the presence of the coalition y, the non-compliant drivers choose their allocation x^1, which is a user equilibrium with respect to the true rider demand functions $\left\{\Lambda_i^M\right\}_i$. Let π^1 be the resulting equilibrium profit per non-compliant driver.

F.3 Proof structure

Let y^* be an optimal Stackelberg strategy for the coalition under the profit function P. Let x^* be the corresponding response from the non-coalition drivers under profit function P.

First, we bound $P(x^* + y^*)$, i.e., the performance of y^*, by its performance under the approximated profit function $\widetilde{P}$. Specifically, if $\widetilde{x}^*$ is the response from the non-coalition drivers to y^* under the profit function $\widetilde{P}$, then:

Step 1: We show that $P(x^* + y^*) \leq (1+\gamma)\widetilde{P}(\widetilde{x}^* + y^*)$.

Step 2: y^* is a feasible Stackelberg strategy under $\widetilde{\Pi}$. The L-NCF algorithm achieves a profit of $\widetilde{P}(x^0 + y)$. Furthermore, L-NCF is an optimal algorithm for this problem, hence

$$\begin{aligned} &\widetilde{P}(\widetilde{x}^* + y^*) \leq \widetilde{P}(x^0 + y) \\ \implies &P(x^* + y^*) \leq (1+\gamma)\widetilde{P}(x^0 + y). \end{aligned}$$

Step 3: Finally, we show that L-NCF's utility under $\widetilde{P}$ transfers to P, namely: $\widetilde{P}(x^0 + y) \leq P(x^1 + y)$.

Putting everything together, we see that L-NCF is guaranteed to obtain a $\frac{1}{1+\gamma}$ fraction of the optimal profit in the Stackelberg game:

$$P(x^* + y^*) \leq (1+\gamma)P(x^1 + y).$$

We prove Step 2 in Section E, so all that remains is to prove step 1 and step 3.

F.4 Proof of Step 1

Let y^* be an optimal stackelberg strategy for the coalition under the profit function P. Let x^* be the driver user equilibrium in response to y^* under profit function P, and $\widetilde{x}^*$ be the driver user equilibrium in response to y^* under profit function $\widetilde{P}$. Let $\pi^*, \widetilde{\pi}^*$ denote the equilibrium profits per driver (for non-coalition drivers) under $x^* + y^*$ and $\widetilde{x}^* + y^*$ respectively. In this section, we will upper bound $P(x^* + y^*)$ in terms of $\widetilde{P}(\widetilde{x}^* + y^*)$.

Define the following sets based on $x_i^* + y_i^*$:

$$E_T^* := \{i : x_i^* + y_i^* > k_i^*\}$$
$$E_P^* := \{i : x_i^* + y_i^* \in (0, k_i^*]\}$$
$$E_N^* := \{i : x_i^* + y_i^* = 0\}.$$

Note that x^* is not necessarily an equilibrium response to y^* under profit function $\widetilde{P}$. Indeed, note that

$$\text{for } i \in E_T^*, p_i \frac{\widetilde{\Lambda}_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = p_i \frac{\Lambda_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = \pi^*,$$
$$\text{for } i \in E_P^*, p_i \frac{\widetilde{\Lambda}_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = p_i \widetilde{\zeta}_{i,1} < p_i \frac{\Lambda_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = \pi^*,$$
$$\text{for } i \in E_P^*, p_i \frac{\widetilde{\Lambda}_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} \geq \frac{1}{1+\gamma} p_i \frac{\Lambda_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = \frac{1}{1+\gamma}\pi^*,$$
$$\text{for } i \in E_N^*, p_i \frac{\widetilde{\Lambda}_i^M(x_i^* + y_i^*)}{x_i^* + y_i^*} = p_i \widetilde{\zeta}_{i,1} \leq p_i \zeta_{i,1} \leq \pi^*.$$

In particular the routes in E_T^* have higher profit per driver than the routes in P^*. Thus the equilibrium $\widetilde{x}^*$ is achieved by moving non-coalition drivers from E_P^* to E_T^*. As a result, we have

$$\widetilde{\pi}^* \in \left[\frac{\pi^*}{1+\gamma}, \pi^*\right] \text{ and } \widetilde{\pi}^* \geq \max_{i \in E_P^* \cup E_N^*} p_i \widetilde{\zeta}_{i,1}.$$

Summing over all non-coalition drivers, we have:

$$\widetilde{P}_x(\widetilde{x}^* + y^*) \geq \alpha m \frac{\pi^*}{1+\gamma} = \frac{1}{1+\gamma} P_x(x^* + y^*).$$

Next, we analyze the profit from the coalition drivers. There are two cases.

1. For $i \in E_P^* \cup E_N^*$, we have $\widetilde{\pi} \geq p_i \widetilde{\zeta}_{i,1}$, meaning that $\widetilde{x}_i^* + y_i \leq k_i^*$. From this we can deduce that
$$p_i \frac{y_i^*}{\widetilde{x}_i^* + y_i^*} \widetilde{\Lambda}_i^M(\widetilde{x}_i^* + y_i^*) = p_i \zeta_{i,1} y_i^* = p_i \frac{y_i^*}{x_i^* + y_i^*} \widetilde{\Lambda}_i^M(x_i^* + y_i^*) \geq \frac{1}{1+\gamma} p_i \frac{y_i^*}{x_i^* + y_i^*} \Lambda_i^M(x_i^* + y_i^*).$$

2. For $i \in T^*$, note that we are in the regime where $\Lambda_i^M = \widetilde{\Lambda}_i^M$, hence
$$p_i \frac{y_i^*}{x_i^* + y_i^*} \Lambda_i^M(x_i^* + y_i^*) = \pi^* y_i^*,$$
$$p_i \frac{y_i^*}{\widetilde{x}_i^* + y_i^*} \widetilde{\Lambda}_i^M(\widetilde{x}_i^* + y_i^*) = \widetilde{\pi}^* y_i^*.$$

3. Thus in either case, we have
$$p_i \frac{y_i^*}{\widetilde{x}_i^* + y_i^*} \widetilde{\Lambda}_i^M(\widetilde{x}_i^* + y_i^*) \geq \frac{1}{1+\gamma} p_i \frac{y_i^*}{x_i^* + y_i^*} \Lambda_i^M(x_i^* + y_i^*).$$

Summing over all i gives $\widetilde{P}_y(\widetilde{x}^* + y^*) \geq \frac{1}{1+\gamma} P_y(x^* + y^*)$

Putting everything together, we have

$$\widetilde{P}(\widetilde{x}^* + y^*) = \widetilde{P}_x(\widetilde{x}^* + y^*) + \widetilde{P}_y(\widetilde{x}^* + y^*) \geq \frac{1}{1+\gamma} P_x(x^* + y^*) + \frac{1}{1+\gamma} P_y(x^* + y^*) = \frac{1}{1+\gamma} P(x^* + y^*),$$

which is the desired result.

F.5 Proof of Step 3

Define the following indices based on x^0:

$$\begin{aligned} E_T &:= \left\{ i : x_i^0 > k_i^* \right\} \\ E_P &:= \left\{ i : x_i^0 \in (0, k_i^*] \right\} \\ E_N &:= \left\{ i : x_i^0 = 0 \right\}. \end{aligned}$$

x^0 being a driver user equilibrium under the tilde demand curves means that

$$\frac{p_i \widetilde{\lambda}_i^M(x_i^0)}{x_i^0} = \pi^0 \text{ for all } i \in E_T \cup E_P.$$

When computing y, the coalition will only allocate drivers to links in $E_P \cup E_N$. Furthermore, when allocating supply to a link $i \in E_P \cup E_N$, it will ensure that $x_i^0 + y_i \leq k_i^*$. Hence for every link $i \in E_T \cup E_P$,

- If $i \in E_T$, then $x_i^0 + y_i = x_i^0$, hence $\frac{p_i \widetilde{\Lambda}_i^M(x_i^0 + y_i)}{x_i^0 + y_i} = \frac{p_i \widetilde{\Lambda}_i^M(x_i^0)}{x_i^0} = \pi^0$.
- If $i \in E_P$, then $x_i^0 + y_i \leq k_i^*$, hence $\frac{p_i \widetilde{\Lambda}_i^M(x_i^0 + y_i)}{x_i^0 + y_i} = \frac{p_i \widetilde{\zeta}_{i,1}(x_i^0 + y_i)}{x_i^0 + y_i} = \frac{p_i \zeta_{i,1} x_i^0}{x_i^0} = \pi^0$.

Hence x^0 is still a driver user equilibrium even in the presence of the coalition vehicles y.

Now let us compare x^1 to x^0. To this end, let us first analyze x^0 under the profit function P, in the presence of the coalition's allocation y. When replacing $\widetilde{P}$ with P, note that

- If $i \in E_T$, then $x_i^0 \geq k_i^*$ and $y_i = 0$. Hence $\frac{p_i \lambda_i^M(x_i^0 + y_i)}{x_i^0 + y_i} = \frac{p_i \widetilde{\lambda}_i^M(x_i^0 + y_i)}{x_i^0 + y_i} = \pi^0$.
- If $i \in E_P$, then $x_i^0 + y_i \in (0, k_i^*)$, meaning that $\frac{p_i \lambda_i^M(x_i^0 + y_i)}{x_i^0 + y_i} > \frac{p_i \widetilde{\lambda}_i^M(x_i^0 + y_i)}{x_i^0 + y_i} = \pi^0$.
- If $i \in E_N$, both of the following cases are possible: $p_i \zeta_{i,1} \geq \pi_0$ and $p_i \zeta_{i,1} < \pi_0$.

Similar to the proof of Step 1, we will show that $\widetilde{P}(x^0 + y) \leq P(x^1 + y)$ by considering the profit contributions from the non-coalition and coalition drivers separately.

For non-coalition drivers, under the allocation $x^0 + y$ and profit function P, routes in E_P (and possibly also in E_N) are more appealing than those in E_T. Non-coalition drivers will thus move from E_T to $E_P \cup E_N$. This will improve the profit per driver of routes in E_T, and the resulting

equilibrium will have a profit per driver π^1 that is at least as large as π^0. Summing over all non-coalition drivers,

$$\widetilde{P}_x(x^0+y) = (1-\alpha)m\pi^0 < (1-\alpha)m\pi^1 = P_x(x^1+y),$$

hence $\widetilde{P}_x(x^0+y) \leq P_x(x^1+y)$.

For the coalition drivers, note that by construction, we know that $y_i = 0$ for all $i \in E_T$ and $x_i^0 + y_i \leq k_i^*$ for all $i \in E_P \cup E_N$. So we can focus the discussion on the routes in $E_P \cup E_N$. Since

$$\pi^1 \geq \pi^0 \geq \frac{p_i \widetilde{\Lambda}_i^M(x_i^0+y_i)}{x_i^0+y_i},$$

we can deduce that $x_i^1 + y_i \leq k_i^*$ for all $i \in E_P \cup E_N$. Indeed, if there was some i that violated this, then it must be the case that $x_i^1 > 0$ (since $y_i \leq k_i^*$), which gives the following contradiction:

$$\pi^1 = \frac{p_i \Lambda_i^M(x_i^1+y_i)}{x_i^1+y_i} = \frac{p_i \Lambda_i^M(k_i^*)}{x_i^1+y_i} < \frac{p_i \Lambda_i}{k_i^*} = p_i \widetilde{\zeta}_{i,1} \leq \pi^0.$$

Therefore $x_i^1 + y_i \leq k_i^*$ for all $i \in E_P \cup E_N$. From this, for any $i \in E_P \cup E_N$,

$$\begin{aligned} p_i \frac{y_i}{x_i^1+y_i} \Lambda_i^M(x_i^1+y_i) &\geq p_i \frac{y_i}{x_i^1+y_i} \widetilde{\Lambda}_i^M(x_i^1+y_i) \\ &= p_i \widetilde{\zeta}_i y_i \\ &= p_i \frac{y_i}{x_i^0+y_i} \widetilde{\Lambda}_i^M(x_i^0+y_i). \end{aligned}$$

Summing over all routes, we see that

$$P_y(x^1+y) \geq \widetilde{P}_y(x^0+y),$$

Combining this with the bound we established for non-coalition drivers gives the desired result: $P(x^1+y) \geq \widetilde{P}(x^0+y)$.

G L-NCF Performance Guarantees (Demand)

Given an allocation of drivers x, define the demand served W and approximate demand served $\widetilde{W}$ as follows:

$$R(x) := \sum_i \Lambda_i^M(x_i),\ \widetilde{R}(x) := \sum_i \widetilde{\Lambda}_i^M(x_i),$$

Where $\widetilde{\Lambda}^M$ is as defined in Section 6.3. In this section, we will show that L-NCF obtains a welfare no worse than $\frac{1}{1+\gamma}\frac{1}{1+\gamma\frac{p_{\min}}{p_{\max}}}\frac{\max_i l_i}{\min_i l_i}$ times the maximum achievable demand served in Stackelberg Routing Informal Transit problem.

G.1 The L-NCF algorithm and its intermediate steps

There are a total of m drivers, α of which are coordinated by a driver union. Here is how the L-NCF algorithm works:

1. The coalition simulates how the $(1-\alpha)m$ non-compliant drivers would allocate themselves, x^0, in the absence of any coalition drivers if the profit function were $\widetilde{P}$. Namely, x^0 is a driver user equilibrium (with respect to the profit objective). All routes with a positive allocation of non-compliant drivers have the same profit per driver. Call this value π^0.

2. Upon observing x^0, the coalition picks an allocation y of its drivers to maximize demand served: $y \in \underset{y \succeq 0, \|y\|_1 = \alpha m}{\operatorname{argmax}} \widetilde{R}(x^0 + y)$.

3. Given the presence of the coalition y, the non-compliant drivers choose their allocation x^1, which is a user equilibrium (with respect to the profit objective) with respect to the true profit function P. Let π^1 be the resulting equilibrium profit per non-compliant driver.

G.2 Proof structure

The proof structure has a very similar structure to the proof for the profit objective. However, due to the misalignment between the rider objective (profit maximization) and system objective (maximize demand served), the performance guarantee achieved here is weaker.

Let y^* be an optimal Stackelberg strategy for the coalition under the objective function R. Let x^* be the corresponding response from the non-coalition drivers under profit function P.

First, we bound $R(x^* + y^*)$, i.e., the performance of y^*, by its performance under the approximated demand served function $\widetilde{R}$. Specifically, if $\widetilde{x}^*$ is the response from the non-coalition drivers to y^* under the profit function $\widetilde{P}$, then:

Step 1: We show that $R(x^* + y^*) \leq (1+\gamma)\widetilde{R}(x^* + y^*)$ and $\widetilde{R}(x^* + y^*) \leq (1 + \gamma\frac{p_{\max}}{p_{\min}})\widetilde{R}(\widetilde{x}^* + y^*)$. Here $p_{\max} := \max_i p_i$ and $p_{\min} := \min_i p_i$. Combining these inequalities gives $R(x^* + y^*) \leq (1+\gamma)(1 + \gamma\frac{p_{\max}}{p_{\min}})\widetilde{R}(\widetilde{x}^* + y^*)$.

Step 2: y^* is a feasible Stackelberg strategy under $\widetilde{R}$. The L-NCF algorithm achieves $\widetilde{R}(x^0 + y)$ riders served. We will show that, even when drivers are optimizing for profit and the platform is optimizing with respect to $\widetilde{R}$, NCF is an optimal algorithm for the Stackelberg routing problem with objective $\widetilde{R}$, hence

$$\begin{aligned} & \widetilde{R}(\widetilde{x}^* + y^*) \leq \widetilde{R}(x^0 + y) \\ \implies & R(x^* + y^*) \leq (1+\gamma)(1 + \gamma\frac{p_{\max}}{p_{\min}})\widetilde{R}(x^0 + y). \end{aligned}$$

Step 3: Finally, we show that NCF's utility under $\widetilde{R}$ transfers to R, namely: $\frac{\min_i l_i}{\max_i l_i}\widetilde{R}(x^0 + y) \leq R(x^1 + y)$.

Putting everything together, we see that NCF is guaranteed to obtain a $\frac{1}{(1+\gamma)(1+\gamma\frac{p_{\max}}{p_{\min}})}\frac{\min_i l_i}{\max_i l_i}$ fraction of the optimal demand served in the Stackelberg game:

$$W(x^*+y^*) \leq \frac{\max_i l_i}{\min_i l_i}(1+\gamma)(1+\gamma\frac{p_{\max}}{p_{\min}})W(x^1+y).$$

We will complete the proof by proving steps 1, 2 and 3.

G.3 Proof of Step 1

The inequality $R(x^*+y^*) \leq (1+\gamma)\widetilde{R}(x^*+y^*)$ is a direct consequence of (10). Thus we spend the rest of this section establishing $\widetilde{R}(x^*+y^*) \leq (1+\gamma\frac{p_{\max}}{p_{\min}})\widetilde{R}(\widetilde{x}^*+y^*)$.

Let y^* be an optimal Stackelberg strategy for the coalition under the objective function R. Let x^* be the corresponding response from the non-coalition drivers under profit function P. Let π_{eq} denote the profit per driver achieved by the non-coalition drivers in this equilibrium. As before, define the following indices based on the optimal solution x^*+y^*:

$$\begin{aligned}
E_T^* &:= \{i : x_i^* + y_i^* > k_i^*\} \\
E_P^* &:= \{i : x_i^* + y_i^* \in (0, k_i^*]\} \\
E_N^* &:= \{i : x_i^* + y_i^* = 0\}.
\end{aligned}$$

Note that if we change the profit function from P to $\widetilde{P}$, the profit per driver of routes in E_T^* remains unchanged, and the profit per driver of routes in E_P^* shrink by at most $\frac{1}{1+\gamma}$:

$$\begin{aligned}
\frac{p_i\widetilde{\Lambda}_i^M(x_i^*+y_i^*)}{x_i^*+y_i^*} &= \frac{p_i\Lambda_i^M(x_i^*+y_i^*)}{x_i^*+y_i^*} = \pi_{\text{eq}} \text{ for all } i \in E_T^*. \\
\frac{p_i\widetilde{\Lambda}_i^M(x_i^*+y_i^*)}{x_i^*+y_i^*} &\geq \frac{1}{1+\gamma}\frac{p_i\Lambda_i^M(x_i^*+y_i^*)}{x_i^*+y_i^*} = \frac{1}{1+\gamma}\pi_{\text{eq}} \text{ for all } i \in E_P^*.
\end{aligned}$$

Therefore, $\widetilde{x}^*$ is achieved by moving drivers from E_P^* to E_T^* until the profit per driver for all routes which have nonzero non-coalition drivers is the same. The equilibrium profit per driver (as measured by $\widetilde{P}$) in the equilibrium $\widetilde{x}^*+y^*$ will be denoted $\widetilde{\pi}_{\text{eq}}$, and

$$\widetilde{\pi}_{\text{eq}} \in \left[\frac{1}{1+\gamma}\pi_{\text{eq}}, \pi_{\text{eq}}\right].$$

This is because moving drivers from E_P^* to E_T^* improves the profit per driver on routes in the former set and reduces the profit per driver on the latter set.

To bound $\widetilde{R}(x^*+y^*)$ by $\widetilde{R}(\widetilde{x}^*+y^*)$, we need to understand how many drivers $\widetilde{x}^*$ moves from E_P^* to E_T^* (compared to x^*). To this end, note that $\widetilde{\pi}_{\text{eq}} \leq \pi_{\text{eq}}$ implies that $\widetilde{x}_i^* + y_i^* \geq x_i^* + y_i^* > k_i^*$ for any $i \in E_T^*$. Hence, for any $i \in E_T^*$ we have:

$$\begin{aligned}
&\frac{p_i\Lambda_i}{\widetilde{x}_i^*+y_i^*} = \widetilde{\pi}_{\text{eq}} \geq \frac{1}{1+\gamma}\pi_{\text{eq}} = \frac{p_i\Lambda_i}{x_i^*+y_i^*} \\
&\iff \widetilde{x}_i + y_i \leq (1+\gamma)(x_i^*+y_i^*)
\end{aligned}$$

The above inequality shows that

$$\sum_{i\in T^*} \widetilde{x}_i^* - x_i^* \leq \gamma \sum_{i\in T^*} x_i^* + y_i^* \tag{11}$$

With (11) in mind, note that:

$$\begin{aligned}
\frac{\widetilde{R}(\widetilde{x}^*+y^*)}{\widetilde{R}(x^*+y^*)} &= \frac{\sum_{i\in E_T^*}\Lambda_i^M(\widetilde{x}_i^*+y_i^*)+\sum_{i\in E_P^*}\Lambda_i^M(\widetilde{x}_i^*+y_i^*)+\sum_{i\in E_N^*}\Lambda_i^M(y_i^*)}{\sum_{i\in E_T^*}\Lambda_i^M(x_i^*+y_i^*)+\sum_{i\in E_P^*}\Lambda_i^M(x_i^*+y_i^*)+\sum_{i\in E_N^*}\Lambda_i^M(y_i^*)} \\
&\geq \frac{\sum_{i\in E_T^*}\Lambda_i^M(\widetilde{x}_i^*+y_i^*)+\sum_{i\in E_P^*}\Lambda_i^M(\widetilde{x}_i^*+y_i^*)}{\sum_{i\in E_T^*}\Lambda_i^M(x_i^*+y_i^*)+\sum_{i\in E_P^*}\Lambda_i^M(x_i^*+y_i^*)} \\
&\overset{(a)}{=} \frac{\sum_{i\in E_T^*}\Lambda_i+\sum_{i\in E_P^*}\widetilde{\zeta}_i(\widetilde{x}_i^*+y_i^*)}{\sum_{i\in E_T^*}\Lambda_i+\sum_{i\in E_P^*}\widetilde{\zeta}_i(x_i^*+y_i^*)} \\
&= \frac{\sum_{i\in E_T^*}\Lambda_i+\sum_{i\in E_P^*}\widetilde{\zeta}_i(\widetilde{x}_i^*+y_i^*)}{\sum_{i\in E_T^*}\Lambda_i+\sum_{i\in E_P^*}\widetilde{\zeta}_i(\widetilde{x}_i^*+y_i^*)+\sum_{i\in E_P^*}\widetilde{\zeta}_i(x_i^*-\widetilde{x}_i^*)} \\
&\geq \frac{\sum_{i\in E_T^*}\Lambda_i}{\sum_{i\in E_T^*}\Lambda_i+\sum_{i\in E_P^*}\widetilde{\zeta}_i(x_i^*-\widetilde{x}_i^*)} \\
&\geq \frac{\sum_{i\in E_T^*}\Lambda_i}{\sum_{i\in E_T^*}\Lambda_i+\left(\max_{i\in E_P^*}\widetilde{\zeta}_i\right)\sum_{i\in E_P^*}(x_i^*-\widetilde{x}_i^*)} \\
&\overset{(b)}{=} \frac{\sum_{i\in E_T^*}\Lambda_i}{\sum_{i\in E_T^*}\Lambda_i+\left(\max_{i\in E_P^*}\widetilde{\zeta}_i\right)\sum_{i\in E_T^*}(\widetilde{x}_i^*-x_i^*)}.
\end{aligned} \tag{12}$$

Here (a) is because $\widetilde{\Lambda}_i^M(x)=\widetilde{\zeta}_i x$ whenever $x\in[0,k_i^*]$, and is equal to Λ_i otherwise. (b) is because all drivers removed from E_P^* are sent to E_T^*, i.e., $\sum_{i\in E_T^*\cup E_P^*} x_i^* = \sum_{i\in E_T^*\cup E_P^*}\widetilde{x}^*$. To wrap up, recall that for any $j\in E_P^*$ and $i\in E_T^*$, we have:

$$\begin{aligned}
p_j\widetilde{\zeta}_j = \frac{p_j\widetilde{\Lambda}_j^M(x_j^*+y_j^*)}{x_j^*+y_j^*} &\leq \frac{p_j\Lambda_j^M(x_j^*+y_j^*)}{x_j^*+y_j^*} = \pi_{\text{eq}} = \frac{p_i\Lambda_i}{x_i^*+y_i^*} \\
&\implies x_i^*+y_i^* \leq \frac{p_i}{p_j}\frac{\Lambda_i}{\widetilde{\zeta}_j} \\
&\implies x_i^*+y_i^* \leq \frac{p_{\max}}{p_{\min}}\frac{\Lambda_i}{\widetilde{\zeta}_j}
\end{aligned}$$

Since this bound holds for all j, we have

$$x_i^*+y_i^* \leq \frac{p_{\max}}{p_{\min}}\frac{\Lambda_i}{\max_{j\in E_P^*}\widetilde{\zeta}_j}.$$

Applying this inequality to (11) gives

$$\sum_{i\in T^*}\widetilde{x}_i^* - x_i^* \leq \gamma\sum_{i\in T^*} x_i^*+y_i^* \leq \gamma\frac{p_{\max}}{p_{\min}}\frac{1}{\max_{j\in E_P^*}\widetilde{\zeta}_j}\sum_{i\in E_T^*}\Lambda_i.$$

Finally, applying this inequality to (12) gives

$$\begin{aligned}\frac{\widetilde{R}(\widetilde{x}^*+y^*)}{\widetilde{R}(x^*+y^*)} &\geq \frac{\sum_{i\in E_T^*}\Lambda_i}{\sum_{i\in E_T^*}\Lambda_i + \left(\max_{i\in E_P^*}\widetilde{\zeta}_i\right)\sum_{i\in E_T^*}(\widetilde{x}_i^*-x_i^*)} \\ &\geq \frac{\sum_{i\in E_T^*}\Lambda_i}{\sum_{i\in E_T^*}\Lambda_i + \left(\max_{i\in E_P^*}\widetilde{\zeta}_i\right)\gamma\frac{p_{\max}}{p_{\min}}\frac{1}{\max_{j\in E_P^*}\widetilde{\zeta}_j}\sum_{i\in E_T^*}\Lambda_i} \\ &= \frac{1}{1+\gamma\frac{p_{\max}}{p_{\min}}}\end{aligned}$$

which establishes the desired result.

G.4 Proof of Step 2

See Appendix E.

G.5 Proof of Step 3

For this analysis, we will again partition the routes into three sets, this time by the allocation x^0+y:

$$\begin{aligned}E_T &:= \left\{i : x_i^0+y_i > k_i^*\right\} \\ E_P &:= \left\{i : x_i^0+y_i \in (0,k_i^*]\right\} \\ E_N &:= \left\{i : x_i^0+y_i = 0\right\}.\end{aligned}$$

Since $\widetilde{\Lambda}_i^M$ is a lower bound for $\widetilde{\Lambda}_i^M$ which is tight on the set $\{0\}\cup[k_i^*,\infty)$, we have:

$$\begin{aligned}\frac{p_i\Lambda_i^M(x_i^0+y_i)}{x_i^0+y_i} &= \frac{p_i\widetilde{\Lambda}_i^M(x_i^0+y_i)}{x_i^0+y_i} = \pi^0 \text{ for all } i\in T^*. \\ \frac{p_i\Lambda_i^M(x_i^0+y_i)}{x_i^0+y_i} &\geq \frac{p_i\widetilde{\Lambda}_i^M(x_i^0+y_i)}{x_i^0+y_i} = \pi^0 \text{ for all } i\in P^*. \\ p_i\left(\Lambda_i^M\right)'(0) &\geq p_i\left(\widetilde{\Lambda}_i^M\right)'(0).\end{aligned}$$

Thus we see that non-coalition drivers on routes in E_P (and possibly some in E_N as well) are weakly better off than non-coalition drivers on routes in E_T, as measured by the profit function P. Therefore, the equilibrium non-coalition driver allocation x^1 will move drivers from E_T to $E_P\cup E_N$, and the resulting equilibrium profit per non-coalition driver π^1 will be at least as large as π^0. In particular, this means that $x_i^1+y_i\leq k_i^*$ for all $i\in E_P\cup E_N$.

Noting that:

$$\frac{\max_i\widetilde{\zeta}_i}{\min_j\widetilde{\zeta}_j} = \max_{i,j}\frac{F(t_2-t_1)}{2l_i}\div\frac{F(t_2-t_1)}{2l_j} = \max_{i,j}\frac{l_j}{l_i} = \frac{l_{\max}}{l_{\min}} \text{ where } l_{\max}=\max_i l_i, l_{\min}=\min_i l_i.$$

With this in hand, we have:

$$\begin{aligned}
\sum_i \Lambda_i^M(x_i^1 + y_i) &\geq \sum_i \widetilde{\Lambda}_i^M(x_i^1 + y_i) \\
&= \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) + \sum_{i \in E_P \cup E_N} \widetilde{\zeta}_i(x_i^1 - x_i^0) - \sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \\
&= \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) - \sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \left(1 - \frac{\sum_{i \in E_P \cup E_N} \widetilde{\zeta}_i(x_i^1 - x_i^0)}{\sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+}\right) \\
&\overset{(a)}{\geq} \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) - \sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \left(1 - \frac{\sum_{i \in E_P \cup E_N} \widetilde{\zeta}_i(x_i^1 - x_i^0)}{\sum_{i \in E_T} \widetilde{\zeta}_i(x_i^0 - x_i^1)_+}\right) \\
&\geq \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) - \sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \left(1 - \frac{\min_i \widetilde{\zeta}_i}{\max_j \widetilde{\zeta}_j}\right) \\
&= \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) - \sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \left(1 - \frac{l_{\min}}{l_{\max}}\right) \\
&\overset{(b)}{\geq} \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) - \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i) \left(1 - \frac{l_{\min}}{l_{\max}}\right) \\
&= \frac{l_{\min}}{l_{\max}} \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i)
\end{aligned}$$

where (a) is because $x_i^0 + y_i > k_i^*$ for $i \in E_T$, and (b) is because $\widetilde{\Lambda}_i^M$ is non-decreasing:

$$\sum_{i \in E_T} \widetilde{\zeta}_i(k_i^* - x_i^1 - y_i)_+ \leq \sum_{i \in E_T} \widetilde{\zeta}_i k_i^* = \sum_{i \in E_T} \widetilde{\Lambda}_i^M(x_i^0 + y_i) \leq \sum_i \widetilde{\Lambda}_i^M(x_i^0 + y_i).$$