Probabilistic Models for Maximizing Service Area in Route Deviation Bus Transit Systems

Madhu Mausam Thapa[1], Ye Chen[2], Shijie Chen[3], Yanshuo Sun[3], Ilya O. Ryzhov[4], Nikola Marković[1]
[1]Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT, USA
[2]Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA, USA
[3]Department of Industrial and Manufacturing Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL, USA
[4]Robert H. Smith School of Business, University of Maryland, College Park, MD, USA
Corresponding author: Nikola Marković, `nikola.markovic@utah.edu`

Keywords: Semi-flexible Transit, Probabilistic Analysis, Optimization, Bayesian Learning

1 INTRODUCTION

Balancing the efficiency of fixed-route transit systems with the flexibility to meet varying passenger demands is a significant challenge in public transit operations, particularly in low-density areas. This paper focuses on the design and optimization of a route deviation bus transit system, which combines the strengths of both conventional fixed-route services and demand-responsive services. In this system, buses follow a fixed route and schedule in general, but can occasionally deviate within a specified buffer zone to accommodate on-demand requests, enhancing flexibility without severely impacting the reliability.

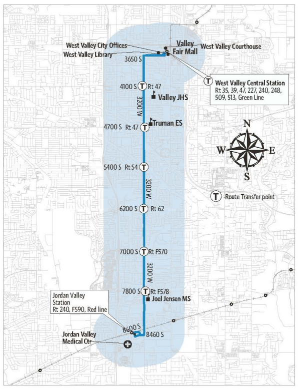


Figure 1 – *F232 Route*

Route deviation systems are widely used to address the limitations of conventional fixed-route services. The Utah Transit Authority (UTA) operates 15 such flex routes, which combine the reliability of fixed-route services with the flexibility of on-demand transport (see Figure 1). In these systems, buses follow a set route but can deviate up to 0.75 miles to accommodate on-demand requests. Passengers must schedule these on-demand requests at least two hours before travel, and only two deviations are allowed per bus roundtrip. Despite these adjustments, buses maintain their schedules and do not depart from designated checkpoints early.

The 0.75-mile deviation limit is based on ADA guidelines and serves as an *arbitrary* threshold balancing accessibility and operational feasibility. To enable a more principled system design, this paper develops a methodology to maximize buffer width while accounting for factors such as the utilization of fixed stops and the spatial distribution of on-demand requests for route deviations. Having underutilized fixed stops that can be skipped creates opportunities for larger deviations for on-demand requests, while heavily used fixed stops may require smaller buffer areas for on-demand requests to maintain the schedule. These factors underscore the need to optimize the buffer zone for maximum efficiency.

The key contributions include an integrated approach—consisting of an analytical probabilistic model and a simulation-based model with approximate Bayesian learning—to determine the *optimal* route-specific buffer width that (i) maintains on-time performance at key checkpoints with high probability and (ii) honors on-demand requests received with adequate notice. These models are applied to Utah routes, improving service efficiency and buffer size.

2 LITERATURE REVIEW

Analytical models have been key in optimizing semi-flexible transit systems with route deviation capabilities. Daganzo (1984) introduced models balancing stop spacing and slack time with passenger wait times and operational costs, while Chang & Schonfeld (1991) compared subscription bus services to traditional feeders, demonstrating cost-effectiveness in low to moderate-demand areas. Chien & Schonfeld (1997) focused on optimizing urban grid routes and stops, and Nourbakhsh & Ouyang (2012) proposed 'bus tubes' to improve coverage in low-demand areas. Errico *et al.* (2021) further refined demand-adaptive design under stationary conditions, using a hierarchical approach for service quality optimization. Recently, Lee *et al.* (2021) developed a two-stage stochastic programming model to incorporate spatial and volume variability, yet most of these models assume deterministic demand, limiting real-world applicability. Our research introduces probabilistic modeling to optimize buffer zones, enhancing adaptability under variable demand.

An advantage of simulation models is their ability to easily account for stochastic demand. Alshalalfah & Shalaby (2012) used discrete-event simulation to assess flexible services in Toronto, revealing potential cost savings, while Quadrifoglio *et al.* (2008) optimized scheduling for Mobility Allowance Shuttle Transit systems with passenger behavior data. Zheng & Li (2019) conducted simulation experiments on flex-route transit systems, examining system efficiency under varying degrees of dynamism to balance pre-booked and real-time requests. Although effective at handling uncertainty, these models can be computationally intensive for large systems. We address this challenge with an approximate Bayesian learning framework that reduces computational time by dynamically adjusting the number of simulation runs, improving overall efficiency.

While existing studies optimize transit at a network-wide level, our research advances this by optimizing buffer sizes for each route individually, making routes more passenger-centric and appealing for independent use. This approach enhances the serviceable area and ensures schedule adherence for on-demand requests made with adequate notice. By bridging deterministic analytical models and intensive simulations with probabilistic elements and Bayesian Learning, our method dynamically optimizes buffers and reduces computational time, providing a scalable solution for reliable semi-flexible transit services.

3 METHODOLOGY

This study proposes a dual-model approach to estimate the maximum allowable buffer width x that keeps the overall duration of the tour below an acceptable threshold with high probability while also honoring requests made on demand within that buffer. First, an analytical model provides insight under stylized settings. Second, a simulation-based model accommodates real-world geometry and demand patterns, with approximate Bayesian learning to reduce run times.

3.1 Analytical Model

The analytical model is developed in a stylized setting, assuming a straight main route as shown in Figure 2 with Manhattan distances used to compute deviations. This approach simplifies the problem, making the derivations more tractable while capturing the key dynamics of route deviation systems. We define a buffer width x allowing up to m on-demand deviations per trip.

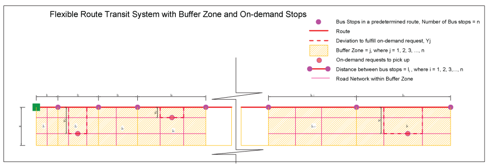


Figure 2 – *Stylized Model*

Let $T^m(x)$ be the total travel time, composed of fixed-route driving time, stochastic dwell times at fixed stops, and additional detour times for on-demand requests. In buffer zone j, each request arises with probability $q_j = x\, l_j\, \lambda_j$, where l_j is the length between the stops and λ_j is on-demand occurrence rate derived from population density and historical demand. The request incurs a detour of $\frac{2}{v} Y_j$, where v is the average travelling speed and $Y_j \sim \text{Uniform}(0, x)$, plus a dwell time s, provided the total number of deviations does not exceed m, a predetermined deviation limit according to policy guidelines.

We optimize the buffer width x while maintaining schedule reliability. The objective is to maximize x subject to constraints on total travel time and early arrivals:

$$\max_{x \in \mathbb{R}_+} \left\{ x \quad \textsf{s.t.} \quad \mathbb{P}\big(T^m(x) \leq c\big) \geq 1 - \beta, \quad \mathbb{P}\big(T^m(x) - \mathbb{E}[T^m(x)] \leq -\tau\big) \leq \alpha, \quad x\, l_j\, \lambda_j < 1, j = 1, \ldots, n-1 \right\}.$$

Here, c represents the time headway threshold. The first constraint ensures that the total bus tour duration $T^m(x)$ remains within c with at least probability $1-\beta$, thereby maintaining schedule reliability. The second constraint ensures that a request made τ minutes in advance of the expected arrival at a preceding stop is honored with high probability at least $1-\alpha$. Finally, the last constraint guarantees feasibility by ensuring that the probability q_j remains within a valid range.

To analyze the probabilistic constraints, we examine the asymptotic behavior of $T^m(x)$. Under the assumption that individual deviations are independent and contribute bounded random delays, the Lindeberg central limit theorem (CLT) applies. As n increases, the variance of $T^m(x)$ grows without any single term dominating, satisfying Lindeberg's condition. This ensures that $T^m(x)$ converges to a normal distribution for large n, allowing us to approximate constraints using normal probability bounds.

For small n, where the normal approximation may not be accurate, we use a one-sided Chebyshev inequality to provide a bound on the probability of early arrivals. Specifically, the likelihood that $T^m(x)$ falls below its expected value by at least τ is limited by the ratio of its variance to the sum of its variance and τ^2. This approach ensures probabilistic control over early arrivals, even when the distribution of $T^m(x)$ deviates from normality.

The resulting optimization problem yields a computationally tractable algorithmic solution to determine the optimal buffer width x that satisfies all constraints.

3.2 Simulation-Based Model

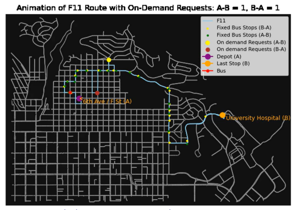
Our simulation-based model complements the analytical framework by mirroring real-world operating conditions and handling route geometries and travel patterns. Rather than imposing strong assumptions about travel times or spatial layouts, this approach takes advantage of real-world data to generate a more nuanced picture of daily operations. In each simulation run, the bus follows its scheduled route, whereas on-demand requests appear randomly in space according to demand densities estimated using population density and historical data. If a request lies within the current buffer width x and the bus has not reached its deviation limit, it is served; otherwise, the request is rejected.

To keep the process computationally manageable, we allow early termination when the tour is clearly infeasible due to exceeding tour durations or other reasons, avoiding unnecessary simulations. Simultaneously, we used approximate Bayesian learning to update uncertain parameters, such as travel speeds and dwell-time distributions, after each run, improving the fidelity of subsequent simulations. By pairing these simulations with a binary search over x, we iteratively identify the largest buffer width that ensures on-time completion with high probability. This data-driven approach balances accommodating more requests with maintaining schedule reliability while considering practical constraints like the maximum number of deviations per trip.

4 RESULTS

This study applies the methodology to optimize the Utah Transit Authority's (UTA) 15-route semi-flexible bus network, where buses can deviate up to 0.75 miles from the main route for on-demand requests. However, this flexibility can affect schedule reliability for downstream passengers. By optimizing the buffer width, this study aims to expand service coverage while maintaining reliable scheduling.

The optimized buffer zones demonstrate significant improvements in operational efficiency and service flexibility. The models dynamically adjust buffer widths based on demand, balancing service area coverage with schedule reliability. For instance, in high-demand routes, narrower buffer zones reduce delays from deviations, helping maintain punctuality. Conversely, in lower-demand routes, wider buffer zones are feasible, allowing buses to serve additional passengers without compromising the schedule. This adaptive approach enhances coverage and accessibility, particularly in areas under-served by traditional transit.



https://youtu.be/RjFoWKsH8Oo.

For instance, on the F11 flex route, we get optimized buffer width of 0.35 miles, significantly less than the original 0.75 miles allowance, but all stops maintain 100% on-time performance. This reduction highlights the trade-off between flexibility and reliability.

At University Hospital (06:34 AM scheduled time), the optimized time window ensures arrivals between 06:31 AM and 06:37 AM. Similarly, 9th Ave / LDS Hospital (WB) at 06:46 AM has a buffered arrival window of 06:36 AM to 06:47 AM, demonstrating that the model prevents deviation-induced delays while maximizing coverage.

To illustrate the effect of optimized buffer zones, we provide a video simulation of the F11 route showing bus deviations in response to on-demand requests. The video demonstrates how the bus deviates within the allowed buffer to accommodate passenger requests while following the primary route. It highlights the extent and location of deviations along the route.

5 CONCLUSION

This study presents a framework for optimizing route deviation systems by integrating analytical modeling with adaptive simulation techniques. The approach provides practical tools for transit agencies to enhance flexibility while maintaining schedule reliability, making it especially valuable in low-demand areas. The analytical model allows quick assessments, highlighting how parameters like demand density and headway affect service coverage, while the Bayesian simulation-based method enables precise optimization under diverse real-world scenarios.

Applied to UTA's flex routes, the framework demonstrates significant improvements in maintaining schedule adherence and expanding service reach. By dynamically adjusting buffer zones, the model balances accessibility in low-demand areas with consistent on-time performance. Wider buffer zones effectively expand service coverage without compromising punctuality, while high-demand routes benefit from tighter buffers to avoid delays. The results underscore the model's scalability and effectiveness in offering transit systems a balance of flexibility and reliability, enhancing both passenger accessibility and operational efficiency across varied transit systems.

References

Alshalalfah, B, & Shalaby, A. 2012. Feasibility of flex-route as a feeder transit service to rail stations in the suburbs: Case study in Toronto. *Journal of Urban Planning and Development*, **138**(1), 90–100.

Chang, Shyue Koong, & Schonfeld, Paul M. 1991. Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science*, **25**(4), 281–298.

Chien, Steven, & Schonfeld, Paul. 1997. Optimization of grid transit system in heterogeneous urban environment. *Journal of Transportation Engineering*, **123**(1), 28–36.

Daganzo, Carlos F. 1984. Checkpoint dial-a-ride systems. *Transportation Research Part B: Methodological*, **18**(4-5), 315–327.

Errico, Fausto, Crainic, Teodor Gabriel, Malucelli, Federico, & Nonato, Maddalena. 2021. The single-line design problem for demand-adaptive transit systems: A modeling framework and decomposition approach for the stationary-demand case. *Transportation Science*, **55**(6), 1300–1321.

Lee, Enoch, Cen, Xuekai, Lo, Hong K., & Ng, Ka Fai. 2021. Designing Zonal-Based Flexible Bus Services Under Stochastic Demand. *Transportation Science*, **55**(6), 1280–1299.

Nourbakhsh, Seyed Mohammad, & Ouyang, Yanfeng. 2012. A structured flexible transit system for low demand areas. *Transportation Research Part B: Methodological*, **46**(1), 204–216.

Quadrifoglio, Luca, Dessouky, Maged M., & Ordóñez, Fernando. 2008. Mobility allowance shuttle transit (MAST) services: MIP formulation and strengthening with logic constraints. *European Journal of Operational Research*, **185**(2), 481–494.

Zheng, Yue, & Li, Wenquan. 2019. Flex-route transit service with different degrees of dynamism. *Pages 4369–4378 of: Proceedings of the 19th COTA International Conference of Transportation Professionals*. American Society of Civil Engineers.