

# Application of Entry-Only Automated Fare Collection System Data to Infer Unlinked Trips and Passenger Miles

*Alla Reddy*

*Alex Lu, Santosh Kumar,*

*Victor Bashmakov, Svetlana Rudenko*

## System Data & Research Operations Planning New York City Transit

Presented at the 88<sup>th</sup> Annual Meeting of the  
Transportation Research Board  
Washington D.C. (2009)



# Urbanized Area Formula Program (§5307 UAFP)

**Federal Transit Administration (FTA) requires transit agencies to provide service statistics to the National Transit Database (NTD) annually, including:**

- **Passenger miles**
- **Unlinked trips**
  - service consumption statistics
  - major data collection effort
  - input to incentive portion of Section 15 formula appropriation
- **Operating expenses**
- **Revenue miles**
- **Fixed guideway miles**



# UMTA 1984 Sampling Methodology

- FTA approved plan requires 50 booths per strata
- Strata with large variations have more samples

Time Period	Sample Size
A.M. Peak (06:00-10:00)	150
Midday (10:00-15:00)	100
P.M. Peak (15:00-19:00)	150
Evening (19:00-24:00)	75
Night (00:00-06:00)	75
Saturday (00:00-24:00)	75
Sunday (00:00-24:00)	75
<b><u>Total</u></b>	<b>700</b>



# Manual Data Collection Issues

- **Random Sample**

- High costs
- Missed assignments

- **Low Survey Yield**

- Large passenger volumes
- Language barriers
- Non-responses
- Responses (2005)  
Surveyor: 59,442  
MetroCard: 195,283

- **Data Interpretation**

- ‘36 St.’: Bklyn or Queens?

- **Inconsistencies in Data Collection**

- **Unavailable Data**

- High Entry and Exit Turnstile (HEET)
- PayPass Contactless Smart Card

**OPERATIONS PLANNING**  
**SECTION 15 - SUBWAY DATA COLLECTION SHEET**

DAY: MON      NAME: MURRAY      PASS#: 31040      DATE: 08/28/06  
 SURVEY TIME: 800 TO 900      STATION: 42 ST-GRAND CNT      CONTROL AREA: R236K  
 LINE: 4,5,6,7,S      DIV: IRT      TIME PERIOD: AM

ARRIVAL TIME: 0740      DEPARTURE TIME: 0802

TURNSTILE READING AT 0750			TURNSTILE READING AT 0735		
TOKENS	CARDS	TOTAL ENTRY	TOKENS	CARDS	TOTAL ENTRY
1	10353	16978	1	40550	17022
2	Special Entry		2	Special Entry	
3	29100	07546	3	29100	071670
4	Not Rec'd		4	Not Rec'd	
5	40115	44151	5	40115	44513
6	58242	26742	6	58242	27258
7	3869	68019	7	3869	67997
8	62138	52158	8	62138	52616
9	10567	57939	9	10567	57939
10	70984	61003	10	70984	61003
11	37358	60007	11	37358	60007
12	89128	69278	12	89128	69278
13			13		
14			14		
15			15		
SP	67818	01695	SP	67518	01754
SP		19684	SP		19743

\*\*\* WHAT STATION WILL YOU BE GETTING OFF THE SUBWAY TO GO TO THE STREET? \*\*\*

#	TIME	STATION	MILES	TIME	STATION	MILES	TIME	STATION	MILES
1	0800	190th St	0805	190th St	0805	N/R			
2	0801	N/R	0805	Canal St	0805	Times Sq			
3	0801	57th - 7th	0805	66th Ave	0805	59th St			
4	0801	Wall St	0805	N/R	0805	N/R			
5	0801	Times Sq	0805	N/R	0805	Grand St (D)			
6	0801	Times Sq	0805	N/R	0805	Franklin (D)			
7	0801	Times Sq	0805	21st St - Penn St	0805	Houston St			
8	0801	Franklin St	0805	Wall St	0805	Wall St (2)			
9	0801	84 St - Penn St	0805	Times Sq	0805	80th St			
10	0802	86th - 110th Ave	0804	73rd - 7th	0806	N/R			
11	0802	Fulton St	0804	73rd - 7th	0806	Wall St			
12	0802	Times Sq	0804	Times Sq	0806	Franklin St			
13	0802	83 St	0804	84 St - Penn St	0807	Wall St			
14	0802	84 St - Penn St	0804	84 St - Penn St	0807	Houston St			
15	0802	Franklin St	0804	N/R	0807	Penn St - 34th			

NON-TURNSTILE PASSENGERS COUNT:

FOR OFFICE USE ONLY
TURNSTILE READING:
NON-TURNSTILE READING:
TOTAL RESPONSE:
TOTAL TRANSFER:
TOTAL MILES:

COMMENTS: TS #4 Not identifiable - City of Transit  
 TS 9-12 No entry - Train Exit Line only

NOTES: 1. INTERVIEW EVERYBODY IF POSSIBLE EVEN THOUGH THERE IS NO RESPONSE.  
 2. WRITE ALL NO RESPONSES (NR).  
 3. MAKE SURE THAT THE CORRECT TURNSTILE READINGS ARE RECORDED.  
 4. SP=SPECIAL ENTRY.  
 5. WHEN INTERVIEWING ENTERING PASSENGER, YOU MUST STAND INSIDE THE PAID CONTROL AREA.  
 6. YOU MUST WEAR YOUR VEST.



# Automated Fare Collection Benefits

- **Higher accuracy and consistent results**
  - Unique serial number on each MetroCard
- **Same source (AFC) for O/D data and annual registrations**
- **All turnstiles measurable**
  - Including HEET and Smart Card
  - Including part-time booths (HEET open, turnstiles closed)
- **Eliminates paper data collection**
- **Monthly ridership activity is consistent with annual submission**
- **Quality assurance processes**
  - electronic data is easier to retrieve for internal and external audits



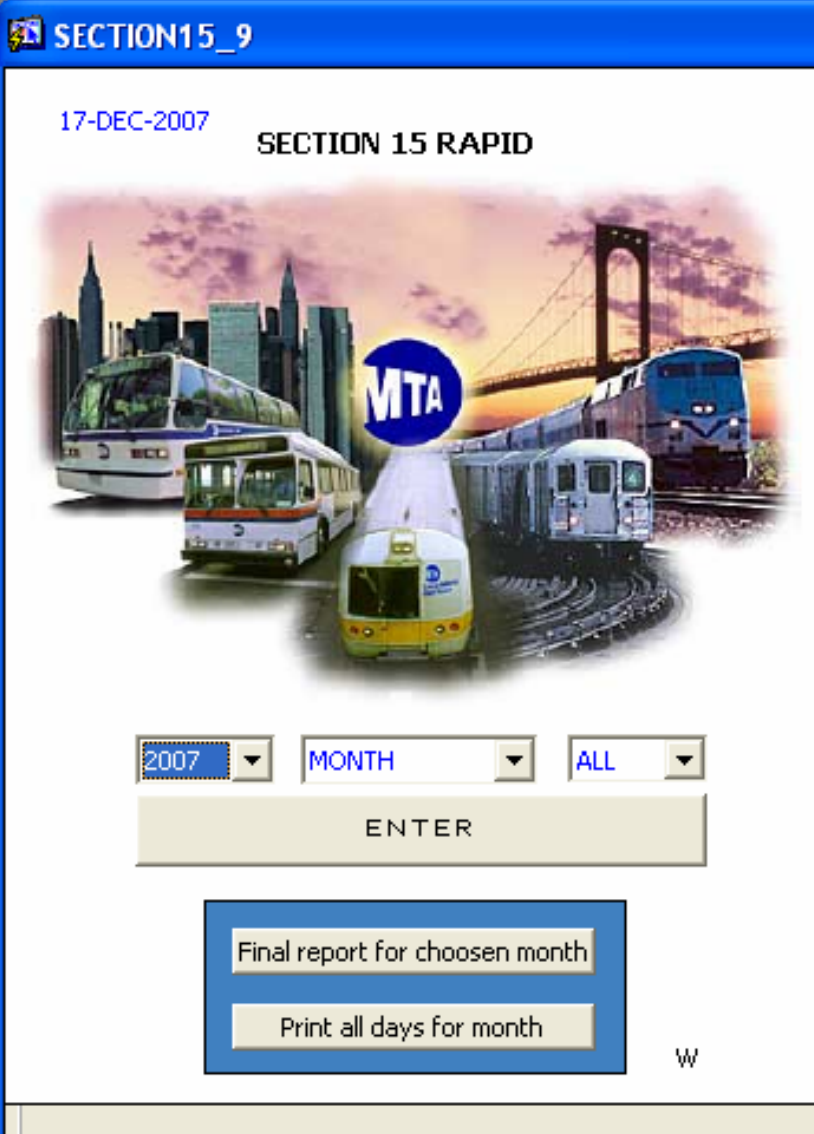
Canal St., IRT Lexington Av. Line



# Shortest Path Model Development

- **The shortest path program, developed in Oracle:**
  - Calculates all possible paths from origin to destination for each trip
  - Uses subway schedules in effect at time of survey to determine applicable service pattern
  - Looks up mileages for all paths using a distance look-up table built into Oracle
  - Selects shortest distance path\* with least transfers
- **Parallel-tested from 2004 to 2007 against original mileage calculation spreadsheet**

\* Where shortest path has more transfers than a longer path, lowest transfers per mile is picked. Where transfer count is equal, preference is given to transfers with shortest walking distance.



The screenshot shows a web application window titled 'SECTION15\_9'. The main content area features a collage of MTA transit vehicles, including a bus, a subway train, and a light rail train, with the MTA logo prominently displayed. The date '17-DEC-2007' and the title 'SECTION 15 RAPID' are visible at the top. Below the collage is a search form with three dropdown menus: '2007', 'MONTH', and 'ALL'. An 'ENTER' button is positioned below these menus. A separate box contains two buttons: 'Final report for chosen month' and 'Print all days for month'. A small 'W' icon is located in the bottom right corner of the interface.

# Daily Section 15 Processing

- **Download AFC data**
  - 600 megabytes of swipe data generated **daily**
- **Split-and-Merge**
  - MetroCard “revenue day” runs from 3 a.m. to 2:59 a.m. the following day
- **Origin-Destination (O/D)**
  - Infer travel destinations
  - Select swipes according to sample
  - Count registrations
- **Shortest Path Program**
  - Calculate trip distance and transfers
  - Compute passenger miles and unlinked trips



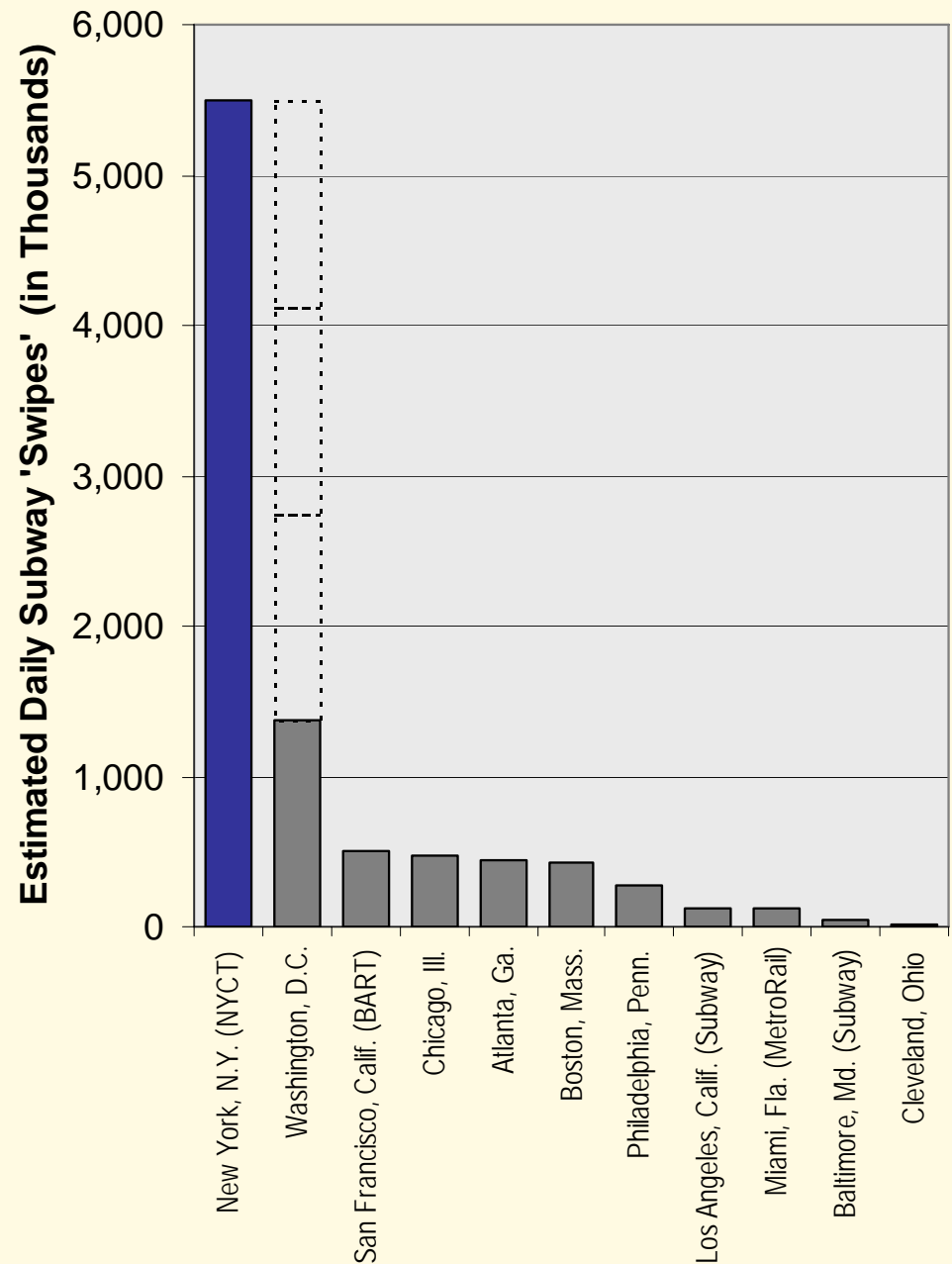
Lexington Av./53 St., IND Queens Blvd. Line

Photo: Richard Panse



# Data Processing Challenge

- By some estimates, New York has **four** times more 'subway swipes' per day than the nearest competitor, Washington D.C.
  - Even though WMATA makes you swipe twice for every trip!
- Mainframe generates **8.5 million records per day**
  - Not easily processed in Oracle
  - **Requires** sequential processing using compiled procedural language programs
  - **Requires** optimized search and sort algorithms



# Destination Inference Algorithm

- **MetroCard data record consists of:**
  - control area (station booth) of entry
  - transaction time, type of swipe (passes, special farecards)
  - fare paid and value remaining (Pay-per-Ride MetroCards)
- **Rider Origin and Destination is determined from MetroCard ‘swipes’:**
  - **Industry standard algorithm**
  - **Origin** is where the passenger swiped into sample booth control area
  - **Destination** is where the second-swipe into the subway occurred during the day



# Reasons for AFC Non-Responses

- **O/Ds of AFC Responses fairly represents O/Ds in AFC Non-Responses**
- **Single Trip in 24 Hour Period**
  - Destinations cannot be inferred, but also theoretically has different travel pattern
  - Accounts for about 5% of daily ridership, including:
    - Very Short Subway Rides (returning via bus or walk)
    - Very Long Multi-Day Trips
    - One-way Express Bus Trips
    - Other Carriers (PATH, AirTrain)

Reason for AFC Non-Response	Approximate % of Subway Ridership	Representative Sample in AFC Responses
Special Tickets and Passes	< 5%	Yes
Fare Abuse	< 3%	Yes
Multi-Person Pay-per-Ride	~ 4%	Yes
New Farecard Purchased	~ 5%	Yes
<b>Single Trip in 24 Hour Period</b>	<b>~ 5%</b>	<b>No</b>

215 St./10 Av., IRT Tenth Av. El.



# Approvals Process

- **Sample highly prescriptive**

- No change to sample
- New data collection technology (AFC)

- **Progressive scope expansion**

- Two pilot months (May and October 2005)
- All of 2005, then 2006

## SECTION 15 SUBWAY DATA: JANUARY - DECEMBER 2005 & 2006

### PASSENGER MILES & UNLINKED TRIPS

#### Surveyors vs. MetroCard Origin/Destination Data

		Total Registration	Total Registration Change (2005-2006)%	Total Response Count	Total Turnstile Entry Count	Passenger Miles	Passenger Miles % (Metrocard O/D Data To Surveyor's Data)	Passenger Miles Change (2005-2006)%	Unlinked Trips	Unlinked Trips % (MetroCard O/D Data To Surveyor's Data)	Unlinked Trips Change (2005-2006)%
2005	2005 Surveyor's Data	1,468,463,274		59,442	273,536	8,402,147,333			1,804,034,331		
	2005 MetroCard O/D Data			195,283	244,922	8,984,910,917	<b>6.94%</b>		2,172,124,952	<b>20.40%</b>	
2006	2006 Surveyor's Data	1,518,946,884	3.4%	63,874	293,973	8,338,041,853		-0.8%	1,874,980,539		3.9%
	2006 MetroCard O/D Data			207,985	256,096	9,381,650,838	<b>12.52%</b>	4.4%	2,272,287,243	<b>21.19%</b>	4.6%

- **Close FTA supervision**

- **Full parallel testing documentation**

- Flowchart, summary comparisons, description of procedure
- Full survey submission and proposed AFC submission
- Side-by-side comparisons including sample survey and AFC data for same booth, same date and time



# FTA Special Request

- In early January 2007, FTA approved AFC data for NYCT Section 15 conditionally on:
  - **Complete re-submission of revised NTD monthly data** based on AFC sampling for 2007 by January 31
- **To meet this challenge:**
  - Section 15 software installed on many PCs, for parallel computation
  - Office was staffed round-the-clock
    - Enterprise server and network offered better performance during off-hours
  - Calculations complete after 14 days
- **Starting 2008, NYCT submits AFC unlinked trips data monthly**



Seneca Av., BMT Myrtle El.

Photo: Fred Guenther



# Lessons Learned & Further Applications

- **Lessons Learned**

- Use Smaller Dataset During Development
- Involve Regulatory Authorities and Data Consumers
- Expect Full Parallel Test
- Design for Scaling Up, Measure Performance Often
- Add New Features Incrementally
- **Focus on Simple yet Effective Automated Routine Reports from 'live' Data Streams**
- De-emphasize Complicated Models Requiring Ad-Hoc Analysis (These don't often get used)

- **Additional Applications**

- AFC-based Cordon Counts and Hourly Link-Loads
- AFC-based Bus Ridechecks
- Operating Performance from Signal System Data



14 St., IRT Seventh Av. Subway





42 St.-Port Authority Bus Terminal, IND Eighth Av. Subway

## Acknowledgements

- **Federal Transit Administration**
  - Gary Delorme, FTA
- **National Transit Database**
  - Sergio Maia, NTD-TSP
- **MetroCard Operations**
  - Miguel Garcia, Daniel Rodriguez
- **NYCT Office of Management and Budget**
  - Larry Hirsch, Qifeng Zeng
- **NYCT Operations Planning**
- **New York City Transit Museum**

