Measuring Subway Service Performance at New York City Transit: A Case Study Using Automated Train Supervision (ATS) Track-Occupancy Data

Brian Levine
Operations Planning,
MTA New York City Transit,
2 Broadway, Cubicle A17.100,
New York, N.Y. 10004-2207
<Brian.Levine@nyct.com>

Alex Lu
Transit Analyst,
Haverhill, Massachusetts 01835-7234
<lexcie@gmail.com>

Alla Reddy (corresponding author)
Sr. Director, System Data & Research,
Operations Planning,
MTA New York City Transit,
2 Broadway, Office A17.92,
New York, N.Y. 10004-2207
<Alla.Reddy@nyct.com>

Word Count: 202 (Abstract) + 5,696 Words + (6 Figures * 250 Words) = 7,398 Words
ABSTRACT

A recurring challenge facing transit managers today is the persistent question of how to do more with less—to maintain and improve service despite deficits of historic proportions. New York City Transit (NYCT) responded by re-tooling performance measurement frameworks and procedures to better capture customers’ perspective, respond to management initiatives, and incentivize proper operating decisions. NYCT’s primary performance measure, Wait Assessment (WA), measures customers’ maximum wait times while waiting to board at stations. Defined as percent of headways between trains not exceeding 125% of scheduled headways, a “Reach and Match” algorithm was developed to account for NYCT’s irregularly scheduled service and ensure customer experienced headways are matched to the specific published scheduled headway in effect at that moment, regardless of which scheduled trip was supposed to arrive. Upgrading sample-based methods that gathered limited data manually, track-occupancy data was downloaded from the Automated Train Supervision (ATS) system for the No.1 through No.6 routes, providing 100% coverage, much lower public reporting time-lag, and the ability to take near-term corrective action. The increase in data availability also allows NYCT to easily consider corridor-level and track-level WA standards for internal diagnostic purposes, analyzing train performance in shared-track territory regardless of route designations, to provide better service.
INTRODUCTION

New York City Transit (NYCT) operates the world’s third largest subway system (by annual ridership), carrying about 5.0 million riders on an average weekday. The subway system extends 830 track miles through four boroughs, covering 321 square miles and serving 8.0 million people 24-hours, seven days a week. NYCT’s predominant role is to ensure trains and buses operate safely, reliably, on-time, and provide convenient services cost effectively. Outside of both operations management and customer advocacy groups, an independent performance audit infrastructure helps to ensure NYCT is properly carrying out its mission. Continuous applied research allows improvements in monitoring methodologies and service delivery.

This paper describes NYCT’s case study in using track-occupancy data to measure subway service performance with Wait Assessment (WA), and to understand how WA is useful in analyzing and improving service. WA is the percentage of actual headways between successive trains less than or equal to prescribed standards. An algorithm was developed that objectively matches actual observed train headways to scheduled service headways specified in the timetable. Prior performance research provided matching either based on trip identifiers or subjectively by schedule or operations experts (1). This algorithm builds upon previous research by considering actual headways customers should experience at that moment in time, instead of using averages, making standardized performance results sensitive to both service delivery and proper schedulemaking. The algorithm also avoids subjectivity or ambiguity in analytical processes matching actual observations to scheduled timeslots where trips were dropped, as discussed in the literature (1). Algorithmic matching results using various measures and standards, and “flash” and outlier reports provided daily to operations management are presented. Results are used to diagnose performance problems and ultimately recommend strategies for improving service delivery. This research is one piece of overall performance measurement and service improvement framework at NYCT, which includes Passenger Environment Survey (2), general performance indicator reporting (3), key performance indicators (16), environmental justice monitoring (17), fare evasion monitoring (18), and others.

New York City Subway’s Characteristics

NYCT uses relay-based interlockings to control train traffic throughout its “A” Division (formerly Interborough Rapid Transit, IRT) network of seven numbered subway routes. ATS (Automated Train Supervision) is a non-vital centralized dispatching system overlaid on existing local relay logic and remote control panels in master towers. ATS provides Operations Control Center (OCC) with real-time track occupancy information and track each train’s identity as they proceed throughout the system, allowing automated route-setting using pre-loaded schedule data (19). Real-time track occupancy information provided by ATS is stored in central servers, providing constant service reliability and performance monitoring capabilities, and aiding service delivery improvement strategies.

NYCT is unique in North America in providing frequent service throughout the day over multiple interconnected routes. Since the subway is an amalgamation of three previously independent systems, many passengers must transfer trains to reach their final destination. Passengers are thus generally concerned with waiting time as opposed to on-time performance at trains’ termini. During rush hours, all routes have typical headways between 2 and 8 minutes, and passengers do not generally arrive according to prescribed schedules; instead they tend to arrive randomly since they know waiting times are short. NYCT’s public timetables do not generally give exact arrival times; instead they say, e.g.,
weekday service is provided every 4–7 minutes on Northbound “5” route. During off-peak hours, standard service headway is 10 minutes or more frequent, whereas overnight, policy headway is every 20 minutes.

Another feature of the NYCT subway schedule is that trunk line headways are rarely uniform, due to interactions between routes with different service frequency at flat junctions or merge points:

- Summer 2006 schedules contained off-peak service requiring southbound “4” and “5” trains to merge at 149 St. in the Bronx on a 2:1 ratio, with “4” trains operating every five minutes north of 149 St., but changing to a four-and-six minute pattern to the south, allowing a “5” train to operate in the six-minute gap between successive “4” trains.
- Winter 2009 schedules required “M” trains to cross over from a division where prevailing headways are multiples of six, to one where prevailing headways are multiples of eight. A schedule conflict therefore occurs predictably every two hours, which must be resolved by introducing irregular intervals.

Performance Indicators

The Performance Indicator (PI) program was established in 1994 in response to research (4) recommending service reliability measures beside traditional Terminal On-Time Performance (TOTP). TOTP is suitable for commuter railroads where many customers are traveling to the central business district (CBD) final stop. However, transit routes tend to drop off and pick-up most passengers at intermediate stations, which requires more sophisticated measures blending waiting and travel time experiences from a customer perspective.

Turnquist and Bowman (5) describe network structure effects on service reliability, finding that controlling link travel time variability and scheduling to ensure easy transfers are both important, and that service reliability is very sensitive to service frequency. This becomes a key factor in NYCT subway system, requiring reliability measures that can distinguish between minor differences in service.

Extensive research had been conducted to understand service reliability from passengers’ and transit managers’ perspectives (6), building on headway variance models (7,8). Furth and Muller (9) describe a method to determine potential passenger waiting time using automated vehicle location data, assuming for short headways, passengers arrive independent of vehicles. While this may be true on single route lines, many NYCT’s CBD stations are large complexes where transfers are possible between numerous routes, hence leading to heavy loading when trains arrive. Statistical service reliability measures like root-mean-squared average passenger wait time (10) were considered too complex for use as public measures; a key criterion of NYCT’s performance measures is that people would not need mathematical backgrounds to understand its significance.

NYCT developed a simplified version of algorithms and ideas described above, more easily understood by passengers and operational management. Wait Assessment (WA), which effectively measures how long customers may wait for given trains at given stations, is calculated by comparing successive headways between trains to prescribed standards. Though this measure does not focus on specific passengers, it ensures that longest passenger wait times are within defined standards, even if many passengers were not waiting as long. In effect, it describes a worst-case scenario. This strict standard promotes consistent service throughout the whole system; in fact, many passengers have wait times
much shorter than this amount, especially at large transfer complexes. Since service during overnight periods is less frequent and many people arrive according to schedule, on-time performance is the primary measure.

PI’s main purpose is to monitor how well NYCT is providing service. Wait Assessment is publicly reported at systemwide and route levels on MTA’s Performance Dashboard (23), and are used by rider advocacy groups in their annual rating of subway routes—the State of the Subways (11). Maintaining a transparent and accountable performance reporting process is critical to achieving public trust. Indeed, stakeholder and watchdog groups have adopted MTA’s official measures for their performance reporting.

WAIT ASSESSMENT “REACH AND MATCH” ALGORITHM

Wait Assessment applies an analytical algorithm on raw data collected by surveyors or ATS system, specifically using departure times of all vehicles passing a location. Data is collected at all en-route timepoints (Figure 1(a)), because majority of riders enter and exit the subway at intermediate stops, not just terminals. It is important to provide reliability measures at these locations. Typical routes have anywhere from 5-15 timepoints (25-50% of all stops), a subset of which are used for WA reporting.

WA reporting locations are agreed upon by management in Operations Planning and Rapid Transit Operations, and generally include major transfer stations, hubs, and originating terminals, since many customers board at these locations. Figure 1(a) shows current WA timing points for ATS-enabled A-Division routes. Timepoints with dashed circles are only used for trains originating there, e.g. the “2” route (Wakefield-241 St., Bronx to Flatbush Ave., Brooklyn) contains 13 timepoints out of 49 total stops. Currently, eight timepoints are assessed for WA: 1-2 in the Bronx, 4 in Manhattan, and 2-3 in Brooklyn. Timepoints for non-ATS territory are shown in Figure 1(b).

Algorithm described below matches actual headways to scheduled headways, then applying one uniform standard across all trains on all routes to ensure consistent reporting. This methodology accounts for varying headways on one route due to operational characteristics, instead of applying the standard against an average headway. Consider a hypothetical service operating every 3-4 minutes during rush hour—implying an average headway of 3½-minutes—when in actuality at any given moment the scheduled headway is either three or four minutes. The goal is to match actual service provided at one specific point in time to scheduled service at that exact moment.

For all actual observations of consecutive trains departing a given timepoint, actual headways must be matched to scheduled headways based on daily operating schedule (with supplemental schedules applied), to identify trains that met WA standards. Supplements are planned schedule alterations due to construction and/or track maintenance necessitating train reroutes and added running time. These schedule changes are published online, and the public should therefore be aware of these alterations. The matching process is governed by the “Reach and Match” algorithm, described briefly below (more details towards the end—Figure 6(a)). Matches are made at timepoint locations only, where scheduled departure times are published. NYCT’s subway schedule features holding times at key transfer stations, thus departure rather than arrival times are assessed. A train may be scheduled to arrive at 10:52am but not scheduled to depart until 10:54am, even though passengers may board the train once it arrives. WA is essentially designed to assure even train departures.
This algorithm’s primary element is the train matching step, whereby actual headways are matched to scheduled headways based on departure time. Actual trains are matched if they fall within acceptable boundaries, based on scheduled headways between itself and adjacent trains in both temporal directions (i.e. prior and following trains). When large gaps exist in service, attempted matches can be made between one actual train and multiple scheduled slots, but the train may only fulfill one single best-fit slot. Actual trains may be unmatched when too much service was provided within a single scheduled slot— extra trains should not improve WA if they do not provide service in distinctly separate intervals. A single train that successfully picked up passengers within one slot cannot be evaluated again, however trains with large gaps may fail to match multiple scheduled slots.

This algorithm was developed to account for the naturally occurring “drift” as actual trains move out of scheduled slots in normal daily operations. The intent is to describe service headways as experienced by customers expecting a train to arrive every $h$ minute(s), where $h$ is timetable-specified headway, i.e. regardless of what specific train was supposed to arrive in that slot. This “Reach” criterion prevents actual trains and scheduled slots from drifting too far apart. While up to one headway of give is allowed to account for operational schedule adjustments (called “flexing”), as soon as actual departure times drift out-of-sync with timetabled slots, it is “Out-of-Reach” and not used to make a “Match”.

After all trains are matched, WA results are calculated by comparing actual service headways to scheduled service headways. If actual headway is greater than scheduled by an allowable margin, that interval is denoted as failing. A discussion of different WA standards and measures follow.

### MEASURES OF WAIT ASSESSMENT

$T$ represents threshold headway which delineates passing and failing WA headways. Exact values used depend on how strict the performance measure is intended to be.

#### Absolute vs. Relative WA

As first conceived, WA was an absolute measure of relative performance. It’s an absolute measure because thresholds of acceptable excess wait time ($T$) is a fixed quantity by time period (+2 minutes peak, +4 minutes off-peak). However, it measures relative performance because it’s headway-based—obtained by comparing train departure time with its predecessor, and not by comparison with fixed schedules. The rationale was to provide customers with a fixed standard of excess wait time above which service headways are considered unacceptable. This type of metric has one interesting property: routes scheduled with shorter headways tend to score higher in Wait Assessment, because probability of a train—any train—achieving that two-/four-minute window above headway is simply higher.

In discussions with operations management, it became apparent that this property does not give dispatchers correct incentives. High frequency service routes are often very congested, where smallest perturbations in headways or ridership volume can quickly snowball into bunched service and big gaps. On lower frequency routes, dispatchers have more latitude to adjust schedules; proper headway management isn’t as critical to maintaining service performance as holding for connections at major transfer points. To prevent such imbalance, standards to which each route is held must be a function of...
prevailing service frequency, with busier routes held to more exacting standards. In turn, tolerable excess wait time must be specified relative to the headway.

New York City’s “7” route (Times Sq., Manhattan to Flushing Main St., Queens) operates every 4-5 minutes even during off-peak periods. With up to four minutes of acceptable excess wait time, dispatchers can still achieve WA scores above 90% even if delayed trains departed in bunches, which does not incentivize dispatchers to properly maintain train regulation. Earlier research (20) showed that dispatching decisions are often driven by crew requirements; absent strong incentives to maintain proper headway, dispatchers sometimes allowed trains following heavily delays to operate without being checked, in an effort to “catch up” to schedule and therefore minimize required real-time crew manipulations at terminals.

After consultation with stakeholders and management, WA threshold was modified to be +25% of scheduled headway, thereby making WA a strictly relative performance measure. This corresponds to \( T = \frac{1}{4}(t_{i+1} - t_i) \), where \( t_{i+1} - t_i \) is scheduled headway between consecutive trains. The strictest standard at +25% (Figure 2) was selected, to assure the public that NYCT is looking for continuous service delivery improvements. The formal WA definition subways is thus: “percentage of actual headways that does not exceed 25% of scheduled headways,” e.g. for scheduled headways of four minutes, actual headways of less than five minutes are permissible.

Revised Wait Assessment shows results of tighter Headway +25% standards without giving any information about distribution of headways not meeting it. Line Managers (responsible for one single route) felt that knowing, say, 20% of their trains were more than 25% outside scheduled headway did not help them pinpoint sources of bunching problems, which often started with an overcrowded train that could have a gap of more than double the headway from its predecessor. The “failing” WA headways were then further broken into three subcategories: 25%-50% more than headway, considered “minor gap”, 50%-100% above headway—“medium gap”, and more than double headway—“major gap”. This reporting (Figure 2) clearly shows different service levels provided, and offers a multi-dimensional picture of actual service quality delivered and system performance without relaxing standards. Though overall WA at 25% for the entire system hovers around 79%, approximately 95% of observed headways were actually less than double scheduled headways, showing that despite incidents and crowding affecting quality, service was consistently available on these routes, albeit at increased headways.

Current development work will allow Figure 2 to become interactive, allowing management to drill down to see location(s) and time(s) of day where WA did not meet 25% standard, and to determine where major gaps tend to occur regularly—which could be proactively mitigated with an appropriately named “gap train”. Reports shown in Figure 5 provide examples of location- and time-based drill-down, which will be expanded to provide distribution data.

Using ATS data, WA can be calculated for all routes at all time periods, allowing peak periods to be reported separately, explicitly monitoring performance during maximum ridership.

**Line-, Corridor-, and Track-Level Wait Assessment**

New York has many subway routes, some of which are co-routed on the same physical line infrastructure. Queens Boulevard Line is a major four-track subway corridor in Queens that actually
hosts two local routes, “R” and “M” Trains, and two express routes, “E” and “F” Trains. Local routes share local tracks, making all stations stops; express routes share express tracks that have platforms only at major transfer points.

This creates a dilemma when measuring WA at stations served by more than one route, where customers have choices as to which route to use. Passengers may take the first train that arrives, to get as close to their destination as that route permits, and make transfers to other routes to complete their journey. Yet other customers prefer one-seat rides and wait longer for the exact route they require. WA is generally route-based, measuring headways only between trains on the same route—and does not consider headways between trains sharing tracks if they are assigned different route letters or numbers.

An internal NYCT debate is ongoing about how such shared-track corridors should be managed. Customers destined for outlying branches (e.g. Concourse, Rockaway, Culver Lines) often require specific routes to reach their final destination, therefore are interested in knowing if their route is having bunching/spacing problems. Customers who use trunk sections exclusively (e.g. 8 Avenue, 6 Avenue Lines), or make inter-divisional transfers at major transfer points (e.g. Times Sq., Union Sq., Fulton St.), are usually more interested in corridor-level measures because route designations are only of passing interest—for they simply require any train headed in the direction they’re traveling. When routes share track, it is sometimes operationally important for crowding reasons to keep even spacing between trains with different route designations, therefore operations management usually are more interested in corridor-level measures.

Furthermore, when dealing with incidents affecting service like sick passengers or an inoperative switch, transit supervisors often reroute trains from express to local tracks or vice versa, to provide service where there otherwise would have been none. This happens quite often in NYCT’s network, making quite a strong case for track-level WA measurements whereby route designations are ignored—rather the fact that service is provided on a particular track segment better reflects what actual customers experience.

Having both route- and track-level WA results provide a complete picture of service experienced by different passenger types. Passengers requiring a particular service are more concerned with route-level results, whereas passengers indifferent to service designation are appropriately concerned with track-level WA. Overall results together are most reflective of customer experience.

Figure 3(a) shows comparative WA results for route-, corridor-, and track-levels for Friday, September 23, 2011 on Brooklyn’s Eastern Parkway (EPK) corridor for the 17:00–21:00 period. EPK (Figure 3(b)) consists of “2” and “3” local routes, and “4” and “5” express routes. These interborough routes travel from Bronx/Northern Manhattan to Brooklyn, with “2” and “5” having geographically proximate Bronx and Brooklyn termini, and “3” and “4” having closely terminals in Brooklyn. Accordingly, service can be adjusted across routes to provide better service when incidents occur.

On September 23 numerous incidents on Manhattan’s 7 Ave IRT affected evening Southbound service in Manhattan and Brooklyn. Department of Subways attempted to balance service by rerouting trains; local “2” and “3” Trains experienced a partial blockage in Manhattan, thus selected express “4” and “5” trains were rerouted at Nevins Street Interlocking to “run local” when arriving in Brooklyn, to serve Bergen Street, Grand Army Plaza, and Eastern Parkway stations. Local track WA was slightly higher
than corridor WA due to express train reroutes to local track (Figure 3(a)). Similarly, express track WA along EPK was lower than corridor WA due to removal of “4” and ”5" trains from express track. Across the whole day, WA differences were minimal (63.2% versus 62.2% in “4/5” corridor versus track; 57.0% versus 57.2% in “2/3” corridor versus track), but track-level performance better reflects what customers experienced that day.

Local track WA 25% was not much higher than corridor WA 25% due to fewer overall trains in service. Since WA is indirectly a function of train throughput at a given stop, regardless of operational changes to alleviate inconsistencies caused by incidents, WA at 25% cannot improve drastically when fewer trains are in service. However, track-level WA 100% is higher than corridor-level WA by over 10%, indicating although fewer trains were available, they were spaced at approximately double scheduled headways, quantifying attempts made to provide service where there otherwise would have been none. Although trains operating on local tracks were officially designated “4” trains, customers used them to reach local stops. Express track WA decreased slightly compared to corridor WA, as expected, but it shows that express service was not severely degraded even though some trains were diverted.

**ALGORITHM RESULTS**

Figure 4(a) shows headway matching algorithm results for Brooklyn’s “2” route, southbound at Atlantic Avenue on 9/23/2011, along with WA calculation at 25% and 100% standards. Each train is assigned a train identifier, which indicates route number (05), origin departure time (1301+, 1309+, etc), followed by plus signs if trains were scheduled to depart on the half minute (e.g. 1301+ is 13:01:30), and also codes for origin (241) and destination (FLA) terminals. The primary terminals for “2” trains are 241 (Wakefield-241 St., Bronx), and FLA (Flatbush Av.-Brooklyn College).

First scheduled headways are determined at that particular station when the train is scheduled to arrive. The “reach and match” algorithm matches headways based on relationships between scheduled time, scheduled headway, and actual departure time, to determine whether an observed departure is within reach of a scheduled slot. Actual headways that are too wide or narrow can easily cause actual times to fall out of reach of scheduled slots. Figure 4(a) show a few interesting and important properties of the algorithm, discussed in Figure 4(b).

From these results, line managers could identify strategies to improve service. First, trains in slots 3, 4, and 5 were clearly ahead of schedule, negatively impacting performance even when later trains arrive on time. Holding trains to scheduled departure times at key timepoints can help alleviate this problem. Alternatively, if those trains run consistently early, schedules can be adjusted.

Sometimes, gaps in service arise during rush hours solely because of congestion and merging, or due to incidents like sick passengers. “Gap” trains can be strategically placed within the system, to be activated when such an incident occurs. Service gap between slots 11 and 13 lasting nearly 24 minutes could have been partially alleviated by filling in service with a gap train. Additionally, during these scenarios, service could be rerouted or diverted from other routes to provide “2” Train service.

**Daily Reporting & Operational Impacts**

Availability of extensive data downloads from ATS provides 100% coverage on NYCT’s IRT division (except “7” route) and yields much lower time-lags for compiling performance measures, allowing near-
Daily “outlier” reports are issued to assist in identifying troublespots. Daily summary reports (Figure 5(a)) provide all performance information about one train route on a single page. For each hour and timepoint location, WA and throughput (train count passing that location) results during that hour is given, allowing line managers to see at a glance how their route performed the previous day—and more importantly, if an incident occurred, what was its performance impact. Route-level results are currently calculated and reported daily; track- and corridor-level results are available experimentally.

Without detailed daily knowledge of incidents, it can be difficult to determine whether lower WA scores during certain hours was typical, and if problems are ongoing and repeating or due to specific non-recurring incidents. Operation managers generally prefer to focus on recurring problems rather than unusual incidents. Figure 5(b) compares today’s WA statistics (by hour and location) with rolling averages over past 30 days (where data is available). “Low outlier” hour-and-location combinations are printed out, allowing managers to investigate further.

Both reports are presented at the 10:00am operations meeting the day thereafter; managers are also provided with previous day’s list of incidents, as to focus on recurring problems (e.g. a slow train operator) and review of how dispatchers responded to an incident. This can be used to debrief supervisors on incident responses to provide better passenger service.

Southbound “1” route’s 09:00 hour shows WA scores of 50-70% for many stations from mid-route through destination terminal. These entries are not present in outlier reports, indicating similar results for previous 30 days; this may imply recurring capacity problems solved by schedule adjustments or capital improvements. Conversely, Southbound 07:00 hour is an outlier, implying an incident causing lower than expected performance. Managers review incident reports and logs to determine whether appropriate actions were taken, and how responses might be improved in future.

**WAIT ASSESSMENT “REACH AND MATCH” ALGORITHM DETAIL**

Before describing WA “reach and match” algorithm in detail, following notation is presented. Let:

- \( I \) = Set of trains in the schedule having the same route identifier, direction, and timepoint location
- \( J \) = Set of trains in the actual data, sorted by actual departure time
- \( i \) = Current scheduled train from set \( I \) being processed
- \( j \) = Current actual train from set \( J \) being processed
- \( n \) = Pointer to actual train from set \( J \) to be used for matching
- \( t_i \) = Scheduled departure time for train \( i \)
- \( t_j \) = Actual departure time for train \( j \)
- \( t_{\text{Min}} \) = Minimum (earliest) matching limit for scheduled departure time of train \( i \)
- \( t_{\text{Max}} \) = Maximum (latest) matching limit for scheduled departure time of train \( i \)
- \( \text{Tag}_j \) = Tag assigned to actual train \( j \) if it has been used in matching process
- \( \text{Match}_{ij} \) = Array of tags assigned to the match between scheduled train \( i \) and actual train \( j \)
- \( WA_i \) = Wait Assessment result for scheduled train \( i \)
\[ T = \text{Headway threshold for Wait Assessment to pass} \]

**Step 0: Input**

For a given date, set \( I \) with corresponding \( t_i \). For a given date, set \( J \) with corresponding \( t_j \). Repeat the steps below for each unique grouping of route, timepoint, and direction.

**Step 1: Sorting**

Sort the set of scheduled trains \( I \) by scheduled departure time, such that \( I = \{1, 2, 3, \ldots, i_{\text{Max}}\} \). Sort the set of actual train observations \( J \) by actual departure time, such that \( J = \{1, 2, 3, \ldots, j_{\text{Max}}\} \).

**Step 2: Initialization**

Initialize \( i \) and \( j \) to the first scheduled and actual trains of the day, respectively. \( i \) is initialized to the first scheduled train after midnight (\( i = 1 \)). Set \( j = 1 \). Then, increment \( j \) until the actual departure time of the train is later than or equal to the first scheduled train after midnight, i.e. until \( t_j \geq t_i \), where \( i = 1 \). Set \( n = 0 \). \( n \) is a pointer representing how far ahead or backward the algorithm looks to find a matching actual train. Set \( \text{Tag}_j = \text{"Not Used"} \) for all trains \( j \). Perform steps 3 to 5 for each scheduled departure \( i \) of the route, direction, and timepoint group.

**Step 3: Train Matching**

Determine acceptable matching boundaries. The maximum acceptable \( t_j \) for a match to \( i \) is \((t_{i+1})\), the scheduled departure time of the next train \( i+1 \), also called \( t_{\text{Max}} \). The minimum acceptable \( t_j \) is \( t_{\text{Min}} = (t_i - (t_{i+1} - t_i)) \), the scheduled train departure time \( t_i \) minus the scheduled headway \( t_{i+1} - t_i \). The reason \( t_{\text{Min}} \) is not equal to \( t_{ij1} \) is because during transition periods between peak hour service and off-peak service, some of NYCT's routes have somewhat irregular headways due to operational reasons. To facilitate proper matching of these irregular headways, each train’s acceptable matching boundary is based on the headway between itself and the following train, and not the prior train, shown in Figure 8(a). If a train falls within acceptable matching boundaries, it is denoted as “Within Reach.” Determine if the actual departure time of train \( j \) (\( t_j \)) is within acceptable matching boundaries and process the train accordingly:

a) If \( t_{\text{Min}} \leq t_{j+n} \leq t_{\text{Max}} \) (i.e. train \( j+n \) within acceptable matching boundaries for scheduled train \( i \)), then \( \text{Match}_{i,j+n} = \text{"Matched—Within Reach"} \), and \( \text{Tag}_{j+n} = \text{"Used"} \). 

b) If \( t_{j+n} \geq t_{\text{Max}} \) or \( t_{j+n} \leq t_{\text{Min}} \) (i.e. train \( j+n \) not within acceptable matching boundaries), check future trains for a potential match:

1. Increment \( n = n + 1 \) until train \( j+n \) satisfies \( t_{\text{Min}} \leq t_{j+n} \leq t_{\text{Max}} \) or \( j+n = j_{\text{Max}} \). If match is found, then \( \text{Match}_{i,j+n} = \text{"Matched—Within Reach"} \); \( \text{Tag}_{j+n} = \text{"Used"} \). Go to Step 4.

2. If no match is found, set \( n = -1 \) and check to see if train \( j+n \) satisfies \( t_{\text{Min}} \leq t_{j+n} \leq t_{\text{Max}} \). If yes:
   a. If \( \text{Tag}_{j+n} = \text{"Used"} \) this indicates actual train \( j+n \) may be matched to multiple scheduled trains \( i \), i.e. \( \text{Tag}_{j+n} \) will be “Used—Repeat” and \( \text{Match}_{i,j+n} \) will be “Matched—Repeat Train”. To determine the best possible match, a “Half Headway” test is applied.
      i. Determine the scheduled headway \( (t_{i+1} - t_i) \) and the actual headway \( (t_{j+n+1} - t_{j+n}) \).
         Let the headway deviation \( (\text{Dev}_{i,j+n}) \) be the difference between scheduled and actual headways, i.e. \( (t_{j+n+1} - t_{j+n}) - (t_{i+1} - t_i) \).
ii. Determine if the headway deviation (Dev_{i,j+n}) is within ±50% of the scheduled headway \((t_{i+1} - t_i)\), or 5 minutes, whichever is less. If
\[-\frac{1}{2} (t_{i+1} - t_i) \leq \text{Dev}_{i,j+n} \leq \frac{1}{2} (t_{i+1} - t_i) \quad \text{and} \quad -5 \text{ min} \leq \text{Dev}_{i,j+n} \leq 5 \text{ min},\]
then the “Half Headway” test passes, and \(\text{Match}_{i,j+1} = \text{“Matched—Best Possible”}\). Since only actual train \(j+n\) fits within the acceptable train matching boundary, this actual train is a best match to previous scheduled train \(i-1\), and since an actual train can only be credited once, this implies scheduled train \(i\) “Autofails”. Set \(WA_i = \text{“Autofail”}\). Go to Step 4.

iii. If not, then \(Tag_{j+n} = \text{“Used—Repeat”}\) and \(\text{Match}_{i,j+n} = \text{“Matched—Repeat Train”}\). This signals that although the previous scheduled headway \(ij_1\) is a technical match, the current scheduled train \(i\) is a better match to the actual train \(j+n\) being considered. Go to Step 4.

b. If \(Tag_{j+n} = \text{“Not Used”}\), then set \(\text{Match}_{i,j+n} = \text{“Matched—Within Reach”}; Tag_{j+n} = \text{“Used”}\). This should never occur, since the algorithm works in increasing order of actual train observations \(J\). Having to set \(n = -1\) (go backwards in time to find a match) implies there is a shortage of trains, i.e. actual train throughput is lower than scheduled throughput, and the previous schedule departure should have matched this actual train as the algorithm looks predominantly ahead in time to find possible matches. However, pedantic implementation of this algorithm usually provides a check to ensure that every train is correctly matched. Go to Step 4.

3. If \(t_{\text{Min}} > t_{j+n}\) or \(t_{j+n} > t_{\text{Max}}\), then for all \(t_{j*} \in \{ t_{j-1}, t_j, t_{j+1}, \ldots, t_{\text{Max}} \}\), \(t_{j*} < t_{\text{Min}}\) and \(t_{j*} > t_{\text{Max}}\) i.e. the departure times of all actual trains \(j*\) are not within the acceptable train matching boundary for scheduled train \(i\). This implies scheduled train \(i\) auto-fails: \(WA_i = \text{“Autofail”}\). There is no possible match to an actual train \(j\), thus scheduled train \(i\) is determined to fail by default. Go to Step 4.

**Step 4: Result Calculation**

After each “Match” result is computed (Matched—Within Reach, Matched—Best Possible, Matched—Repeat Train), a Wait Assessment (WA) result is calculated using the current scheduled train \(i\).

1. If \(\text{Match}_{i,j+n} = \text{“Matched—Repeat Train”} \quad \text{and} \quad WA_{i-1} = \text{“Pass”}\), then \(WA_i = \text{“Fail”}\). This is to prevent the same actual headway \(j\) for being credited against two scheduled headways \(i_1\) and \(i_2\). If \(WA_{i-1} = \text{“Fail”}\), then scheduled train \(i\) has the opportunity to pass Wait Assessment.

2. Otherwise, calculate Wait Assessment (WA). Recall \(T\) is the headway threshold by which Wait Assessment passes:
   a. If \((t_{j+n+1} - t_{j+n}) \leq T\), then \(WA_i = \text{“Pass”}\). If the actual headway is less than or equal to the permissible threshold, then Wait Assessment is “Pass”.
   b. If \((t_{j+n+1} - t_{j+n}) > T\), then \(WA_i = \text{“Fail”}\). If the actual headway is greater than the permissible threshold, signifying a gap in service, then Wait Assessment is “Fail”.

**Step 5: Increment Counter**

If \(i+1 \neq \{\}\) (i.e. there is another scheduled departure), increment \(i = i+1; j = j+n+1\). Note that \(n\) may be negative or positive. After \(i\) and \(j\) are incremented, reset \(n = 0\). Return to Step 3.
CONCLUSIONS

MTA New York City Transit (NYCT) responded to challenges of “doing more with less” by re-tooling performance measurement frameworks to better capture service reliability from customers’ perspectives, respond to system improvement initiatives, and incentivize operating decisions that deliver excellent service.

“Reach and Match” algorithm is a crucial piece in WA calculation processes. By applying uniform standards across all trains and routes, consistent public reporting is ensured. The algorithm takes schedules into account but allows flexibility for “on-the-fly” changes made daily by dispatchers to improve service. Recent improvements to WA standards made it a more meaningful relative performance measure, stricter for more frequent routes. Formerly binary pass/fail standard is now replaced with distributions of failing headways, providing customers more detailed views of system performance. NYCT continues to improve performance standards by understanding how WA could be fairly and best applied to shared-track territories where different routes can be treated as a service corridor and train performance analyzed without reference to route designation. Upgrading previous sample-based methods gathering limited data manually, extensive data is downloaded from ATS, providing 100% coverage and much lower reporting time-lag, allowing near-term corrective action by operations supervisors.

These improvements to NYCT’s customer-centric service performance indicators were developed with extensive operations management consultations, have been ratified by MTA’s Board, and endorsed by stakeholders and advocacy groups. In the tradition of improved reporting, NYCT continues to explore new ways of assessing performance for both internal diagnostic purposes and public accountability. WA measures must be consistent with customer experience in the system; if customers experience worse than scheduled service, WA should drop accordingly.

Future Research

Future research involves applying “Reach and Match” algorithms to bus operations using data from on-board GPS devices tracking bus locations. Difficulty in this endeavor is that buses do not necessarily arrive at given stops in same sequences as they leave the terminal, whereas this is the case for trains travelling without overtaking maneuvers. In New York, many bus routes operate at frequencies higher than trains (20, 21), and WA is a crucial performance measure (3, 15). Additional research focuses on determining travel paths of individual passengers, allowing us to compute weighted waiting time measures reflective of individual passenger experiences.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of Barry Greenblatt, Paul McPhee, Dan Mazzella, Herbie Lambert, Tom Calandrella, Bill Shrage, Theresa Cheung, Karl Steel, Leon Coakley, Nancy Yeoh, Bill Fitzgerald, the Subway On-Time Performance Task Force, Hercules Mack, Jean-Raymond Theobal, Chi Chan, Tanya Lipsmann, Aaron Berkovich, Larry Gould, Glenn Lunden, Patrick Diskin, John Cucarese, Anthony Cramer, H. Robert Menhard, and Steven Aievoli during the development phase of this algorithm.
REFERENCES


LIST OF FIGURES

FIGURE 1. New York City Subway System Wait Assessment Timing Points: (a) ATS-A enabled A-Division (numbered lines); (b) Non-ATS territory.

FIGURE 2. Wait Assessment Results with Distribution. Can someone please provide more comprehensive description of this figure?

FIGURE 3. Wait Assessment Experiment on the IRT Eastern Parkway Corridor in Brooklyn: (a) Single-route, Corridor-, and Track-level Wait Assessment Results for Afternoon Peak Period on September 23, 2011; (b) Functional Track Layout of the Segment Discussed.

FIGURE 4. Wait Assessment Detail Results from Signal System Data: (a) Raw Results; (b) Description of Results as it Relates to the Wait Assessment Algorithm.

FIGURE 5. Wait Assessment Daily Flash Reports: (a) Report by Hour and Location; (b) Outlier Report Indicating Worst Performing Locations.

FIGURE 6. Wait Assessment Processing: (a) Flowchart of the Analytical Process; (b) Acceptable Matching Boundaries in the Wait Assessment Algorithm.
New York City Subway System
Automated Train Supervision
A-Division (ATS-A)
Wait Assessment Timing Points

**FIGURE 1.** New York City Subway System Wait Assessment Timing Points: (a) ATS-A enabled A-Division (numbered lines); (b) Non-ATS territory.
FIGURE 2. Wait Assessment Results with Distribution
**FIGURE 3** Wait Assessment Experiment on the IRT Eastern Parkway Corridor in Brooklyn: (a) Single-route, Corridor-, and Track-level Wait Assessment Results for Afternoon Peak Period on September 23, 2011; (b) Functional Track Layout of the Segment Discussed.

<table>
<thead>
<tr>
<th>Route/Line</th>
<th>Scheduled Headways</th>
<th>Passing Headways</th>
<th>Percentage Passed</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA 25% 9/23/2011 at Atlantic Ave 1700-2100 Southbound</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>27</td>
<td>69.20%</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>43</td>
<td>64.20%</td>
</tr>
<tr>
<td>4/5 Corridor</td>
<td>67</td>
<td>43</td>
<td>64.20%</td>
</tr>
<tr>
<td>4/5 Express Track</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>40</td>
<td>59.70%</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>12</td>
<td>44.40%</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>9</td>
<td>33.30%</td>
</tr>
<tr>
<td>2/3 Corridor</td>
<td>56</td>
<td>23</td>
<td>41.10%</td>
</tr>
<tr>
<td>2/3 Local Track</td>
<td>56</td>
<td>25</td>
<td>44.60%</td>
</tr>
<tr>
<td>WA 100% 9/23/2011 at Atlantic Ave 1700-2100 Southbound</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>34</td>
<td>87.20%</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>24</td>
<td>92.30%</td>
</tr>
<tr>
<td>4/5 Corridor</td>
<td>67</td>
<td>54</td>
<td>80.60%</td>
</tr>
<tr>
<td>4/5 Express Track</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>67</td>
<td>52</td>
<td>77.60%</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>13</td>
<td>48.10%</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>13</td>
<td>48.10%</td>
</tr>
<tr>
<td>2/3 Corridor</td>
<td>56</td>
<td>30</td>
<td>53.60%</td>
</tr>
<tr>
<td>2/3 Local Track</td>
<td>56</td>
<td>36</td>
<td>64.30%</td>
</tr>
</tbody>
</table>

(Underground grade separation generally not shown, for simplicity)
Slots 1 through 11 indicate trains and corresponding scheduled headways matched to trains scheduled to provide service during those headways, as indicated by sched_train_id = matched_train_id. Even though matching is successful, certain slots fail wait assessment at the 25% standard. The 1325+ 241/FLA train arrives approximately 4 minutes early, therefore over 10 minutes ahead of its follower. Compared to the scheduled 8 minute headway between trains, this headway fails under the 25% standard.

Slots 11 through 13 indicate a large timeframe without any actual trains passing by (15:32:30 to 15:56:01). This causes slot 11 to fail wait assessment due to the higher than scheduled headway, and slot 12 fails because no train departed within the acceptable range of the algorithm.

Beginning at slot 13, actual trains have drifted out of sync with the scheduled trains; however this does not necessarily cause the slots to fail Wait Assessment. As long as any actual train arrives within reach of the scheduled train headway, the algorithm matches these trains because scheduled headways change from train to train, WA 100% fails for the first slot, but WA 100% passes for the second slot. This feature of the algorithm allows more opportunities for passing credit to be given.

Slots 25 and 26 indicate another repeat match. In this case, slot 25 passes wait assessment at WA 100%, therefore the following slot 26 must automatically fail wait assessment, to prevent a single train from being credited towards two distinct service slots. Extra trains within a given scheduled headway do not help the wait assessment metric.

During the beginning of the rush hour, beginning around 1600 hours, the scheduled headway ranges from 5 to 10 minutes and varies greatly from interval to interval. Actual headway intervals are matched to scheduled headway intervals that are in effect at the time the train actually arrives.
FIGURE 5. Wait Assessment Daily Flash Reports: (a) Report by Hour and Location; (b) Outlier Report Indicating Worst Performing Locations.
For each unique grouping of line, timepoint, and direction:
Sort set of scheduled trains $I$ by scheduled departure time: $I = \{1, 2, 3, ..., i_{\text{max}}\}$
Sort set of actual observations $J$ by actual departure time: $J = \{1, 2, 3, ..., j_{\text{max}}\}$
Initialize $i$ and $j$ to first scheduled ($i = 1$) and actual ($j = 1$) trains after midnight.
Set Tag = "Not Used" for all trains in set $J$. Initialize $n = 0$.

Increment $j$

If $t_j \geq t_i$?

Yes

Increment $i$

$t_{\text{Max}} = (t_i+1)$
$t_{\text{Min}} = (t_j - (t_{i+1} - t_j))$

No

$t_{\text{Min}} \leq t_{j+n} \leq t_{\text{Max}}$?

No

Increment $n$

Yes

$j_{\text{Max}}$ Reached?

No

Set $n = -1$

Yes

$t_{\text{Min}} \leq t_{j+n} \leq t_{\text{Max}}$?

No

Tag$_{j+n}$ = "Used"?

Yes

Duplicate match

Dev$_{i,j+n} = (t_{j+n+1} - t_{j+n}) - (t_{i+1} - t_i)$

No

Match$_{i,j+n}$ = "Matched—Within Reach"  Tag$_{j+n}$ = "Used"?

Yes

$t_{j+n} - t_{j+n} \leq T$?

No

WA$_{i,j+n}$ = "Pass"?

Yes

WA$_i$ = "Fail"

No

WA$_{i,j+n}$ = "Matched—Repeat Train"  Tag$_{j+n}$ = "Used—Repeat"?

Yes

"Half Headway Test"  
$(-\frac{1}{2} (t_{i+1} - t_i) \leq \text{Dev}_{i,j+n} \leq \frac{1}{2} (t_{i+1} - t_i))$

and $(-5 \text{ mins} \leq \text{Dev}_{i,j+n} \leq 5 \text{ mins})$?

No

Yes

Match$_{i,j+n}$ = "Matched—Best Possible"  WA$_i$ = "AutoFail"

(b)

\[
\begin{align*}
\text{Scheduled Trains} & \quad i-2 & i-1 & i & i+1 & i+2 \\
\text{Schedule Departure Time} & \quad t_{i-2} & t_{i-1} & t_{\text{min}} & t_i & t_{\text{max}} = t_{i+1} \\
\text{ } & \quad t_{\text{min}} = (t_i - (t_{i+1} - t_i))
\end{align*}
\]

FIGURE 6. Wait Assessment Processing: (a) Flowchart of the Analytical Process; (b) Acceptable Matching Boundaries in the Wait Assessment Algorithm