Subway Service Performance Measurement at New York City Transit: Recent Advancements

Alla Venugopal Reddy
Sr. Director, System Data & Research, Operations Planning
MTA New York City Transit, 2 Broadway, Office A17.92, New York, N.Y.
Alla.Reddy@nyct.com

Abstract ID: 63

Abstract. A recurring challenge facing North American transit managers in today's economic climate is the persistent question of how to do more with less – not only to maintain but to improve service in the face of deficits of historic proportions, and against a background of public pressure to reduce government costs. MTA New York City Transit (NYCT) has responded to that challenge by re-tooling its performance measurement frameworks to better capture performance from a customer's perspective, respond to management system improvement initiatives, and better incentivize operating decisions that deliver excellent customer service.

NYCT's key operating performance indicators as measured from a customer's perspective have traditionally been in two areas: on-time performance (OTP), and service quality indicators (SQI). The main OTP measure method is called Wait Assessment (WA), designed to measure wait times experienced by customers waiting to board at a station stop. WA applies an analytical algorithm on departure times of all vehicles passing a time point. Initially defined as % of intervals between trains that does not exceed a +2 minutes peak/+4 minutes off-peak threshold more than the scheduled headway, the algorithm has undergone numerous refinements.

WA standard was modified to within +25% of the scheduled headway, thereby making it relative performance measure that is stricter for lines with more frequent service. However, this pass/fail standard does not give any information on the distribution of service intervals not meeting standard, thus it was further refined to use a distribution for failing intervals provides customers a more detailed view of system performance, planned to be effective beginning January 2012. At this time NYCT is considering setting WA standards for shared-track territory, treating different routes as the same service corridor and analyzing performance of trains sharing track together regardless of their route designation.

Upgrading a previous sample-based method that gathered limited data manually, extensive data was downloaded from the Automated Train Supervision (ATS) to provide 100% coverage and much lower time-lag for compiling performance measures. This allowed near-term corrective action by operations supervisors.

Introduction

Public transit systems in the United States are generally owned and operated by state, county, or municipal agencies or authorities receiving financial operating support from the local government parent entity and capital investment from the Federal government. New York City Transit (NYCT) is an operating agency supported by the Metropolitan Transportation Authority (MTA), State of New York. The primary role of the MTA is to provide financial management and oversight, bonding authority, project management for large-scale capital projects, and to balance the needs of various modes of transportation serving the New York metropolitan region. Today the MTA oversees NYCT's subway and city bus networks, two commuter railroads, a suburban bus network, numerous toll bridges and tunnels at key river crossings, and several large construction projects including the Second Avenue Subway, Manhattan's East Side Access, and the new Fulton Street Transit Center. NYCT provides day-to-day operational management of the subway and bus networks, including supervision of operating and maintenance crews, service planning, implementation of smaller capital projects (such as rolling stock procurement), and certain back-office management functions [1].

NYCT's predominant role is to ensure that trains and buses operate safely, reliably, on-time, and provide convenient services to the customer in a cost effective manner. One of the tools we have to ensure the mission is being properly carried out is an independent performance audit infrastructure—independent from both the operations management and the customer advocacy groups—and continuous applied research and improvements in not only monitoring methodologies but also how the service can be improved.

Our experience with deferred maintenance and service deterioration in the 1970s serves as a cautionary tale of how service quality, and eventually system maintenance can quickly deteriorate in the absence of an independent performance monitoring program. Furthermore, performance monitoring can provide the needed justification and basis for obtaining the funds and grants needed for continued system operations and upkeep.

About New York City Transit

NYCT operates the third largest subway system in the world (by annual ridership), carrying about 5.0 million riders on an average weekday. The subway system extends 830 track miles through four boroughs, covering a service area of 321 square miles and serving a population of 8.0 million people 24-hours, seven days a week. The subway is equipped with 6,375 electric passenger cars stored in 13 yards and two heavy maintenance facilities, travelling a combined total of over 354 million miles a year on 24 routes and three permanent shuttles. The system operates over 468 stations with a total of 5,105 stairways, 906 platforms, 192 elevators, and 178 escalators.

On the bus side, the 199 local and 24 express routes serve 11,995 stops and provide almost 55,000 weekday scheduled trips, carrying 2.4 million weekday riders throughout the 1,852 mile route network. The 1.83 billion passenger miles consumed by New Yorkers each year in 926 million discrete trips requires a fleet of 4,406 buses maintained in 19 depots by nearly 17,000 dedicated NYCT Bus employees.

About the Subway Service Performance Indicator (PI) Program

The Performance Indicator (PI) program was established in 1994 in response to the MTA Inspector General's research [2] recommending the need for measures of service reliability other than the traditional Terminal On-Time Performance (TOTP). A detailed description of the NYCT PI program is found in Cramer et al. [3]. TOTP is a good operational measure for commuter railroads where the majority of customers are traveling to the final stop in the central business district (CBD). However, transit lines tend to drop off and pick-up many passengers at intermediate stations, which requires more sophisticated measures capable of blending waiting time and travel time experiences from a customer perspective. Extensive research had been conducted to understand transit service reliability from the passenger and the transit manager's perspectives [4], building on prior models of headway variance [5,6].

Statistical measures of service reliability, such as root-mean-squared average passenger wait time [7], were considered too complex for use as public measures. NYCT developed simplified version of the algorithms that are more easily understood. The result was Wait Assessment (WA). This design of the PI program achieves a duality of purposes:

1. To provide the public with measurements that are clearly defined, easily verifiable, readily understandable, and realistically represent the many factors that impact their riding experience, such that NYCT can be held accountable to its core mission.

2. To provide quantitative information to operating personnel that can be used to diagnose and correct service problems, and to improve overall performance.

All subway lines are monitored. Principal bus routes representing the greatest ridership, demographic, and geographic areas within each borough of the city were selected. A stratified statistical sample, designed to prevent sample bias by route, is generated using a fully automated system and achieves an accuracy of 95 ± 5% at the route level.

Wait Assessment Development

The PI program's main purpose is to monitor how well NYCT is providing service to the public [8]. Wait Assessment is publicly reported at the systemwide, bus by route and subway by line level. These results are routinely used by rider advocacy groups for their annual rating of subway lines in their *State of the Subways* report. Maintaining a transparent and accountable performance reporting process is critical to achieving public trust in the performance audit infrastructure. Indeed, stakeholder and watchdog groups have adopted the MTA's measures as the basis of its performance reporting:

> *The MTA and its Operating Agencies provide some of the most transparent and detailed operational metrics among U.S. transit agencies; and this information is readily available on the MTA website. With respect to MNR and LIRR, no major commuter railroad comes close to their level of operational performance disclosure, especially with the recent addition of metrics on delayed and canceled trains in Board materials and on the website. In addition, the NYCT is to be lauded for the improvement of its performance indicators over the last 15 years, particularly with the implementation and refinement of its Wait Assessment metric.* [9]

Wait Assessment applies an analytical algorithm on raw data collected by surveyors (including fleet number and departure time of all vehicles passing a location.) Data is collected at en-route timepoints (not just terminals), because the majority of riders enter and depart the system at intermediate stops along the route.

As it was first conceived, Wait Assessment was an absolute measure of relative performance. It's an absolute measure because the thresholds of what constitutes an acceptable excess wait time [10] is a fixed quantity for a given time period (+2 minutes peak/+4 minutes off-peak). However, it is a measure of relative performance because it is based on headways between trains—obtained by comparing a train's departure time with its predecessor, and not by comparing a train's departure time with the fixed schedule. The rationale for this was to provide a customer with a "Bill of Rights", a fixed standard of excess wait time above which the service interval is considered unacceptable.

This type of measurement metric has one interesting property: lines scheduled with shorter headways tend to score higher, because there is simply a higher probability of a train—any train—achieving that two-/four-minute window above headway.

Refinement – Absolute versus Relative

In discussions with operations management, it became apparent that this property does not give dispatchers correct incentives. The lines that have high frequency service are often very congested, where the smallest perturbation in headways or ridership volume can quickly snowball into bunched service and big gaps [11]. On lower frequency lines, dispatchers have a little more latitude to adjust schedules, and the headway isn't so critical to maintaining proper service. To prevent such imbalance, the standard to which each line is held must be a function of service frequency, with busier lines held to more exacting standards. In turn, the tolerable excess wait time must be specified relative to the headway.

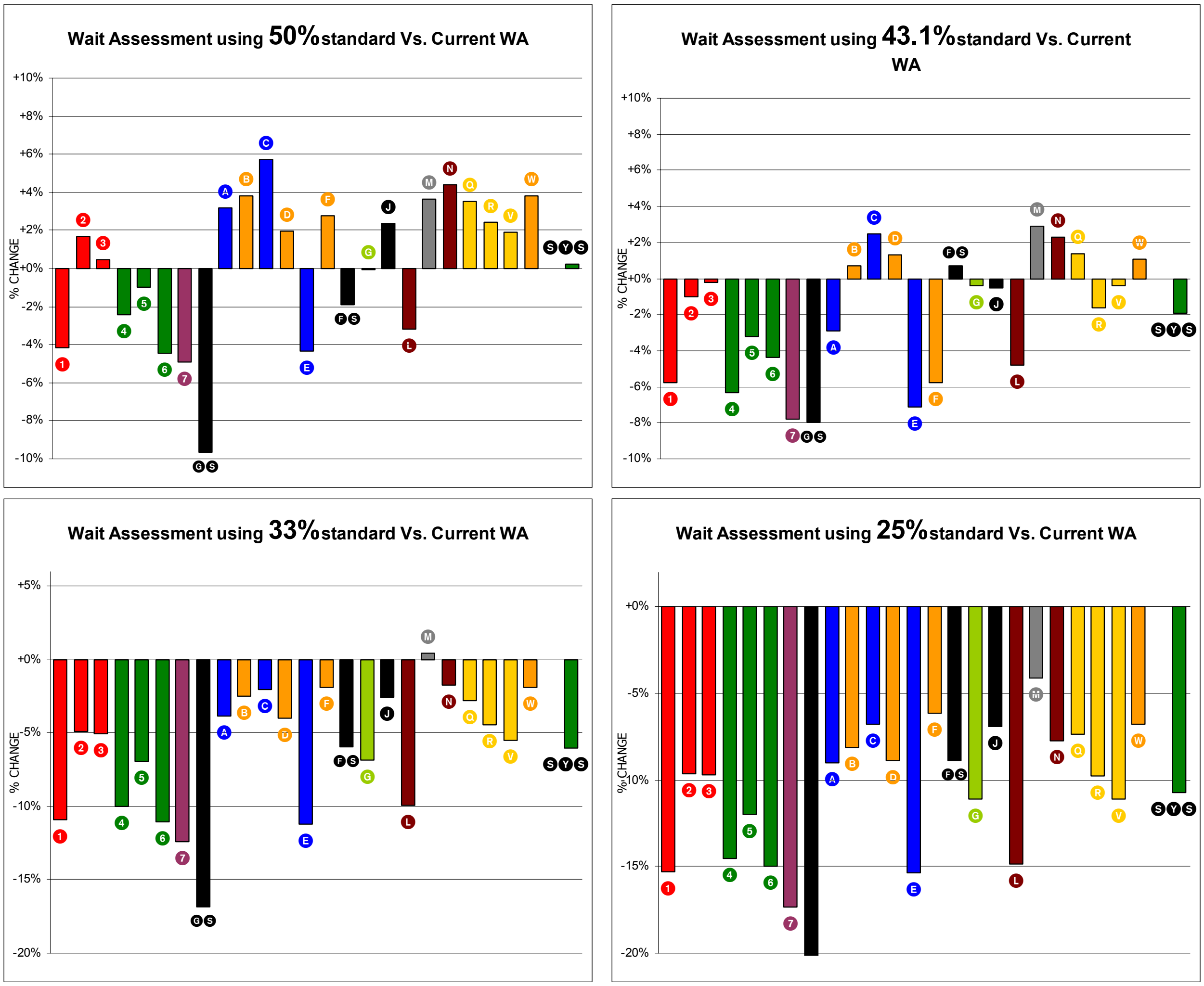
To determine an appropriate percentage, the difference between a +2/+4 standard and a percentage standard was computed at a line level at numerous levels (Figure 1). As NYCT's typical headway is 4 minutes peak, 8 minutes off-peak, a standard of headway +50% seems reasonable. Indeed, Figure 1 shows a standard of 50% would leave overall scores approximately unchanged, lowering scores on frequent lines like 1, 4, 6, 7, and E, while improving scores on infrequent lines like B, C, and N. Because of differences in total train volumes between lines, historical data shows that a 50% standard would actually increase systemwide Wait Assessment scores slightly, and a standard of 43.1% would actually maintain the systemwide result compared to while re-distributing pass and fail scores amongst different lines.

After much consultation with stakeholders and management, the final decision was reached to modify the Wait Assessment threshold to be within +25% of the scheduled headway, thereby making Wait Assessment a relative performance measure. The strictest standard at +25% was selected in an effort to assure the public that NYCT is looking for continuous improvements in service delivery. The formal definition of Wait Assessment for subway lines thus became:

% of intervals between trains that does not exceed 25% of scheduled headway.

For example, for a scheduled headway of four minutes, an actual headway of less than five minutes would be permissible. It essentially measures how evenly distributed actual service is in relation to schedule. Gaps in service (widely spaced intervals between departures) are penalized by this measure. Each scheduled interval is matched to an actual observed headway; only one observed headway is compared with each scheduled interval. Scheduled intervals that contain no actual service are considered automatic failures.

Figure 1 Wait Assessment with Relative Excess Wait Time—Options



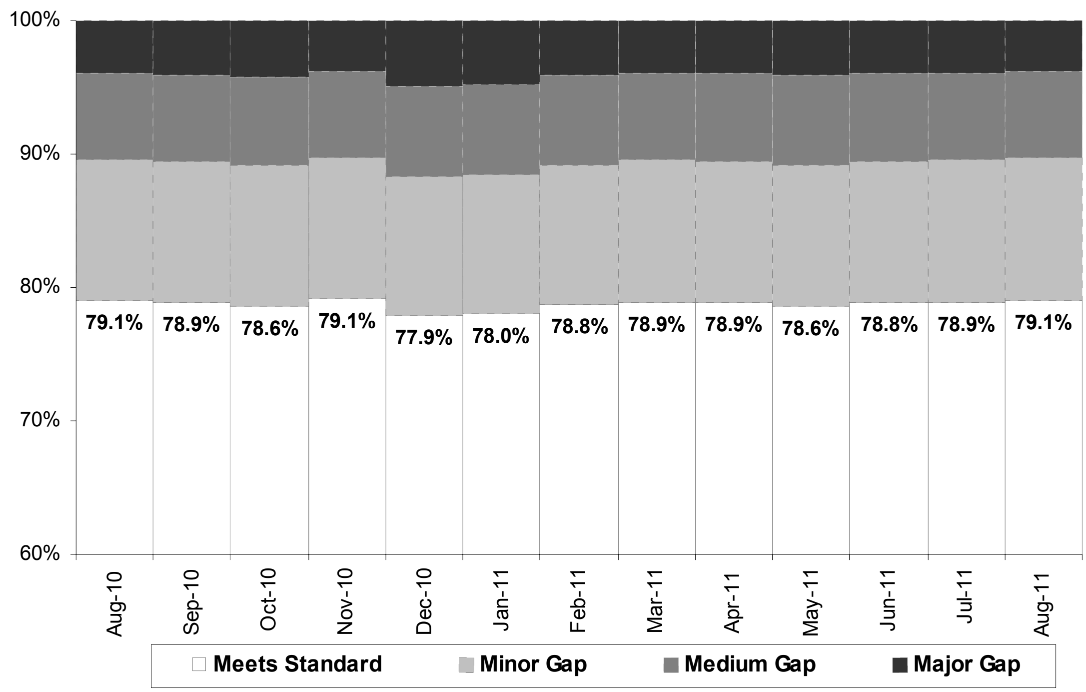
Refinement – Distribution of Failing Intervals

Wait Assessment performance measure shows results of the tighter Headway +25% standard, however, it does not give any information on the distribution of service intervals not meeting standard. The refinement to use a distribution for failing intervals provides customers a more detailed view of system performance without modifying the existing standard. Peak time periods will now be included to allow performance to be monitored during maximum ridership. Weekend Wait Assessment now reporting for all lines with prior year data for comparison and will also utilize new reporting format. These report formats are planned to be issued in January, 2012.

While the public welcomed NYCT's efforts to raise the bar on its performance, the line managers felt that simply knowing 20% or more of their trains were more than 25% outside the scheduled headway did not help them pinpoint the source of the bunching problems, which often started with an overcrowded train that sometimes became more than double the headway later than its predecessor.

To assist the management, the "failing" Wait Assessment intervals were further broken into three subcategories: 25%-50% more than headway, considered a "minor gap", 50%-100% above headway, a "medium gap", and more than double the headway, a "major gap". This type of reporting clearly identifies a bunching and spacing condition, and helps the management to identify lines and times-of-day when many major gaps in service occur—which could be pro-actively filled in with an appropriately named "gap train". For the public, this provides a multi-dimensional picture of actual service quality delivered.

Figure 2 Wait Assessment Results with Distribution



Proposed Refinement – Corridor Level Wait Assessment

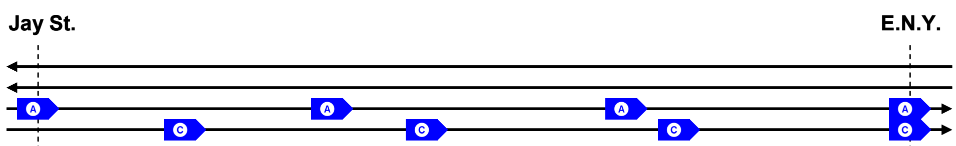
New York has many subway routes, some of which are co-routed on the same physical line infrastructure for major sections. The Queens Boulevard Line, which is a four-track major subway corridor in Queens actually hosts two local routes, the "R" and "M" Trains, and two express routes, the "E" and "F" Trains. In terms of track-sharing, the local routes share the local tracks, which makes all stations stops, and the express routes share the express tracks that has station platforms only at major transfer points.

This creates a dilemma when measuring Wait Assessment at stations that are served by more than one route. At those stations, customers have a choice as to which route they would like to use. We believe relatively few customers exercise this choice, based on our origin-destination modeling studies; many will simply take the first train that arrives, using it to get as close to their destination as the route permits, and make a transfer to a different train later on in their journey. Yet other customers prefer a one-seat ride and will wait on the platform for the exact route they require. Wait Assessment is a route-based measure; that is, it measures the intervals between trains of the same route—and does not consider the intervals between trains that are sharing the same track if they are assigned different route letters or numbers.

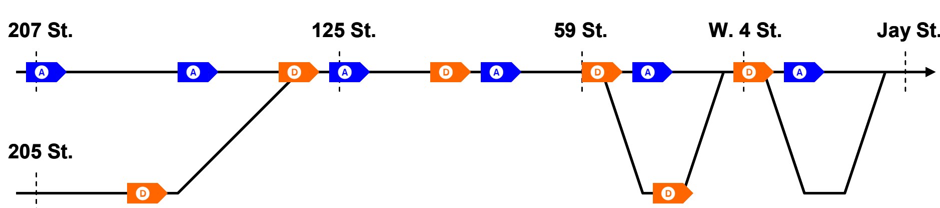
There is a debate about how such shared-track corridors should be managed. Customers who are destined for the branching section of trunk lines often require a specific route to reach their final destination, and therefore are interested in knowing if their route is having bunching and spacing problems. Customers who use the trunk section exclusively, or make inter-divisional transfers at major transfer points, are usually more interested in the corridor-level measures because routes letters or numbers are only of passing interest to them—for they simply require a train going in the same direction they're travelling. When routes share track, it is operationally important to keep even spacing between trains even if they have different route letters or numbers, therefore line management usually are more interested in track-level measures.

After much discussion, it was determined that Wait Assessment, which concerns the spacing between subsequent trains, only makes sense on trunk sections where different routes share the same physical track while passing through multiple timing points. The reasoning is as follows:

1. On four-track infrastructure where there is only one local route and one express route (e.g. the "A" and "C" Trains on the Fulton St. Line in Brooklyn), it does not make sense to measure intervals between "A" and "C" Trains because the different running times along local and express tracks will make "A" and "C" Trains appear bunched up at one timing point even if they were perfectly spaced at the previous timing point. In fact, having "A" and "C" Trains arrive and depart at the same time from a timing point (i.e. "bunched", pulsed, or banked departure) is actually a timetable feature—this is done at major transfer points to facilitate cross-platform transfers between local and express trains, such as at Broadway Junction in East New York.



2. Where two routes share one track for a short segment consisting of two timing points or less (e.g. the "A" and "D" Trains on the 8 Avenue/Central Park West Line express tracks in Manhattan), it does not make sense to measure the intervals between them because the facility is not necessarily operated as a corridor. From the passenger perspective, the choice of "A" and "D" Trains depends on their ultimate destination. Southbound, those who are headed to 8 Avenue or Downtown will choose an "A" Train even if a "D" Train was the first to arrive; those headed to 6 Avenue or Lower East Side will choose the "D". Northbound, Bronx-bound passengers use "D" exclusively while Upper Manhattan-bound passengers use "A". Operationally speaking, the Central Park West express tracks are used to get trains as quickly as possible from 53 St. Junction (59 St. timing point) to 135 St. Junction (125 St. timing point) on a first-in-first-out basis, rather than being held to specific intervals.



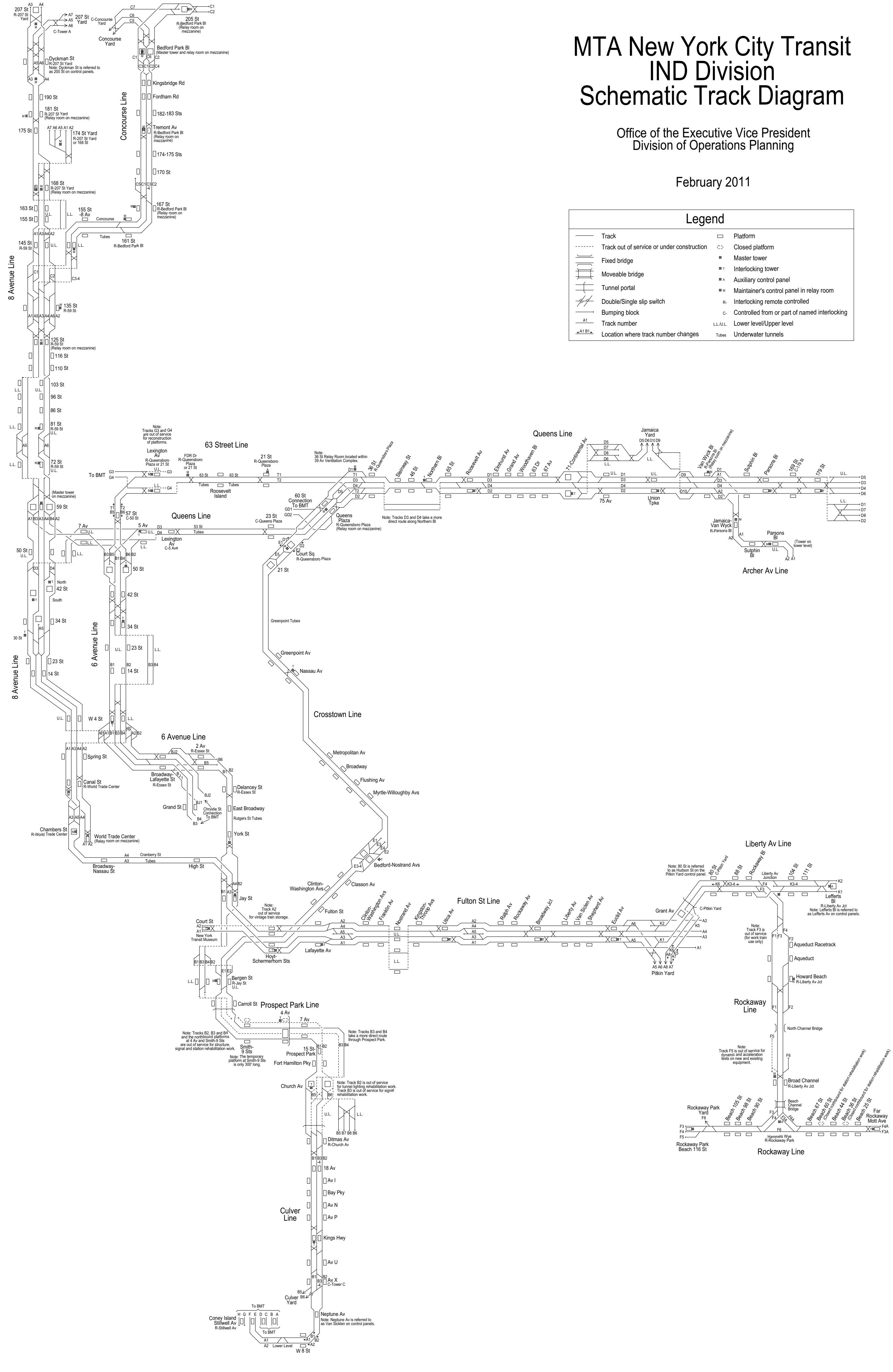
MTA New York City Transit
IND Division
Schematic Track Diagram

Office of the Executive Vice President
Division of Operations Planning

February 2011

Legend

- Track
- Track out of service or under construction
- Fixed bridge
- Moveable bridge
- Tunnel portal
- Double/Single slip switch
- Bumping block
- Track number
- Location where track number changes
- Platform
- Closed platform
- Master tower
- Interlocking tower
- Auxiliary control panel
- Maintainer's control panel in relay room
- R- Interlocking remote controlled
- C- Controlled from or part of named interlocking
- L.L./U.L. Lower level/Upper level
- Tubes Underwater tunnels



Based on these criteria, only three major corridors were identified where it made sense to measure Wait Assessment on a corridor level. Those corridors are:

1. Lexington Ave IRT express tracks (host to "4" and "5" Trains) between 125 St., Manhattan and Franklin Av., Brooklyn, with timing points at 125 St., 86 St., 42 St., Brooklyn Bridge, Bowling Green, Borough Hall, Flatbush Av., and Franklin Av.

2. West Side IRT express tracks (host to "2" and "3" Trains) between 135 St-Lenox, Manhattan and Franklin Av., Brooklyn, with timing points at 135 St., 96 St., 42 St., Chamber St., Borough Hall, Flatbush Av., and Franklin Av.

3. Queens Boulevard Line (host to "E" and "F" Express Trains, and "M" and "R" Local Trains) between Continental Av., Queens, and 36 St. Interlocking, Queens, with timing points at Continental, Roosevelt, and 36 St.

All other shared-track segments were too short to be properly operated as a corridor, and customer train choice tend to be route based. Because of the availability of automatically collected signal system data, the West Side IRT Corridor was chosen as a test case and the Wait Assessment results were computed using different algorithms.

Figure 4 Corridor-Level Wait Assessment Pilot Results

	❷		❸		❷❸ Total as Separate Routes		❷❸ Assessed as Combined Corridor	
	Old	Current	Old	Current	Old	Current	Old	Current
Passing Intervals	1,355	1,215	1,357	1,224	2,712	2,439	2,823	2,261
Failing Intervals	359	499	332	465	691	964	580	1,142
Percentage	**79.1%**	**70.9%**	**80.3%**	**72.5%**	**79.7%**	**71.7%**	**83.0%**	**66.4%**

The results are interesting in several ways. The prior Wait Assessment (+2/+4) improved in the combined corridor compared to the average of separate routes, because the +2/+4 minutes margin is applied to a more frequent service, making it easier to pass that standard. However, the current Wait Assessment (+25%) declined, because the +25% margin on a combined corridor with higher effective service frequency is tighter compared to if the services were considered two separate routes. At this time, NYCT is continuing to evaluate how best to analyze and report corridor-level results.

The Wait Assessment "Reach and Match" Algorithm

For all actual observations of consecutive trains passing timepoints, every actual departure interval must be matched to scheduled intervals based on the daily schedule with supplements applied. The matching process is governed by the "Reach and Match" algorithm, given below. The matches are made for ***timepoint*** locations only

This algorithm was developed to account for the "drift" that occurs as actual service intervals move out of their scheduled slots in normal daily operations. The intent of the algorithm is to describe the service intervals as experienced by customers who are expecting a train to arrive every *n* minute, where *n* is the headway specified in the timetable.

1. For each date headway is processed separately for each unique grouping of: Line, Timepoint and Direction, sorted by the scheduled departure time.

2. For the first interval, the first scheduled departure time after midnight is the starting point for Schedule (i). Increment Actual (j+1…) until departure time >= Schedule (i) departure time.

3. All interval matches are made by checking if the departure time first train of the Actual (j) interval is within the range of current schedule departure time, from "Min" (scheduled interval start [i] - scheduled interval length [i to i+1]) to "Max" (the first train departure time of the next scheduled interval [i+1])

   a. If (j) fits, it is tagged as "Matched—Within Reach"
   b. If Actual (j) does not fit, increment forward Actual (j+1...) interval until fit is found in "Min" / "Max" range or having run out of actual observations.
   c. If (j+1…) does not fit, check if previous (j-1) fits in "Min" / "Max" range
      i. If yes, check if (j-1) was used to calculate a previous Scheduled interval ("Already Used").
      ii. If Already Used, this indicates a "Matched—Repeat Interval" could potentially exist. To determine best match, a "Half Interval" test is applied to see if the actual interval (j to j+1) is between ±50% of the first matched Scheduled interval (i to i+1), or 5 minutes, whichever is less. (j-1) interval can be used (called a "Matched—Repeat Interval") ***only*** if the "Half Interval" test fails, signalling that although the previous Scheduled interval is a technical match, the current Scheduled interval is a better match to the Actual interval being considered.
      Generally speaking, Actual intervals that pass the Half Interval test against a specific Scheduled interval ("Matched—Best Possible") cannot be used again to make a match, since it has already been granted the best match, while those that do not pass the Half Interval test can be reused. In other words, a single vehicle that has successfully picked up passengers cannot be evaluated again, however a vehicle with a long actual interval may fail to pick up passengers in multiple scheduled (i, i+1…) intervals.
      iii. If no, then the interval "Autofails"

   The intent of this "Reach" criterion is to prevent actual and scheduled intervals from drifting too far away from each other. While up to one interval of give is allowed, to account for operational schedule adjustments (called the "flex"), as soon as the Actual intervals drift out-of-sync with the timetable, it is considered "Out-of-Reach" and is not used to make a "Match".

4. "Autofail" (2), or automatic failure, is a Scheduled interval with no Actual interval match having exhausted all of the possible actual observations. This can happen if:

   a. Current, Next and Previous Actual interval departure times are not within the Min and Max (i+1) "Reach" boundary.
   b. Only the Previous Actual (j-1) interval fits, but has been used and was a "Matched—Best Possible" in another (Previous) scheduled interval. An actual interval can only be credited once, i.e. deemed "Matched—Best Possible" of one specific scheduled interval.
   c. An Autofail record will have a "Null" actual headway.

5. After each "Match" result is computed (Matched—Within Reach, Matched—Best Possible, Matched—Repeat Interval, and Automatic Failure), a Wait Assessment (WA) result is calculated using the current interval:

    a. If interval matching Autofails, then WA also Autofails
    b. If Actual interval is a "Matched—Repeat Interval" (j = j-1), WA fails if WA for the Previous (j-1) interval was "Pass". This is to prevent the same Actual interval for being credited against two Scheduled intervals.
    c. Else, Calculate Wait: Wait Assessment Pass/Fail Criteria:
        i. Actual interval is less than or equal to 25% of Scheduled interval: Pass
        ii. Actual interval is greater than 25% of Scheduled interval: Fail

6. This process continues until the last scheduled departure (i) of the line, direction and timepoint group, which is neither evaluated nor stored, since (i+1) is Null.

Wait Assessment Detail Results

Figure 5 shows an extract of Wait Assessment results from NYCT's West Side IRT local track, which hosts the "1" Train. Each train is assigned a TRAIN_ID, which indicates the line number of the train (01), the departure time from the origin (1504, 1509, etc.), followed by a plus sign if the train is scheduled to depart on the half minute (e.g. 1509+ = 15:09:30), and also the codes for origin terminal (SFT) and destination terminal (242). SFT is South Ferry in Lower Manhattan, and 242 is the street number code for Van Cortland Park Terminal located on 242 Street in Upper Manhattan.

Figure 5 Signal System based Wait Assessment Detail Results

SVC_DATE	SCHD_TRAIN_ID	STA_NAME	SCHD TIME	SCHD HDWY	ACT TIME	ACT HDWY	WA24	WA25
20110329	01 1504 SFT/242	59 ST	152530	330	152247	276	PASS	PASS
20110329	01 1509+ SFT/242	59 ST	153100	330	152723	477	PASS	FAIL
20110329	01 1515 SFT/242	59 ST	153630	330	153520	220	PASS	PASS
20110329	01 1520+ SFT/242	59 ST	154200	330	153900	502	PASS	FAIL
20110329	01 1526 SFT/242	59 ST	154730	330	154722	314	PASS	PASS
20110329	01 1531+ SFT/242	59 ST	155300	330	155236	140	PASS	PASS
20110329	01 1537 SFT/242	59 ST	155830	330	155456	617	FAIL	FAIL
20110329	01 1542+ SFT/242	59 ST	160400	330	160513	156	PASS	PASS
20110329	01 1548 SFT/242	59 ST	160930	330	160749	321	PASS	PASS
20110329	01 1553+ SFT/242	59 ST	161500	330	161310	383	PASS	PASS
20110329	01 1559 SFT/242	59 ST	162030	330	161933	375	PASS	PASS
20110329	01 1604+ SFT/242	59 ST	162600	330	162548	209	PASS	PASS
20110329	01 1609+ SFT/242	59 ST	163130	300	162917	325	PASS	PASS
20110329	01 1614+ SFT/242	59 ST	163630	270	163442	282	PASS	PASS
20110329	01 1618+ SFT/242	59 ST	164100	300	163924	370	PASS	PASS
20110329	01 1623+ SFT/242	59 ST	164600	300	164534	98	PASS	PASS
20110329	01 1628 SFT/242	59 ST	165100	300	164712	386	PASS	FAIL
20110329	01 1633 SFT/242	59 ST	165600	300	165338	314	PASS	PASS
20110329	01 1638 SFT/242	59 ST	170100	300	165852	190	PASS	PASS
20110329	01 1642+ SFT/242	59 ST	170600	300	170202	296	PASS	PASS
20110329	01 1647 SFT/242	59 ST	171100	270	170658	529	FAIL	FAIL
20110329	01 1651+ SFT/242	59 ST	171530	240	171547	103	PASS	PASS
20110329	01 1655+ SFT/242	59 ST	171930	240	171730	222	PASS	PASS
20110329	01 1659+ SFT/242	59 ST	172330	240	172112	296	PASS	PASS

Based on the algorithm described previously, the schedule is first used to work out the scheduled headway, and matched to the actual time and actual headway. The "match" part of the algorithm determines whether the interval is within reach, and some intervals that are too narrow or too wide easily falls out of reach (e.g. the 1520+ departure from SFT was out of reach by the time the train had moved north to 59 St.) Initially, Wait Assessment was defined as +2/+4 minutes, with results shown in the WA24 column. Note that it is possible for an interval to be out-of-reach yet for Wait Assessment to pass, because service was provided during that time. When the stricter "+25%" standard came into effect, with results shown in the WA25 column, there were more failed intervals.

Daily Reporting

Upgrading the previous sample-based method that gathered limited data manually, extensive data was downloaded from the Automated Train Supervision (ATS) system to provide 100% coverage and much lower time-lag for compiling performance measures. This allowed near-term corrective action by operations supervisors. Daily "outlier" reports are issued to assist managers to identify troublespots.

Figure 6 shows a daily summary report, with all information about a single train line's performance summarized on one page. For each hour and for each timing point location, the Wait Assessment result during that hour is given, as is the throughput (count of trains passing that location during that hour). This report allows a line manager to see at a glance how the line performed yesterday—and more impotantly, if there were an incident on the line, what the performance impact was for that incident.

Figure 6 Daily Flash Report by Hour and Location

Daily Wait Assessment by Hour and Location DRAFT

Service Date **11/03/2011** *Line* ❶

Direction S **Hourly Results at Gap Locations (Wait Assessment 25%/Thruput)**

	AM Peak (0600-0859)			Off-Peak (0900-1559)							PM Peak (1600-1859)			Off-Peak (1900-2359)				
Station	**06**	**07**	**08**	**09**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
51020 238 ST	86%/7	60%/10	89%/9	91%/11	92%/12	30%/10	90%/10	60%/10	100%/10	92%/12	67%/12	83%/12	67%/12	91%/11	80%/10	89%/9	100%/6	100%/5
51100 168 ST	67%/9	47%/15	77%/13	90%/10	75%/12	30%/10	90%/10	70%/10	100%/10	83%/12	87%/15	75%/12	75%/12	92%/12	70%/10	80%/10	100%/6	100%/5
51130 137 ST	67%/9	50%/14	64%/14	70%/10	75%/12	20%/10	100%/10	90%/10	90%/10	83%/12	79%/14	79%/14	58%/12	92%/12	70%/10	78%/9	100%/6	100%/6
51170 103 ST	86%/7	50%/14	61%/18	54%/13	67%/12	0%/10	100%/10	90%/10	70%/10	92%/12	64%/14	69%/13	58%/12	92%/12	82%/11	67%/9	100%/7	100%/5
51210 96 ST	71%/7	36%/14	72%/18	50%/12	83%/12	10%/10	100%/10	100%/10	80%/10	75%/12	73%/15	77%/13	83%/12	83%/12	80%/10	90%/10	83%/6	80%/6
51250 66 ST	86%/7	46%/13	72%/18	54%/13	83%/12	0%/10	100%/10	80%/10	80%/10	58%/12	79%/14	64%/14	67%/12	67%/12	80%/10	80%/10	86%/7	75%/5
51280 42 ST	88%/8	45%/11	84%/19	54%/13	92%/12	0%/11	100%/10	100%/10	90%/10	91%/11	86%/14	50%/14	67%/12	67%/12	82%/11	80%/10	83%/6	80%/6
51380 CHAM ST	71%/7	40%/10	71%/17	73%/15	67%/12	18%/11	50%/10	90%/10	90%/10	91%/11	77%/13	67%/15	67%/12	67%/12	73%/11	80%/10	75%/8	75%/6

Direction N **Hourly Results at Gap Locations (Wait Assessment 25%/Thruput)**

	AM Peak (0600-0859)			Off-Peak (0900-1559)							PM Peak (1600-1859)			Off-Peak (1900-2359)				
Station	**06**	**07**	**08**	**09**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**
51380 CHAM ST	100%/6	50%/10	53%/15	82%/17	83%/12	55%/11	30%/10	100%/10	100%/10	100%/11	100%/12	73%/15	85%/13	83%/12	82%/11	100%/10	100%/9	100%/6
51280 42 ST	50%/6	88%/8	57%/14	56%/18	100%/12	50%/12	20%/10	80%/10	100%/10	90%/10	75%/12	60%/15	92%/13	83%/12	83%/12	90%/10	80%/10	100%/6
51260 59 ST	40%/5	67%/9	54%/13	67%/18	92%/13	45%/11	10%/10	80%/10	90%/10	82%/11	73%/11	67%/15	77%/13	83%/12	83%/12	82%/11	67%/9	100%/7
51210 96 ST	33%/6	86%/7	54%/13	61%/18	77%/13	67%/12	10%/10	70%/10	80%/10	70%/10	67%/12	86%/14	64%/14	83%/12	100%/12	80%/10	70%/10	83%/7
51130 137 ST	40%/5	71%/7	56%/9	53%/17	64%/14	75%/12	20%/10	60%/10	80%/10	70%/10	73%/11	86%/14	64%/14	75%/12	100%/12	82%/11	90%/10	50%/7
51100 168 ST	50%/4	63%/8	38%/8	59%/17	64%/14	42%/12	40%/10	60%/10	90%/10	70%/10	64%/11	93%/14	57%/14	75%/12	100%/12	82%/11	80%/10	86%/8
51070 DYCK ST	50%/4	86%/7	33%/9	60%/15	67%/15	67%/12	30%/10	50%/10	80%/10	80%/10	55%/11	100%/13	47%/15	75%/12	92%/12	83%/12	80%/10	83%/7
51030 231 ST	100%/4	71%/7	50%/8	58%/12	54%/13	58%/12	60%/10	30%/10	90%/10	60%/10	64%/11	92%/13	47%/15	92%/12	83%/12	75%/12	80%/10	86%/8
51020 238 ST	100%/4	71%/7	38%/8	58%/12	58%/12	55%/11	73%/11	30%/10	90%/10	82%/11	45%/11	83%/12	79%/14	82%/11	90%/10	58%/12	70%/10	86%/8

Figure 7 is a slightly different daily report designed to improve troubleshooting. Without detailed knowledge of the day's incidents, it can be difficult to determine whether a lower Wait Assessment score during a certain hour was typical, and whether problems were nogoing and recurring or if they're due to a specific non-repeating incident. Managers generally prefer to focus on recurring problems rather than unusual incidents. This report compares today's Wait Assessment statistics (at the hourly and location level) with the rolling average over the past 30 days (where data is available). The "low outlier" hour-and-location combinations are printed out, allowing managers to take corrective action.

Figure 7 Daily Outlier Report Indicating Worst Performing Locations

Daily Flash Report: Outliers* by Location by Hour DRAFT

Date* *11/03/2011 *Sample Period Begins* *10/05/2011*

Line and Service Type ❶ **Direction** S

Location	Time (Hour)	Today's WA25	Month Rolling Avg at This Location	# of Days
51100 168 ST	06 00	67%	91%	28
51130 137 ST	06 00	67%	90%	30
51210 96 ST	06 00	71%	89%	30
51020 238 ST	07 00	60%	88%	28
51100 168 ST	07 00	47%	78%	28
51130 137 ST	07 00	50%	89%	30
51170 103 ST	07 00	50%	85%	30
51210 96 ST	07 00	36%	86%	30
51250 66 ST	07 00	46%	83%	30

Line and Service Type ❶ **Direction** N

Location	Time (Hour)	Today's WA25	Month Rolling Avg at This Location	# of Days
51070 DYCK ST	06 00	50%	89%	28
51100 168 ST	06 00	50%	90%	28
51130 137 ST	06 00	40%	92%	30
51210 96 ST	06 00	33%	94%	30
51260 59 ST	06 00	40%	92%	30
51280 42 ST	06 00	50%	95%	30
51260 59 ST	07 00	67%	90%	30

* "Worst Case" Wait Assessment scores, i.e., lower than the 90th Percentile of the 30-Day Rolling Average by Location by Hour.

Friday, November 04, 2011 Page 1 of 9

Operational Impacts

In its capacity as an oversight mechanism, the PI program achieves continual performance improvements through the usual goal-setting and management processes. In addition to this auditing function, data is also used to specifically identify operations issues.

Monthly reports are made available about a week after the conclusion of the month, allowing managers to investigate reasons for performance changes prior to formal public reporting.

NYCT initiates bus dispatcher programs in response to declines in Wait Assessment, or community concerns about service reliability. Extra dispatchers are allocated to specific routes to monitor field performance and recommend operational changes if necessary. One dispatcher program on a key interborough bridge route led to a rewrite of operating schedule that improved reliability substantially. Another dispatcher program monitored outbound express bus routes, resulting in a partnership with local agencies to reconfigure street access to a key tunnel to reduce bus delays.

Wait Assessment can also alert subway management to suboptimal dispatching strategies. Early train departures can result in missed connections, uneven spacing, and poor performance. After reviewing reports showing many early trains, the district management instituted an initiative to hold all trains to time. This resulted in improved service reliability, which was substantiated by Wait Assessment improvements post-implementation.

Conclusions

MTA New York City Transit (NYCT) has responded to the challenge of "doing more with less" by re-tooling its performance measurement frameworks to better capture performance from a customer's perspective, respond to management system improvement initiatives, and better incentivize operating decisions that deliver excellent customer service.

Recent improvements to the WA standard have made it a more meaningful relative performance measure that is stricter for lines with more frequent service. The former pass/fail standard has now been replaced with by a distribution for failing intervals that provides customers a more detailed view of system performance. NYCT is continuing to improve its performance standards by understanding how WA could fairly and best applied to shared-track territories where different routes can be treated as the same service corridor and train performance analyzed without reference to route designation.

Upgrading a previous sample-based method that gathered limited data manually, extensive data was downloaded from the Automated Train Supervision (ATS) to provide 100% coverage and much lower time-lag for compiling performance measures. This allowed near-term corrective action by operations supervisors.

These improvements to NYCT's customer-centric service performance indicators were developed with extensive consultation with operations management, have been ratified by the MTA Board, and endorsed by stakeholders and public advocacy groups. In the tradition of improved reporting, NYCT will continue to explore new ways of assessing its own performance and reporting it for both internal diagnostic purposes and for public accountability.

Acknowledgements

The author would like to thank Alex Lu and John Cucarese for assistance in preparing this manuscript. Responsibility for errors or omissions remains with the author. Opinions expressed are the author's and do not necessarily reflect official policy of Metropolitan Transportation Authority or MTA New York City Transit.

References

[1] Reddy, Alla V., A. Lu, and T. Wang. Subway Productivity, Profitability, and Performance: A Tale of Five Cities. TRB Paper #10-0487. In *Transportation Research Records 2143*, pp.48-58, Transportation Research Board of the National Academies, Washington, D.C., 2010.

[2] Office of the Inspector General, State of New York. Regularity Indices for Evaluating Transit Performance. MTA/IG Technical Report 90-32. New York, N.Y., 1990.

[3] Cramer, Anthony, J. Cucarese, M. Tran, A. Lu, and A.V. Reddy. Performance Measurements on Mass Transit: New York City Transit Authority Case Study. In *Transportation Research Records 2111*, Transportation Research Board of the National Academies, Washington, D.C., 2009.

[4] Henderson, G., P. Kwong, and H. Adkins. Subway Reliability and the Odds of Getting There On-Time. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1297, Transportation Research Board of the National Academies, Washington, D.C., 1991, pp.10-13.

[5] Adebisi, O. A Mathematical Model for Headway Variance of Fixed Bus Routes. Transportation Research B, Vol. 20B, No. 1, 1986, pp 59-70.

[6] Henderson, G., P. Kwong, and H. Adkins. Regularity Indices for Evaluating Transit Performance. In Transportation Research Record: Journal of the Transportation Research Board, No. 1297, Transportation Research Board of the National Academies, Washington, D.C., 1991.

[7] Wilson, N. H. M., D. Nelson, A. Palmere, T. Grayson and C. Cederquist. Service Quality Monitoring for High-Frequency Transit Lines. In Transportation Research Record: Journal of the Transportation Research Board, No. 1349, Transportation Research Board of the National Academies, Washington, D.C., 1992, pp. 3-11.

[8] Kittelson & Associates Inc. *Transit Cooperative Research Program (TCRP) Project G-6 Interim Report: A Guidebook for Developing a Transit Performance Measurement System.* Transportation Research Board of the National Academies, Washington D.C., 2002.

[9] Permanent Citizens Advisory Committee to the MTA. *Research Report—Minutes Matter: A Review of Performance Metrics at the MTA*. New York, N.Y., 2011.

[10] Graham, Daniel, X. Liu, and M. Trompet. Development of Key Performance Indicator to Compare Regularity of Service Between Urban Bus Operators. Paper #11-0540, Presented at the Transportation Research Board 90th Annual Meeting, Washington, D.C., 2011.

[11] Wilson, Nigel H.M., R. Macchi, R. Fellows, and A. Deckoff. Improving Service on the MBTA Green Line through Better Operations Control. In *Transportation Research Record* 1361, pp. 296-304, Washington, D.C., 1992.