Transforming Bus Service Planning Using Integrated Electronic Data Sources at NYC Transit

Jeffrey Hanft, Shrisan Iyer, Brian Levine, and Alla Reddy
New York City Transit

Abstract

The installation of an Automatic Vehicle Location (AVL) system alongside existing Automated Fare Collection (AFC) data spurred development of an inferred bus boarding and alighting ridership model at New York City Transit (NYCT), allowing for 100% passenger origin-destination (O-D) data citywide. Analysis techniques that relied primarily on professional judgment due to lack of data were replaced by more sophisticated statistical techniques. This paper describes two case studies and the resulting service planning potential from having access to fully-integrated big data sources: a neighborhood-wide analysis of performance and ridership, where 100% data allowed planners to pinpoint specific, low-cost reroutes and stop changes to better serve riders, and identification of an optimal route split location for a long route with poor performance by minimizing passenger impact using modeled O-D data. In both examples, new data sources allowed for novel analysis throughout problem investigation as well as forecasting ridership and cost impacts of proposed service adjustments. As the agency’s ability to leverage these data improves, it will support Title VI obligations as well as performance monitoring.

Key Words: Automatic vehicle location, AVL, automated fare collection, AFC, transit performance, O-D data, ridership forecasting, service adjustments

Introduction

MTA (Metropolitan Transportation Authority), New York City Transit (NYCT), and the MTA Bus Company operate 316 routes across the 5 boroughs of New York City (NYC). The transit network operates 24 hours per day and serves almost 3 million bus riders and 6 million subway riders daily. Scheduling, headway determination, and route planning historically have been conducted using sampled data collected by traffic surveyors whose primary role is to collect boarding and alighting for bus trips as well as timings for schedule-making and performance measurement. The rollout of the MTA
Bus Time Automatic Vehicle Location (AVL) system across NYC allows for continuous monitoring of performance and running time for every route, with data approaching 100% completeness. Furthermore, by integrating Automated Fare Collection (AFC) data with AVL data, an inference model that produces passenger boarding and alighting locations was developed, yielding trip level loads and passenger origin-destination (O-D) data at both the trip and neighborhood levels on single-mode and inter-modal trips.

Due to budgetary constraints with traffic surveyors, MTA Board guidelines required ridership and running time analysis of weekday bus service every two years, with weekend service checked every four years. Similarly, for operating performance reporting, only 42 key routes were analyzed using a stratified sample. This yielded limited amounts of data to be used as input into the service planning process and high reliance on professional judgment and community input. Now, with AFC and AVL data collected daily for every route, the problem has shifted to determining how to use these data to make sound decisions based on quantitative evidence. This paper discusses the resulting service planning potential from having access to fully-integrated, rich, big data sources. This planning potential is contingent on the availability of high-quality, validated ridership data sources. The specific ridership data sources were developed in-house by Zeng et al. (2015) at NYCT. The development and validation of these data sources are significant efforts and have been detailed in separate papers and, therefore, are not in the scope of this paper. The focus is on two case studies that illustrate how these data sources were combined to analyze the following topics: route performance, running time, dispatcher-initiated service changes, boarding and alighting locations, average passenger trip length, passenger transfers, passenger type analysis, O-D patterns, and ridership impacts. The first case-study describes a neighborhood-wide bus service analysis in the Co-op City section of the Bronx, and the second details the passenger-optimal re-design of a Manhattan bus route with performance problems.

Background

The in-house-developed MTA Bus Time AVL system was installed gradually on the fleet of more than 6,000 buses between 2012 and 2014. The system’s primary function is a customer information system, but from the project’s inception, Operations Planning has been involved in testing and using AVL data for internal planning purposes. Data were used primarily to track vehicle locations and match actual movements to schedule to report On-Time Performance (OTP) and Wait Assessment, NYCT’s main publicly-reported headway-based performance indicator (a measure of the number of headways that pass an acceptable waiting threshold) defines on-time as arriving within -1 to +5 minutes of schedule time, and a headway passing wait assessment is within +3 (peak) or +5 (off-peak) of scheduled headway (Cramer et al. 2009). Schedule matching was accomplished via several algorithms, tying 30-second reported AVL data to scheduled times at timepoint locations along the route and later refined to include depot pull-in/pull-out information, bus status, and route dispatcher-initiated service changes from other sources to report 100% data on all buses with a high degree of accuracy (Levine et al. 2014). AVL data replaced a system of manual traffic surveys, providing sample data for performance reporting conducted on just 42 key routes.
Following the success of matching AVL data to schedule, these data were used for running-time calculations to feed the schedule-making process. This greatly improved the accuracy of schedules while providing a mechanism to analyze bus running times associated with: stop and route modifications, temporal traffic variability, roadway construction, bus lane implementations, traffic signal timing, special events, and shuttle operations.

NYCT took significant steps in expanding usage of AVL data by developing a model connecting AVL with AFC data to estimate boarding and alighting locations of all passengers riding buses (Zeng et al. 2015). The model relies on matching payment time and bus position provided every 30 seconds from AVL data to locate customer boardings. An algorithm was developed to determine customers’ transfer or alighting locations from subsequent swipes of the same MetroCard. The tap-on only structure of the NYCT fare payment systems requires that alighting locations be inferred from subsequent farecard activity; thus, there is no guarantee that the modeled journey is precisely the one actually made. The small minority of ridership that uses cash and other non-MetroCard payment methods are assigned travel distributions that mirror that larger MetroCard population. Although this breakthrough provided O-D and route-choice information for nearly three million passenger bus trips on a daily basis, this ridership detail is best suited for higher-level analysis of ridership patterns. Other agencies and academic researchers have developed similar methodologies for estimating bus passengers’ boarding and alighting locations from AVL and AFC data (Munizaga et al. 2011; Alsger et al. 2015; Lu et al. 2011; Reddy et al. 2009). This breakthrough provided O-D and route-choice information for nearly three million passenger bus trips on a daily basis. Ridership information inferred from the integrated AFC and AVL model was validated against manually-collected ride check data, which was the previous source for scheduling and planning ridership information. An in-house reporting server was enabled to automate comparisons of load profiles reported by each data source to ensure that automated model results were credible. This further enabled simple identification of when bus route model results that deviated from manual checks to determine if algorithm adjustments were warranted, while also acknowledging that manual data sources were also subject to uncertainty. Greater detail surrounding the development and validation of the boarding and alighting model used for NYC buses can be found in Zeng et al. (2015).

Transit agencies deriving similar information with Automated Passenger Count (APC) data are able to use those data to support numerous planning and operational objectives (Hammerle et al. 2005). Boarding and alighting data developed from AFC models are derived from passenger-specific travel patterns and can be connected with transfers to other bus services or subways. APC data do not connect boardings with alightings and require an iterative fit process or other approach to gather O-D detail (Furth et al. 2005; Mishalani and McCord 2013).

Other agencies have leveraged boarding and alighting information for improved understanding of passenger usage within a transit network (Gokasar et al. 2015). NYCT began using these data for similar purposes along with incorporating them into the entire planning process, beginning with problem investigation through forecasting.
ridership and cost implications of proposed service adjustments. As big data sources and boarding and alighting inference models become more commonplace at transit agencies, there is significant potential for peer agencies to use them as valuable tools in planning projects. By using multi-day averages to estimate typical ridership, the information obtained is more current and no longer subject to single-day ridership variations due to weather, local events, and road disruptions. Multi-day results are more representative of the total population of riders that use the system over the collection period, thus improving the statistical validity of any subsequent analyses performed with the output data. Increased data sophistication allows for new planning approaches at NYCT, ranging from route modifications for operational improvement to neighborhood-wide studies of customer travel patterns.

**Neighborhood Bus Service Planning Using AVL and AFC Data**

With complete operational and ridership data available for all routes, the assessment of bus performance and potential improvements to service supplying the needs of an entire neighborhood could be evaluated. With manual surveys, complete data at this scale and temporal consistency were not previously feasible, prohibiting investigation of neighborhood-level service as well as comparability across routes. Integration of AFC and AVL data not only allowed for on-board ridership estimation, but a deep dive into AFC data also enabled tracking of passenger journeys across multiple unlinked trips. This is dependent upon consistent passenger use of the same MetroCard, which includes a unique serial number in the transaction record. The fraction of passenger trips for which fare payment is cash is known and used to scale-up results that are dependent on electronic media only. To serve the needs of residents of the community, full journey patterns were tracked based on anonymized AFC data with estimated boarding and alighting locations.

**Case Study: Co-op City Neighborhood Analysis**

Co-op City is a neighborhood of high-rise apartments in eastern Bronx with a population of about 35,000 people and consisting of 5 residential sections, 3 shopping centers, and additional commercial facilities throughout. It has a mix of students, workers, families and retirees, and a large number of older adults live in the residential towers but account for just under 20% of the population. There are three distinct, contiguous areas; approximately 75% of the population lives in the northern part, which is encircled by a large loop of primary roadways and accessed by several smaller loop, and the remaining 25% of the population lives in the southeast corner of the neighborhood. The Bay Plaza Shopping Center, a large regional shopping center, occupies the area between the two major residential areas. In 2010, as a result of a significant shortfall in the MTA budget, a comprehensive cost reduction program, including staff reductions, renegotiation of contracts, and service cuts, was undertaken. Changes in Co-op City were substantial and included re-routing several routes and discontinuing one. The service changes reduced operating costs; however, they also provided more direct access to traffic generators outside of Co-op City (MTA NYCT 2011, 2014). In the years following cuts, community dissatisfaction with bus service was voiced. The 2013 Co-op City bus map is presented as Figure 1.
At the request of community leaders, a complete bus service planning study for Co-op City was undertaken in 2013 to evaluate bus service in the neighborhood. The timing of the study coincided with completion of NYCT’s bus ridership estimation model, allowing for these data to be used for planning and route change analysis for the first time (MTA NYCT 2014). A key study goal was to evaluate current service on 100% ridership data for Co-op City, as well as to track journeys of passengers to/from Co-op City. Automated ridership data, comprising about 17,500 average weekday records for an entire month, were used in the study alongside a passenger interview survey with about 1,300 respondents conducted over several days. Although vocal community members had clarified the bus service they felt would best serve Co-op City, complete ridership data revealed potential service that would benefit the entire community. Previous service planning managed through community complaints or

FIGURE 1. 2013 Co-op City bus service map
requests potentially could result in service that is advantageous to a small population but not the greatest number of riders. Planning with 100% ridership data presents a more complete and unbiased reality of the existing ridership needs and, by extension, community needs. Costly and timely stop-intercept surveys have a low sample rate, as well as selection bias of respondents (Doxsey 1984). Planning initiatives also may change substantially from original to final proposals, making surveys a poor source of data.

The study objectives were organized into several broad categories:

- Analyze whether existing service was provided as scheduled and assess the quality of service being provided.
- Determine if bus routing within Co-op City provided sufficient intra-neighborhood travel options.
- Study bus boardings and alightings to identify key stops and areas that were under or overserved.
- Analyze O-D patterns of residents and visitors to determine if service was designed optimally to meet their needs and whether a high number of passengers were forced to make several transfers to reach their destinations.

AVL data revealed that bus service in Co-op City was provided at or above service standards within the Bronx. By collecting a statistically-significant number of running time records via an AVL system, the mean as well as 15th and 85th percentiles of actual running times for routes could be compared for every time of day with what was provided in the schedule. Some routes were identified for inaccurate running times by comparing with averages and ranges of the actuals and adjusted in the next schedule.

To study ridership patterns of Co-op City residents and visitors, only passengers boarding or alighting within Co-op City boundaries were considered. This subset was further split into residents and visitors by determining if the first fare payment of the day was made within Co-op City boundaries (classified as a resident) versus elsewhere in the system (classified as a visitor). O-D pairs of residents helped determine the most important areas to serve. Figure 2a, visualized by the widths of the travel arcs presented, shows the frequencies of O-D pairs for destinations outside of Co-op City, and Figure 2b shows frequencies for travel within Co-op City. Visualizing major O-D links made it clear that the primary function of transit in this neighborhood is delivering residents to destinations outside the neighborhood, functioning as feeders to subways. By tracking passenger journeys, it was estimated that 92% of all passenger trips beginning in Co-op City were bound for destinations outside Co-op City. From morning rush through evening rush, there were more than 10,000 bus–subway and subway–bus transfers coming from Co-op City, highlighting the significance of subways as traffic generators for Co-op City and the importance of providing connectivity via the bus network. This finding, corroborated by survey results, was valuable, as it departed from community feedback that indicated that a route providing better service within the neighborhood was the primary need (Cruz 2013). This highlighted the benefit of electronic data to better communicate actual needs, and the matching of O-D pairs provided from survey results to estimations provided by the ridership model highlighted the ability of the data to provide representations of ridership and service patterns.
FIGURE 2. Origin-destination arcs in and around Co-op City and proposed re-route
Although most Co-op City passenger trips had O-Ds outside the neighborhood, trips made within Co-op City were examined to determine if improvements on behalf of the community were possible. A pattern of trips from the south end of the neighborhood towards Asch Loop (in the center of Co-op City), proximate to several commercial areas such as the Bartow and Bay Plaza malls (Figure 2b), emerges from the data, which would not have been uncovered with manual ridecheck data, as they lack O-D components. A minor service revision resulted, with six more stops along four routes, including the re-route of one route through the loop to better serve demand highlighted by this data (Figure 2c).

**Inter and Intra-modal O-D and Transfer Analysis**

O-D patterns of riders leaving Co-op City were studied to determine if service provided matched observed journeys, and particularly if they were served under a single fare by NYCT’s fare policy. Current transfer policy allows one free bus-bus or bus-subway transfer within two hours of initial fare payment, and community complaints indicated that many riders were negatively affected by service changes that introduced double-transfer trips. Although the ridership estimation model makes use of subsequent AFC transactions to determine a passenger’s boarding and alighting stops on a single trip, tracking a user’s transactions throughout the two-hour transfer window can determine true O-Ds of a continuous journey with potentially multiple modes. Tracking journeys identified that only 8% of Co-op City residents’ journeys began and ended within Co-op City, another 55% were bound for Bronx destinations outside of Co-op City, 35% were bound for other boroughs, and the remainder were bound for neighboring counties (Figure 3a). This analysis illustrated that many residents were undertaking very long trips, often with transfers. Since a large proportion of weekday morning peak commuting trips were bound for subway lines close to the neighborhood (1, 2, 4, 5, 6, A, D), bus-subway transfers were analyzed (Figure 3b).

The analysis found that new double transfer trips created by the 2010 service changes impacted fewer than 50 riders. These journeys could have been satisfied by a single transfer and a very moderate increase in walking distance. The analysis also showed that customers originating or destined to stops along Dreiser Loop in northern Co-op City were using a transfer with the Q50 bus, which did not serve Dreiser Loop. Customers were making an extra transfer to go to/from the Pelham Bay Park subway station or other destinations in the Bronx or Queens served by the Q50. In response, a new Q50 stop was created at Dreiser Loop to serve those customers on a single fare and without a transfer.
FIGURE 3.
Co-op City borough destinations and weekday morning bus to subway transfers
Positive Reception and Future Neighborhood Analyses

Service changes were communicated to the community with the expectation that service would improve within Co-op City (Rocchio 2014). The positive reception to the conclusions of the Co-op City study generated interest in extending this process to other NYC areas with even larger geographic and ridership scopes (Rivoli 2014). Expanding to an area of Northeast Queens with more than 35 bus routes, multiple subway connections, and daily ridership in excess of 300,000, both challenged the processes used in Co-op City and highlighted the benefit of big data sources. Passenger trip detail was quickly taken from individual passenger and route-level detail to macro detail for evaluating the quality of service in large areas. Figure 4 shows inter-neighborhood O-D arcs in excess of 1,000 average daily weekday riders. Neighborhood boundaries were defined by the NYC Department of City Planning (DCP), and stop-level ridership was aggregated to larger areas via geographic information systems (GIS). The significant travel subway stations generate confirmed the expectation that providing subway feeder service in this region is one of the most important objectives for NYCT to meet in this area.

FIGURE 4. Significant Northeast Queens neighborhood O-D arcs
NYCT’s ability to analyze big data has permitted for scaling up the size of such studies, along with the ability to be more responsive to community complaints regarding service. Future plans will expand the scope of neighborhood bus studies to the entire borough of Staten Island (Barone 2015).

**Single Route and Redesign and Analysis Using AVL and AFC Data**

Along with the success of the Co-op City study, a single route in Manhattan, the M5, was analyzed at a more micro level of detail. The M5 bus route runs from the George Washington Bridge Terminal (Broadway and W 178 St) in Upper Manhattan to South Ferry in Lower Manhattan. Average weekday ridership exceeds 13,000 riders and is one of the longest bus routes in Manhattan (12 miles). The M5 operates limited-stop service on weekdays and local service overnights and weekends. The same service cuts that resulted in changes to Co-op City in 2010 included discontinuing service on one southern Manhattan route, M6, and extending the M5 to the southern end of Manhattan to cover the gap in service. Figure 5 shows the route profile and location of original Houston Street terminal with average weekday estimated boarding and alightings.

![M5 Ridership profile with route path](image)

**FIGURE 5.** M5 Ridership profile with route path

After the M5 was extended to South Ferry in June 2010, the Department of Buses’ Manhattan Road Operations reported that the route experienced significant delays. Additionally, there has been an increasing number of concerns from the community and elected officials about lateness, gaps in service, and the number of buses being short-turned at various points along the route. OTP issues would have caused surface line dispatchers (SLDs) to short-turn certain bus trips to provide service in the opposite
direction, thus not serving certain stops with scheduled frequency. Operational
difficulties and community concerns identified the M5 as a potential candidate to
explore the novel ways in which service could be redesigned using big data.

**Performance Reporting**
After community complaints regarding service quality had been raised on the M5,
they were confirmed using AVL data. The existing performance-measuring framework
developed in tandem with the AVL rollout at NYCT confirmed that the M5 was
consistently one of the lowest-performing buses in the network. Over an entire month,
en-route OTP at pre-determined timepoints was below 50% for limited (rush-hour and
midday) service. Wait Assessment at these stops was just over 70%. At Houston Street,
buses were late 53.5% of the time in southbound direction, and at 9 AM were late over
70% of the time. Poor performance drove the decisions to short-turn buses to maintain
service in the opposite direction by the SLDs.

**Service Changes**
A majority of the real-time service adjustments made by SLDs to improve service
reliability are short turns, dark-to movements (beginning service a later stop than the
schedule origin), or skipping stops through part of the bus trip. These actions are done
to address lateness and maintain even spacing, as problems arise frequently and tend to
persist between trips. The service changes reported by passengers were also validated
through a novel application of AVL records. By tracing the movement of a specific bus
trip along its path and identifying cases when arrivals at expected stops were absent
for the remainder of the trip, the dispatcher intervention to short-turn a bus and the
approximate location at which it happened could be identified. The associated dark-to
movements in the opposite direction of short turns were also identified through this
process. An electronic booking system (EBS) exists for manually-recording service
changes such as short turns and records of other unplanned events. However, the most
important information from this system would be expected to be logged during periods
of greatest system duress and is then dependent on consistency of input by hundreds
of dispatchers during the busiest periods of managing bus service. When cross-checking
the inferred service changes from AVL data with EBS records, it was found that AVL
data provided a more reliable and more thorough source of service interventions. These
data are also consistently queriable and comparable across a large number of records,
which is not true of manually-entered data.

This process identified that over the course of four months, 9.8% of weekday
southbound scheduled trips on the M5 were short-turned at some point along the
course of their journey. An unequal distribution of service interventions by time of day
led to the 8 AM service hour having 18.2% of scheduled trips short-turned. Although
an external system for tracking service changes existed, the more comprehensive AVL
system discovered that there were almost one-third more service changes than initially
thought. In this hour, the combination of short turns and low operational performance
would lead to a significantly higher headway experienced by passengers than is
scheduled for this service. In the absence of the processed AVL data, NYCT would have
underestimated the extent of the performance problems that resulted from short turns.
This process proved that passengers voiced credible concerns.
Boarding and Alighting Estimation
To address passenger concerns on service changes and gaps in service, ridership data inferred from AFC and AVL data were used to guide the planning process. Due to the length of the M5, thorough checks of ridership for every stop on the route formerly were cost-prohibitive, even for a single day. Point checks at a limited number of stops cost-effectively collect the requisite data for service planning, but do not provide the same level of accuracy. Automated ridership model data quickly provided an overview of route usage and identified important service areas, as shown in Figure 5. Distinct maximum load points in different sections of the route, as well as areas of high turnover, highlighted areas of important service delivery and also served to confirm the appropriateness of the short-turn locations in the southbound direction.

Repeating an analysis procedure done for Co-op City, the most popular O-D pairs were identified. Although the M5 traverses most of Manhattan, 75% of the ridership was found to be making trips that extended no more than two neighborhoods beyond the boarding location. Neighborhoods were defined by the DCP and grouped or divided as appropriate so they contained approximately the same number of stops, about 20–25 blocks long. With specific boarding and alighting information available, average passenger trip length could be estimated easily. It was found that the passenger trip length distribution skewed heavily towards shorter trips, with the most common trip length 0.4 miles and the average trip length 2.2 miles. The ratio of average trip length to total route length for the M5 was 0.179, meaning an average rider uses only 18% of the route. The typical ratio for Manhattan routes is 0.27 and is 0.28 for all NYC. The M5 ratio was the fifth smallest of the 38 Manhattan routes, implying that the M5 length may be out of proportion, but also that other Manhattan routes may benefit from route rationalization. From this, it was determined that the M5 could be split while still satisfying most rider needs.

Passenger Types
AFC data capture the nature of the farecard purchased (e.g., pay per ride, 7-day unlimited, 30-day unlimited, etc.) along with details about discounted fares (e.g., senior, disabled) and are included in ridership data. This allows for targeted analysis of older adult riders who may require specialized bus service. This can also be extended to understanding the nature of different riders such as tourists or infrequent riders using single rides as opposed to unlimited rides.

Although a large majority of M5 riders took short trips, there were some long trips that traversed a majority of the route despite the fact that the M5 has a maximum run time of almost two hours and is redundant with the faster subway for most of its path. Analyzing rider behavior based on the few attributes present in AFC data is another improvement enabled by automated data sources. It was hypothesized that older adult or disabled riders may prefer or require a bus trip due to a preference to avoid stairs and longer access distance to underground subways. Analyzing riders with senior or disabled discounted fare structures found a slightly larger average trip distance was found for these riders, but at only 0.1 mile longer than the average population, it was determined that these riders trips were similar with the general population and would also benefit from a rationalized M5 route. This hypothesis could be disproved with minor data
analysis efforts and would have formerly remained unanswerable or require timely and costly surveys to determine the hypothesis validity.

**Passenger Transfers**

AFC data reveal that approximately one-quarter of the ridership of the M5 transfers to or from the subway as part of their journey. In the part of NYC that the M5 operates, local subway service is spaced about every 10 blocks or closer and almost always is parallel to the direction of them M5. These are presumably journeys that could be replaced easily by walking, but free transfers permit a faster or more comfortable journey to the subway via bus. This may explain why the mode for travel distance is only 0.4 miles.

The most common bus-subway transfer points for the M5 accounted for less than 8% of all bus-subway transfers for the route, implying that subway feeder passenger trips are spread throughout the route with no one station dominant in bus-subway transfers, as indicated in Figure 6a. In the case of the M5, understanding the geographic distribution of subway transfers does not change the planning decisions inferred from single-trip data, although it does serve to explain M5 ridership mode choices and trip lengths given other transit options available.

![FIGURE 6. M5 subway transfers and optimal split point](image-url)
Route Split Approach
To address operational issues associated with running a long route through a city environment, splitting the route was proposed. With complete boarding and alighting trip information for each passenger, the negative impacts of a route split could be quantified by determining those passengers that would need to transfer to complete their journey (induced transfers). Although boarding counts and average loading may help guide a split point, this fails to consider the full journeys of passengers and does not minimize the number of induced transfers.

A short shuttle route with few stops may achieve high performance, but would not serve passengers in a practical manner. An optimal solution could be approximated from these data in the absence of other operational considerations such as depot location, operational feasibility, existence of layover space, etc., by setting an objective function to minimize the number of riders that would require a transfer under the new route design. A non-linear Excel solver was used to iterate through each set of potential new terminals, with stops defined by their relative stop order. The nature of the solver used does not guarantee that a mathematical minimum is reached, although for this application the finite number of feasible combinations gives confidence that the returned solution is the optimal split point. Different objectives could be defined, such as distributing ridership evenly. To solve this objective function, three constraints were defined, including northern boundary for southern split of route, southern boundary for northern split of route, and allowable size of overlap (Figure 6b). Any solution that minimizes induced transfers will include a solution for which the maximum allowable overlap of two routes is a limiting constraint.

Defining the new route terminal locations and allowable size of route overlap yielded optimal split points in midtown Manhattan. Under these constraints, a new southern terminal was found that would induce new transfers for 6.9% of riders, exhibiting uneven length ratios (1.8:1) and ridership ratios (4.7:1). The optimization was repeated with new terminal constraints moved north and the same size of route overlap, producing a new optimal split in the Upper West Side. This solution impacted 11.2% of riders, but distributed ridership between the new routes in more equal ratios (1.3:1). This approach considers only the optimal solution with respect to ridership. In practice, the bus planning process must address other considerations beyond just optimal ridership solutions. While attempting to minimize passenger impacts, operational and cost considerations also were noted, and a proposed split point for the M5 was identified nearby layover space in midtown Manhattan within the first-pass optimal split bounds. Considering operational constraints and a more cost-effective constraint of a single stop overlap, ridership impacts were recalculated. The change in the tolerable route overlap constraint and the assignment of the split point resulted in an impact on 10.2% of riders. Operational feasibility would have to take precedence over solutions minimizing ridership impacts. In the absence of this input to the planning process, professional judgment or anecdotal evidence may have guided where the split location should occur with limited manual data sources to support conclusions. The inclusion of a robust ridership data set helps balance professional judgment along with historic precedent and community input. The data supporting these conclusions had the
additional benefit of being free, as they were a usable byproduct of fare collection and customer-facing AVL data. This is in notable comparison with what may formerly have been necessary, such as an intercept survey to understand passenger travel patterns.

**Estimating New Route Running Time**

Schedule implications and running times could be predicted from existing bus data. Ridership in the northern section of the route had governed frequency, and a new schedule for the southern route could be written from ridership data. Historic AVL data could be used to predict the expected running time for the new routes. Running time and schedule frequency predictions allowed the proposed service change to be evaluated on the basis of cost in addition to performance and ridership impacts.

The M5 currently operates a limited service for most of the day but overnight has a local service pattern. Feedback that the limited and local divide was confusing prompted the idea to rationalize service patterns while the route length was being modified. Although AVL data could be used to estimate running times of two new routes from existing service, no such data existed for a local-only route traversing this path. By measuring local running times on segments of other routes overlapping the M5, limited pattern running times for a proposed new route path and service pattern could be estimated, including the new bus and cost requirements. The new northern route would overlap with other routes that run local-only service the entire day. This allowed for the predicted running times of a local-stopping pattern to be predicted for the M5. For northbound service, the M7 and M5 have overlapping service patterns from Sixth Avenue and 37th Street to Broadway and 70th Street; the difference in limited and local running time could be evaluated from AVL data. A similar overlap existed with the M5 and M4 so that an average increase in running time of 8.3 minutes or 12% was determined for the new northern route. This process was repeated with the M1, M2, M3, M4, and M7 routes to produce southbound estimates, which suggested a 9% increase in running times. Figure 7 better clarifies where local running time data were recorded to produce new running time estimates. The predicted increase in travel time due to the service change led to the abandonment of the proposed idea as planning moved forward.

Big data sources were important in every step of the proposed route modification—problem declaration, optimal solution from a passenger perspective, performance improvements, schedule, running time, and resulting cost implications of the proposed change. In the absence of big data, none of these processes would have been supported and would have involved far greater uncertainty and data of a much greater cost.
Conclusions and Next Steps

Big data sources at NYCT allow for improved bus service planning capabilities in several areas. Most significant was the new information about rider boarding and alighting locations, allowing for estimations of passenger impacts on route re-designs and other route characteristics. The data provide a more complete picture of ridership without relying on intuition or anecdotal or incomplete sources. Replacing manual data with automated data improved the confidence in the information being communicated, and freed up staff time to use and analyze data rather than processing it. Specific rider detail can be obtained without costly and time-consuming surveys or checks while also getting a more robust picture. This allows for more studies to be undertaken and scaled to higher-level network analyses, going from route to neighborhood to borough-level detail. Novel data sources were able to result in service and operational improvements in a Bronx neighborhood, including a re-route and stop addition. This also allowed for a ridership-optimal Manhattan route split at 38th street while supporting the Community Board and MTA Board procedural pieces of route modification.
Integrating ridership with demographic data will also enhance NYCT’s ability to perform Title VI analyses that currently rely on outside data sources such as the Census Journey to Work surveys to infer travel patterns. The different geographic and ridership sizes of the bus planning projects undertaken and planned by NYCT means that most bus-transit providers could extend these kinds of approaches to their planning processes.

Most of the data that supported these efforts have been incorporated into web-based reports, meaning that an analytical intermediate is not required to facilitate this kind of planning work. Boarding-alighting, O-D pairs, performance, running time, and others can be queried via these web-based reports by route, date, and time without intimate knowledge of the data or technologies used, increasing the applicability, flexibility and ease with which these analyses are conducted. This allows for the work described in this paper to be repeated on an ongoing basis, while development of other potentially useful data sources such as linked-trip O-Ds or non-revenue service analysis can be explored.

References


About the Authors

JEFFREY HANFT (jeffrey.hanft@nyct.com) is a Principal Transportation Planner in the System Data & Research unit of Operations Planning at NYCT. He has an undergraduate degree in Engineering from the University of Delaware and a master’s degree in Transportation Systems Analysis and Planning from Northwestern University.
Shrisan Iyer (shrisan.iyer@nyct.com) is a Principal Transportation Planner in the System Data & Research unit of Operations Planning at NYCT. He holds an M.Sc. in Transportation Engineering from Rutgers University and has co-authored several papers in the fields of transportation economics, modeling and simulation, and big data.

Brian Levine (brian.levine@nyct.com) is currently a manager of Operations Research in the System Data & Research unit of Operations Planning at NYCT, whose primary function is to provide analytical tools that can help monitor and improve operational performance. He has interests in aviation and public transit and undergraduate and graduate degrees from Cornell University in Transportation Systems Engineering.

Alla V. Reddy (alla.reddy@nyct.com) is the Senior Director of System Data & Research in NYCT’s Division of Operations Planning. He holds an M.S. in Operations Research from NYU Polytechnic in Brooklyn. In his more than 30 years of NYCT experience, he has practiced in the areas of industrial engineering, internal auditing, materials management, performance analysis, and environmental justice. His research interest focuses on using quantitative research and emerging technologies to solve problems and improve system productivity and efficiency.