On the Value of Dynamism in Transit Networks

J. Carlos Martínez Mori[a,*], Samitha Samaranayake[a], and M. Grazia Speranza[b]

[a] Cornell University, Ithaca NY, USA
jm2638@cornell.edu, samitha@cornell.edu
[b] University of Brescia, Brescia, Italy
grazia.speranza@unibs.it
[*] Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 4, 2022

Keywords: transit systems, on-demand mobility, network design, stochastic optimization

1 INTRODUCTION

The design of transit systems is a classical yet persistently challenging problem. Part of the difficulty stems from the complexity of the application domain; it has a lot of moving parts. These include the design of the underlying physical infrastructure network (e.g., the network of bus lanes), the design of the operational network (e.g., the set of bus lines), timetabling, crew and fleet scheduling, and so on. To add to this, there may be multiple objectives and/or quality of service targets, often without a clear mathematical description and in conflict with one another. As a result, there is no single "global" optimization problem, and if there were it would not be tractable. Rather, the problem is typically decomposed into a sequence of steps, starting from the bare bones—the design of the physical infrastructure network—and continuing on toward increasingly operational bells and whistles. Even when the process is decomposed, the optimization problems arising in each step are usually NP-hard (see Desaulniers & Hickman (2007) and Schöbel (2012) for an overview of the transit system design process).

The rise of on-demand mobility technologies over the past decade has sparked interest in the integration of traditional transit and on-demand systems—the number of *microtransit* (i.e., high-capacity on-demand shuttles) pilot programs conducted by transit agencies across the United States is a testament to this (see Westervelt *et al.* (2018) for a compilation of experiences). One of the main reasons behind this is the potential for microtransit to address a fundamental trade-off in transit: the *ridership versus coverage* dilemma. It is well-known that, given a limited budget, transit networks that maximize ridership and transit networks that maximize coverage (e.g., the geographical service area) tend to be vastly different (see Walker (2012) for a practitioner-oriented discussion). Intuitively, integrated systems may bridge this gap by letting each sub-system do what it does best: transit should focus on ridership, microtransit should extend coverage as a first/last mile service, and the two should be jointly optimized.

However, unlike purely fixed systems or purely on-demand systems, integrated systems are not well-understood. In particular, their planning and operational problems are significantly more challenging, and their broader implications are the source of a heated debate. Some transportation researchers and practitioners have suggested that on-demand systems can be complementary and benefitial to traditional transit (e.g., Shaheen & Chan (2016), Feigon & Murphy (2016), Hall *et al.* (2018), Alonso-González *et al.* (2018), Stiglic *et al.* (2018), Liu & Ouyang (2021)). At the same time, others have raised concerns about or even flat-out dismissed the supposed benefits (e.g., Walker (2012), Rayle *et al.* (2016), Walker (2018), Westervelt *et al.* (2018), Merlin (2019)).

Motivated by the debate we introduce the *dynamicity gap*, a general concept in multi-stage optimization. This concept quantifies the *attainable* benefit of allowing (but not requiring) dynamic components in the response strategy to a multi-stage optimization problem. We study the dynamicity gap within the context of designing transit infrastructure networks; the first step in the the transit system design process, and arguably the most decisive one since all subsequent steps depend on it. However, we note that the concept is more generally applicable in domains where goals can be met through a combination of static and dynamic (i.e, stage-specific) decisions.

2 DESIGN OF TRANSIT NETWORKS

We study the dynamicity gap within the design of integrated transit networks. To this end, we first describe the *Steiner forest problem*; the prototypical problem in network design. We are given a weighted graph $G = (V, E)$ and a set $I \subseteq V \times V$ of origin-destination pairs. The problem is to find a minimum cost subset $X \subseteq E$ of edges supporting a path between each origin-destination pair. We extend this abstraction to the design of *integrated networks* by allowing us to pick two subsets $X, Z \subseteq E$ of edges such that $X \cup Z$ supports a path between each origin-destination pair. In the context of transit, G represents the underlying topology (e.g., a road network), I represents the travel demand, and X and Z represent different "modes" of transportation, so that $X \cup Z$ represents an integrated multi-modal network. In particular, we treat X as a transit network and Z as a microtransit network. Going forward, we represent subsets $X, Z \subseteq E$ of edges by their characteristic vectors $x, z \in \{0, 1\}^m$, where $m = |E|$.

Admittedly, this abstraction ignores important operational aspects of transit. One can in principle enhance it with a number of transit-specific features and constraints. However, it is not our goal to provide a detailed operational model of transit networks. Indeed, producing and efficiently solving such models is a research area in and of itself. Instead, our focus is on fundamental abstractions that lend themselves to mathematical analysis and comprehensive experimentation while capturing the essence of the first step of the transit system design process: the design of the physical infrastructure network.

3 DYNAMICITY GAP

We consider a multi-stage version of the design of integrated transit networks, wherein the transit network is static but the microtransit network is dynamic. The temporal planning horizon is partitioned into $T \in \mathbb{N}$ stages indexed by $[T] \coloneqq \{1, 2, \ldots, T\}$. It is implicit that each stage has a (say uniform) duration $\delta > 0$; we elaborate on this shortly. Let $\mathcal{I} \coloneqq \{I_1, I_2, \ldots, I_k\} \subseteq 2^{V \times V}$ be the collection of possible travel demand realizations (i.e., the possible sets of origin-destination pairs). For each $t \in [T]$, let $I^t \in \mathcal{I}$ be the travel demand during the tth stage and $\mathcal{P}^t$ be the finite and non-empty set of integrated network configurations that can serve it. Let $x \in \{0, 1\}^m$ be decision variables corresponding to the static network (e.g., transit) paid for at cost $c_s \coloneqq c_s(\delta) \in \mathbb{R}^m_{\geq 0}$ on every stage. Let $z^1, z^2, \ldots, z^T \in \{0, 1\}^m$ be decision variables corresponding to the dynamic network (e.g., microtransit) over the stages, each paid for once at cost $c_d \coloneqq c_d(\delta) \in \mathbb{R}^m_{\geq 0}$. Then, the design of integrated transit networks can be posed as a multi-stage optimization problem of the form:

$$\begin{aligned} \min_{x, z^1, z^2, \ldots, z^T} \quad & \sum_{t=1}^{T} \left(c_s \cdot x + c_d \cdot z^t \right) \\ \text{s.t.} \quad & (x, z^t) \in \mathcal{P}^t, \qquad \forall t \in [T] \end{aligned} \tag{1}$$

The constraints $(x, z^t) \in \mathcal{P}^t$ for $t \in [T]$ ensure the static network x and the dynamic network z^t during the tth stage together serve the travel demand I^t during the tth stage.

We emphasize the general dependency of the per-stage costs $c_s \coloneqq c_s(\delta)$ and $c_d \coloneqq c_d(\delta)$ on the stage duration δ. If a system operates at a time-scale different from δ, it is crucial that its per-stage cost is pro-rated. Going forward, let $c \in \mathbb{R}^m_{\geq 0}$ be the cost of a single run of the static system and let $\eta \cdot c$ be the cost of a single run of the dynamic system, for some $\eta > 0$. The surcharge coefficient η captures the notion that static systems and dynamic systems have different operational costs on a per mile basis, *independent* of their frequencies. Let $\delta_d > 0$ be the batching interval of microtransit (e.g., 5 minutes). If the stage duration is tied to the microtransit batching interval (i.e., $\delta := \delta_d$), then the microtransit cost $c_d := \eta \cdot c$ is in the same scale as δ—no pro-rating is needed. However, transit may have a headway $\delta_s > 0$ independent of δ_d (e.g., 15 minutes). If so, we need to pro-rate its per-stage cost as $c_s := (\delta_d/\delta_s) \cdot c$.

Let OPT denote the cost of an optimal trade-off solution to (1) and OPT^Σ denote the cost of an optimal static solution to (1), that is one in which we additionally require $z^1 = z^2 = \cdots = z^T = 0$. We define the *dynamicity gap* α of (1) as the unit-less coefficient

$$\alpha := \frac{\text{OPT}^\Sigma}{\text{OPT}} \geq 1. \tag{2}$$

Large values of α indicate large gains from introducing dynamism.

4 RESULTS

The dynamicity gap α quantifies the value of dynamism, but computing it generally involves solving (1), which may be intractable. Moreover, we observe from (1) that α is influenced by implicit and explicit parameters such as the costs c_s and c_d, the stage duration δ, and the relationship between the sets of feasible configurations $\mathcal{P}^1, \mathcal{P}^2, \ldots, \mathcal{P}^T$. In transit network design, the latter depends on the underlying topology and on the sequence of travel demands.

The overarching goal of this work is to *parametrically* study the behavior of α without the need of solving the underlying multi-stage optimization problem. To this end, we initially assume $c_s := c$ for some $c \in \mathbb{R}^m_{\geq 0}$ and $c_d := \theta \cdot c$ for some relative cost coefficient $\theta > 0$, regardless of δ. That is, we initially restrict our study to problems of the form:

$$\begin{array}{lll} \displaystyle\min_{x, z^1, z^2, \ldots, z^T} & \displaystyle\sum_{t=1}^{T} \left(c \cdot x + \theta \cdot c \cdot z^t\right) & \\ \text{s.t.} & (x, z^t) \in \mathcal{P}^t, & \forall t \in [T] \end{array} \tag{3}$$

We first show that if the input scenarios $I^1, I^2, \ldots, I^T$ are sampled i.i.d. from a probability distribution $\mathcal{D}$ over $\mathcal{I}$, and moreover $T \to \infty$, then we can reformulate the horizon-normalized version of (3) as a two-stage stochastic optimization problem. This intuitive result is closely related to the convergence of the Sample Average Approximation (SAA) method shown by Kleywegt *et al.* (2002). As a corollary, in this case, the dynamicity gap α of (3) reduces to the *stochasticity gap* (see Bertsimas & Goyal (2010)) of the resulting two-stage stochastic problem.

This result allows us to treat the dynamicity gap $\alpha := \alpha(\theta)$ as a function $\alpha : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 1}$ of the relative cost coefficient θ that is implicitly parametrized by $\mathcal{D}$. In this way, our second contribution is a certificate of the value of dynamism (i.e., a certificate that $\alpha > 1$) whenever the relative cost coefficient does not exceed a certain value. Although this certificate is not tight in general, we show that producing it does not involve solving a two-stage stochastic problem.

Theorem 1. *Let $\theta^\dagger := \alpha(1)$ be the dynamicity gap when $\theta = 1$. We have $\theta^\dagger \leq \theta^*$, where $\theta^* := \arg\min_{\theta > 0}\{\alpha(\theta) = 1\} \geq 1$ is the critical relative cost coefficient after which dynamism is no longer valuable. More generally, for $\theta > 0$ we have*

$$\max\left\{\theta^\dagger/\theta, 1\right\} \leq \alpha(\theta).$$

In other words, if the relative cost coefficient θ is such that $\theta < \theta^\dagger$, then dynamism is certifiably valuable. We can strengthen this result to estimate α to any arbitrary precision, provided we solve a finite number of two-stage stochastic problems.

We tie our results back to transit with the following corollary.

Corollary 2. *Let $\delta := \delta_d$, $c_s := (\delta_d/\delta_s) \cdot c$, and $c_d := \eta \cdot c$; as described in Section 3. Then, for any fixed $\delta_s, \delta_d > 0$, we can transform (1) into a scaled version of (3) with $\theta = \eta \cdot (\delta_s/\delta_d)$ so that the condition $\theta < \theta^\dagger$ from Theorem 1 is equivalent to $\eta \cdot (\delta_s/\delta_d) < \theta^\dagger$.*

We view this as a quick, high-level "rule of thumb" giving a green light for the full-blown integrated transit system design process: given a microtransit batching interval $\delta_d > 0$ and the corresponding probability distribution $\mathcal{D}$ over $\mathcal{I}$ (e.g., by aggregating historical data into bins of size $\delta := \delta_d$), we set $\eta \cdot (\delta_s/\delta_d) = 1$ and (relatively) tractably compute $\theta^\dagger$. We then use δ_d and the computed $\theta^\dagger$ to certify the value of microtransit for certain combinations of transit frequency δ_s and surcharge coefficient η, namely whenever $\eta \cdot \delta_s < \theta^\dagger \cdot \delta_d$.

Higher values of $\theta^\dagger$ lead to a larger regime wherein dynamism is certifiably valuable. Therefore, we use $\theta^\dagger$ as a proxy for the value of dynamism and study how it is influenced by parameters implicit in (3). We conduct two sets of computational experiments.

1. Stylized Experiments. We conduct stylized experiments involving a multi-stage version of the Steiner tree problem (i.e., Steiner forest with a common root) on all unweighted, connected simple graphs on $2 \leq n \leq 7$ nodes and various distributions over input scenarios. We find:

- If demand arises at a node with probability proportional to its "centrality" (as quantified by closeness centrality Bavelas (1950)), dynamism tends to be more valuable on sparsely connected graphs. Tying our work back to transit, given that road networks are far from complete graphs, this supports the notion that microtransit may be beneficial when demand is concentrated on "central" areas but still appears sparingly on "peripheral" areas.
- If demand arises at a node with probability inversely proportional to its "centrality," dynamism tends to be more valuable on well-connected graphs. For example, if the graph is a tree, certain edges may be used repeatedly on most input scenarios, rendering dynamism unnecessary. The converse holds for complete graphs, where direct connectivity is possible.

2. Realistic Experiments. We conduct more realistic experiments using data from New York City and a multi-stage version of the Steiner forest problem as a proxy for the design of integrated transit networks. We focus on the area of Manhattan roughly south of the Flatiron Building using a crowd-sourced, distance-weighted graph G representing the road network. We take a subset[1] of weekday morning trip records from the Taxi and Limousine Commission throughout June 2016 and partition it by timestamp into uniform stages of duration $\delta := \delta_d$ for various choices of microtransit batching interval $\delta_d > 0$. To bring our abstraction closer to reality, and for tractability purposes, we impose detour constraints on pairwise connectivity: if $(s,t) \in I$ and the shortest-path length between s and t in G is $\ell(s,t)$, the shortest-path length between s and t in the solution must be less than $\Delta \cdot \ell(s,t)$ for some allowable detour factor $\Delta \geq 1$. We find:

- Dynamism tends to be slightly more valuable with a larger allowable detour factor Δ.
- We can produce quantitative results of the following style: if microtransit is batched for $\delta_d = 15$ minutes and passengers tolerate a $\Delta = 3/2$ factor detour, integration is worthwhile for a transit headway of $\delta_s = 15$ minutes and surcharge coefficients η less than ≈ 1.5.

5 CONCLUSION

Our goal with work is to provide a principled and (relatively) tractable analytical framework with which to study the value of dynamism, as quantified by the dynamicity gap. We hope this style of characterization enables accessible insight for both researchers and practitioners: given the problem at hand, leverage knowledge about the input parameters to quickly assess whether dynamism is worthwhile investment.

[1] We filter out trips starting or ending outside the area of study as well as trips shorter than 1000 meters.

References

Alonso-González, María J, Liu, Theo, Cats, Oded, Van Oort, Niels, & Hoogendoorn, Serge. 2018. The potential of demand-responsive transport as a complement to public transport: An assessment framework and an empirical evaluation. *Transportation Research Record*, **2672**(8), 879–889.

Bavelas, Alex. 1950. Communication patterns in task-oriented groups. *The journal of the acoustical society of America*, **22**(6), 725–730.

Bertsimas, Dimitris, & Goyal, Vineet. 2010. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, **35**(2), 284–305.

Desaulniers, Guy, & Hickman, Mark D. 2007. Public transit. *Handbooks in operations research and management science*, **14**, 69–127.

Feigon, Sharon, & Murphy, Colin. 2016. *Shared mobility and the transformation of public transit.* Transportation Research Board. TCRP Research Report, Issue Number: 188.

Hall, Jonathan D, Palsson, Craig, & Price, Joseph. 2018. Is Uber a substitute or complement for public transit? *Journal of Urban Economics*, **108**, 36–50.

Kleywegt, Anton J, Shapiro, Alexander, & Homem-de Mello, Tito. 2002. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, **12**(2), 479–502.

Liu, Yining, & Ouyang, Yanfeng. 2021. Mobility service design via joint optimization of transit networks and demand-responsive services. *Transportation Research Part B: Methodological*, **151**, 22–41.

Merlin, Louis A. 2019. Transportation sustainability follows from more people in fewer vehicles, not necessarily automation. *Journal of the American Planning Association*, **85**(4), 501–510.

Rayle, Lisa, Dai, Danielle, Chan, Nelson, Cervero, Robert, & Shaheen, Susan. 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, **45**, 168–178.

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

Shaheen, Susan, & Chan, Nelson. 2016. Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. *Built Environment*, **42**(4), 573–588.

Stiglic, Mitja, Agatz, Niels, Savelsbergh, Martin, & Gradisar, Mirko. 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research*, **90**, 12–21.

Walker, Jarrett. 2012. *Human transit: How clearer thinking about public transit can enrich our communities and our lives.* Island Press.

Walker, Jarrett. 2018 (February). *Microtransit: What I Think We Know.* Human Transit, https://humantransit.org/2018/02/microtransit-what-i-think-we-know.html. Accessed 2021-06-27.

Westervelt, Marla, Huang, Emma, Schank, Joshua, Borgman, Nolan, Fuhrer, Tamar, Peppard, Colin, & Narula-Woods, Rani. 2018 (January). *UpRouted: Exploring Microtransit in the United States.* Tech. rept. Eno Center for Transportation.