MINNESOTA DEPARTMENT OF TRANSPORTATION

RESEARCH

2008-09

Improving Capacity Planning for Demand-Responsive Paratransit Services

Take the steps... Research...Knowledge...Innovative Solutions!

Transportation Research

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.
MN/RC 2008-09		
4. Title and Subtitle Improving Capacity Planning for Demand-Responsive Paratransit Services		5. Report Date April 2008 6.
7. Author(s) Diwakar Gupta, Hao-Wei Chen, Lisa Miller, and Fajarrani Surya		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Mechanical Engineering Graduate Program in Industrial & Systems Engineering University of Minnesota 111 Church ST SE Minneapolis, Minnesota 55455		10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. (c) 81655 (wo) 169
12. Sponsoring Organization Name and Address Minnesota Department of Transportation 395 John Ireland Boulevard Mail Stop 330 St. Paul, Minnesota 55155		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes http://www.lrrb.org/PDF/200809.pdf		

16. Abstract (Limit: 200 words)

This report proposes and evaluates two ideas for improving efficiency and service quality of paratransit operations. For carrying out this analysis, the authors use data from Metro Mobility, the agency responsible for providing ADA-mandated transportation services in the Twin Cities. However, the underlying principles, mathematical models, and algorithms are applicable to a variety of similar transportation operations in urban and rural areas.

The first idea is to re-optimize routes developed by Metro Mobility's route-building software (a commercial product named Trapeze) at the end of each day of booking operations to reduce the total time it takes to serve booked trips. The second idea evaluates the selective use of non-dedicated vehicles and service providers (e.g. taxi services) for lowering operational costs. Mathematical models and computer algorithms are developed for each of these approaches. These are then tested on actual operational data obtained from Metro Mobility.

The report shows that a conservative estimate of savings from re-optimization would be 5% of Metro Mobility's operating costs. Additional savings from the use of taxi service would be in the hundreds of dollars per day. The actual magnitude of these savings would depend on the proportion of customers who agree to travel by taxi.

17. Document Analysis/Descriptors		18. Availability Statement	
Paratransit Metro Mobility	Transportation services ADA transportation Cost	No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 78	22. Price

Improving Capacity Planning for Demand-Responsive Paratransit Services

Final Report

Prepared by:

Diwakar Gupta
Hao-Wei Chen
Lisa Miller
Fajarrani Surya

Department of Mechanical Engineering
Graduate Program in Industrial & Systems Engineering
University of Minnesota

April 2008

Published by:

Minnesota Department of Transportation
Research Services Section
395 John Ireland Boulevard, MS 330
St. Paul, Minnesota 55155-1899

This report represents the results of research conducted by the authors and does not necessarily represent the views or policies of the Minnesota Department of Transportation and/or the Center for Transportation Studies. This report does not contain a standard or specified technique.

The authors and the Minnesota Department of Transportation and/or Center for Transportation Studies do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to this report.

Acknowledgements

The research team is grateful to Ms. Sarah B. Lenz of MnDOT's Office of Transit who served as the Technical Liaison, Dr. Alan Rindels, Program Development Engineer at MnDOT's Office of Investment Management, Research Services, who served as the Administrative Liaison, Mr. Dana E. Rude, Project Administrator at Metro Mobility, Mr. Dave Jacobson, former chief of Metro Mobility and Mr. Matthew Yager, Metro Mobility's system administrator. All these individuals supported this project by unstintingly giving their time, effort, and advice. The team is particularly indebted to Mr. Dana Rude for his firm support of this study starting from its conception to completion, and his invaluable help in obtaining the necessary data and arranging meetings with contractors. Finally, the research team is grateful to the staff at the Center for Transportation Studies, especially Ms. Linda Priesen and Ms. Laurie McGinnis for their encouragement and support of this project through its fruition.

Table of Contents

List of Tables

List of Figures

Executive Summary

State agencies responsible for ADA-eligible paratransit services are increasingly under pressure to contain costs and maximize service quality. Many do not themselves operate vehicles; instead, they contract out the provision of services. The contractors are paid for each hour of service. They are responsible for hiring crew, operating vehicles, forming routes, dispatching, and maintaining the vehicles.

This report describes the outcome of a research project, supported by the Minnesota Department of Transportation and the University of Minnesota's Center for Transportation Studies that used data from Metro Mobility (a Twin Cities based provider of paratransit services) to identify opportunities for improving efficiency and service quality. The research project also developed mathematical models and computer algorithms that can be used to implement the outlined approaches. Although the analysis is based on Metro Mobility data, it is possible to apply the underlying principles to a variety of other situations involving ADA transportation in different parts of the state and the country.

The project develops two key ideas for improving efficiency. The first idea is to reoptimize routes developed by Metro Mobility's route-building software (a commercial product named Trapeze) at the end of each day of booking operations to reduce the total time it takes to serve booked trips. The second idea evaluates the selective use of non-dedicated vehicles and service providers (e.g. taxi services) for lowering operational costs. Mathematical models and computer algorithms are developed for each of these approaches. These are then tested on actual operational data obtained from Metro Mobility.

The report shows that a conservative estimate of savings from reoptimization would be 5% of Metro Mobility's operating costs. Additional savings from the use of taxi service would be in the hundreds of dollars per day. The actual magnitude of these savings would depend on the proportion of customers who agree to travel by taxi.

Chapter 1

Introduction

Planning for paratransit services in most counties and multi-county metropolitan areas of the United States is driven by the requirements of the Americans with Disabilities Act (ADA)–see `http://www.ada.gov/pubs/ada.htm` for a text of the ADA. There are, broadly speaking, six factors to consider when planning delivery of such services. These are:

1. Service must be provided in a corridor that roughly matches the fixed-route transit system, such as bus or rail, and extending three quarters of a mile on either side of the fixed-route corridor.

2. Next-day service should be made available and reservations, up to 14 days in advance, should be permitted.

3. Reservations must be taken all days when service is provided and reservation hours must be comparable to service hours.

4. The fare charged should be no more than twice the fare paid by a person without disability on a fixed-route system.

5. Prioritizing of trips, e.g. by trip purpose, is not permitted.

6. Matching paratransit service must be offered during the same days and hours as the fixed-route system is in operation. Public entities are prohibited from limiting the amount of paratransit services provided to ADA-eligible persons.

Some States and metropolitan areas go beyond the minimum ADA requirements. For example, ADA requires paratransit service to be at a minimum a curb-to-curb service. However, a Minnesota state statute requires "first-door-through-first-door" service. This means, for example, that if a customer's pick up is from an apartment complex and drop off is at a shopping mall, then the driver will pick the customer up from the main entrance of the apartment complex and drop him/her off at the main entrance to the mall. Another example is that ADA requires service in a three quarter mile corridor on either side of the fixed-transit routes, whereas Minnesota statutes define service area as the political boundaries of the municipalities served by the fixed-route system (Metropolitan Council 2001).

In short, both Federal and State legislation affect planning for paratransit services. Special lift-equipped vehicles are used to make this service accessible to ADA-eligible users. Such users are called *certified users* in this report. They constitute the pool of clients from which demand originates. For planners in each county/metropolitan area, the size of this pool is known, although it changes over time.

State agencies responsible for ADA-eligible paratransit services are increasingly under pressure to contain costs and maximize service quality. Many agencies do not themselves operate vehicles; instead, they contract out the provision of services. However, they do own dedicated vehicles that are leased to contractors at a nominal cost. The contractors are responsible for hiring crew, operating vehicles, forming routes, dispatching, and maintaining the vehicles. In the Metropolitan Twin Cities area, they are paid on the basis of the number of "revenue" hours. The number of revenue hours accrued on a given route is the amount of time elapsed (in hours) from the first pick up to the last drop off of continuous operation of the route. Total revenue hours that can be billed are capped in each two-week period.

The need to match hours of operation to fixed-route transit system hours, minimize capacity denials (interpreted by the Federal Transit Authority (FTA) to mean zero denials due to capacity shortages), and maximize service quality measures as well as efficiency requires careful attention to all aspects of capacity management. In addition to capacity denials, different county and metropolitan governing bodies may use a variety of additional service quality measures to determine if contractors are providing adequate service. In the Twin Cities metropolitan area, important service quality metrics include the percent of on-time pick-ups, missed trips, on-time drop-offs, acceptable ride times, and excessive ride times.

The primary focus of the research project, which forms the basis of this report, was to analyze Metro Mobility data and to develop mathematical models and algorithms to improve operational efficiencies. [Metro Mobility is responsible for providing ADA paratransit services in the Twin Cities area. It reports to the Metropolitan Council.] Specifically, this report contains answers to the following questions:

1. Is there a statistically significant pattern of operational performance that could uncover sources of operational efficiency?

2. How do contractors respond to the different incentives and disincentives offered in typical service contracts?

3. Is it possible to improve the efficiency of routes? If so, by how much?

4. Can state agencies reduce costs by using a mix of dedicated and non-dedicated vehicles/services?

Whereas the first two questions above can be answered after analyzing historical data with the help of standard statistical techniques, it is much harder to answer questions 3 and 4 because of a variety of complicating factors that make these problems difficult. We describe the key factors in the next section.

1.1 Complicating Factors

Many ADA paratransit service providers use commercial software to perform trip planning and scheduling. Metro Mobility uses a software package called *Trapeze* — see `http://www.trapezegroup.com/solutions/pt_pt.php` for details. Whereas such tools have helped improve accountability and efficiency, a lot more can be accomplished by using advanced Operations Research techniques. We will describe some of these methods later in this proposal. But first, let us examine some factors that make the route planning problem non-trivial.

1. Dynamic trip scheduling: Clients need to be assigned to vehicles, and pick-up and drop-off times need to be determined, without knowledge of requests from all clients who will need services on any given day. Typically, clients are quoted a pick up time when they call. This requires dynamic trip scheduling and route formation. Dynamic scheduling problem is inherently more challenging and requires the use of either heuristic or rule-based scheduling approaches, which inevitably results in suboptimal use of vehicle capacity.

2. Variable demand: The demand for services is not stationary in time. Demand rate changes by the hour of the day, by the day of the week, and by changing seasons. In addition, demand during a particular hour of a particular weekday is random.

3. Variable capacity: The realized capacity of each dedicated vehicle depends significantly on the decision processes that lead to the formation of vehicle routes and schedules. Trip durations, which in turn determine the number of trips that each vehicle can complete each day, also depend non-trivially on the chosen route, the time of day, and the weather conditions.

Next, let us qualitatively examine some pros and cons of using dedicated and non-dedicated vehicles. Dedicated vehicles provide ease of scheduling, complete control over availability, and by and large, better service. The latter comes from the fact that drivers of dedicated vehicles are specially trained to serve ADA clients. For example, in the Twin Cities area, drivers usually pick up clients inside the outermost door of their pick-up spot, drop off the client inside the first door of their destination, and usually help carry items such as groceries. This is especially important during winter months and when the client is in a wheelchair and does not have a personal care assistant. The major drawback of a strategy of providing service with dedicated-only vehicles is the overall system cost and the lack of agility in responding to changing demand patterns. In order to maintain service quality, a sufficiently large number of vehicles must be placed in service to meet peak demand. This creates excess capacity during low demand hours during which there is no alternate use for the equipment/personnel.

In contrast, non-dedicated vehicles, e.g., vehicles operated by taxi service providers, volunteers, and those used for multiple customer populations, can be used to take some of the peak load off the dedicated system. Such vehicles can complement dedicated capacity while remaining available to serve other types of demand during non-peak hours. Thus, non-dedicated vehicles can be an economical solution to the problems caused by the time varying demand pattern. However, the availability of a sufficiently large number of non-dedicated vehicles is not guaranteed because the demand from other sources may peak at the same time when ADA demand peaks.

In what follows, we have summarized key issues that need to be resolved in order to perform a cost-benefit analysis of the non-dedicated option. This list and the rest of this report assumes that the source of non-dedicated vehicles are taxi service providers. The assumption is consistent with the possible sources of such vehicles in the Twin Cities area at this point in time. Our analysis can be adapted to suit different needs, e.g. when the non-dedicated vehicles are multi-use county vehicles, or volunteer-driven vehicles.

1. Which clients should be offered an opportunity to travel by taxi? This requires finding the customer whose removal from a dedicated route leads to the highest cost savings overall.

2. What should the service provider do if the next customer in the list turns down an offer to travel by taxi?

3. The demand for paratransit services usually peaks during the morning and afternoon/evening rush hours. These are also the times during which demand is high for taxi services. Given this reality, it is important to ask how many taxis (non-dedicated vehicles) are likely to be needed to serve a given population of certified users?

In addition to the financial costs and benefits, there are a host of service quality related factors that may affect a state agency's decision to use taxi service. This report does not address the quality issues, some of which are listed below. However, a service provider that is considering using non-dedicated option should work with the providers of non-dedicated vehicles to resolve such issues. It is reasonable to assume that the vast majority of these issues can be resolved via negotiation.

1. Can the taxi service guarantee availability of a vehicle on return trips when the client is dropped off in an area that is not usually served by the taxi service?

2. How can the taxi service ensure compliance with procedures for driver training, proper maintenance of vehicles, and drug and alcohol testing?

3. In case of an accident, is there a transfer of risk from taxi service provider to the ADA paratransit service provider when clients are transported in taxis? What would it cost the state agency to obtain insurance against this risk?

Finally, some counties/metropolitan areas are considering same day service, which goes beyond the ADA mandated level of service. [Note that at present only next-day service is required by ADA policy.] Same-day service makes it difficult to plan for capacity usage. It is therefore more likely to cause capacity shortages and overages. With very little advance notice, same day service providers need to develop decision processes to decide when to insert a client's pick-up/drop-off within an already scheduled route and when to start a new route. This type of service is likely to benefit even more from the use of non-dedicated transportation options. The approach outlined in this report can also help decision makers choose the right mix of different service/equipment types for same-day service.

Readers should note that although we use Metro Mobility data to shine light on avenues for improving operational efficiency, such analysis is applicable to paratransit operations in metropolitan and rural areas across the country. For example, in the metropolitan areas, non-dedicated services may be offered by taxis, whereas in rural areas, such services may be offered by a team of volunteer drivers.

1.2 Organizational Details

The remainder of this report is organized as follows. Chapter 2 contains statistical analysis of Metro Mobility data and answers questions 1 and 2 described earlier in this Chapter. Chapter 3 contains a reoptimization approach that can lead to more efficient route formation. Chapter 4 describes an algorithm to determine which customers should be offered taxi service and conducts a cost-benefit analysis of the non-dedicated option as a function of the fraction of customers who accept the offer to travel by taxi. A summary of our findings, possible future directions for research, and concluding remarks can be found in Chapter 5.

Chapters 2, 3 and 4 are written in a manner that they can be read independently. That is, each chapter has its own Introduction and Conclusion section. We do not present mathematical models in the main body of the report. These are included in the Appendix. Finally, copies of the computer programs used to perform the reported analysis can be obtained from the Research Services Section of the Minnesota Department of Transportation.

Chapter 2

Understanding Performance

For the purpose of analyzing Metro Mobility's (MM) operational performance, we received data from January 2001 through August 2005 pertaining to cancelations, no shows, and overall productivity. We also obtained data on the booking behavior of customers and contractor productivity for 2004 and 2005. Finally, we obtained financial data on bonuses and damages for all of 2004 and from January–August of 2005. The purpose of the analysis presented in this Chapter was to identify patterns of performance that could lead to operational improvements. We present results pertaining to each performance metric under a different heading in the remainder of this Chapter.

Metro Mobility uses two contractors in the Twin Cities Metropolitan area. These are Laidlaw and Transit Team. Each contractor is responsible for a particular region of the metro area and is provided with MM vehicles that are purchased with the help of federal grants. Contractors are paid a fixed rate for each hour of service provided.

2.1 Analysis of cancelations

We studied the variation in the percent of trips that were canceled by year and by month. The results are shown in Figures 2.1 and 2.2 respectively. Yearly comparisons controlled for variations from one month to another, whereas monthly comparisons controlled for variations across years.

Upon performing a test of equality of mean percent cancelations, we found that the cancelation rate was different in at least one year (p-value=0). In particular, percent mean cancelations for 2005 were significantly smaller (p-value=0.05). There was no significant difference in the month to month cancelations (p-value=0.951), although visually it appears from Figure 2.2 that there was a drop in cancelations during the period from October 2004 through July 2005.

We investigated whether there was a correlation between demand volume and percent cancelations. Our analysis revealed a weak positive correlation — the correlation coefficient was 0.270 with a p-value of 0.044. Therefore, the correlation is statistically significant at a 5% significance level. Although the true reason why there were proportionally fewer cancelations when demand was lower remains unknown, one possible reason is as follows. When demand

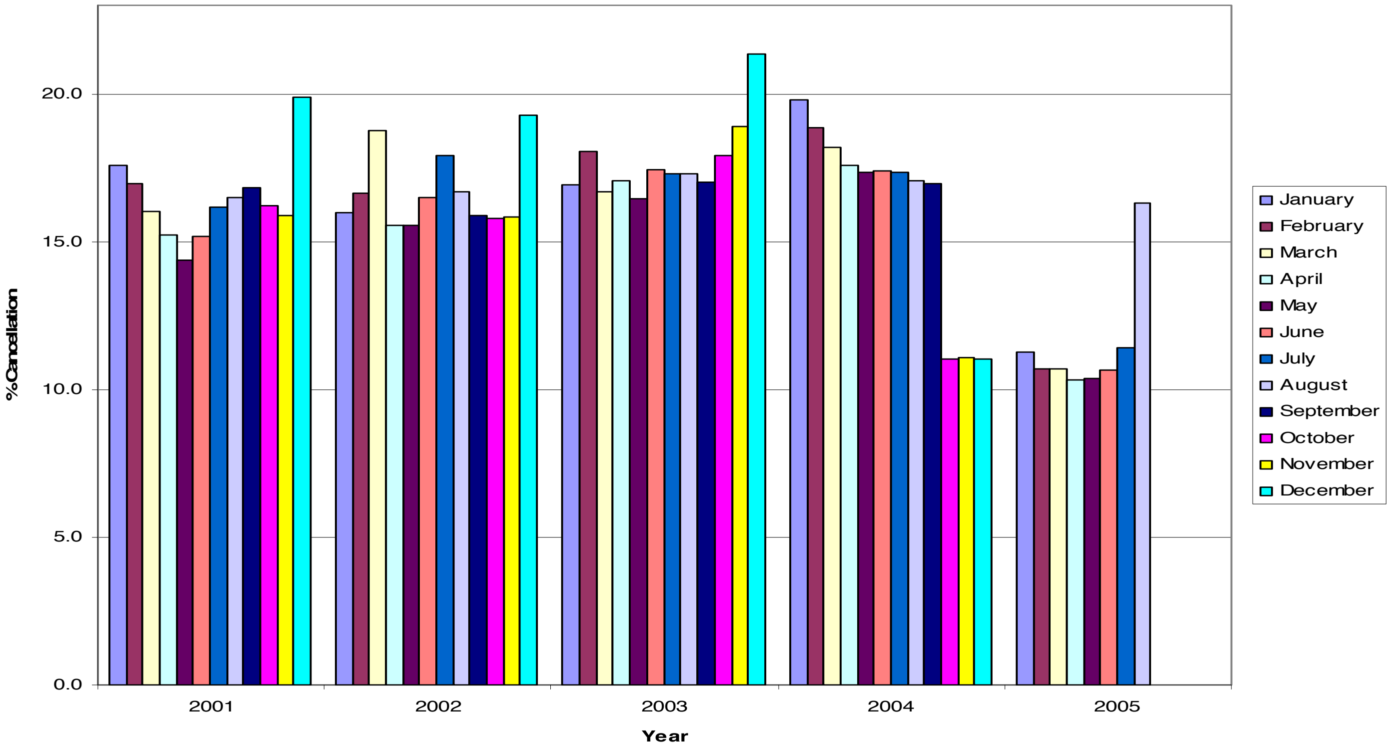


Figure 2.1: Yearly cancelation analysis

is high, customers need to book well in advance and more customers receive less than ideal pick-up and drop-off time schedules. This leads to greater cancelations. From an operational efficiency viewpoint, this is bad news because cancelations increase variability precisely when it is important to manage capacity carefully.

2.2 Analysis of No Shows

Our analysis of percent no-shows followed the same approach that we used for analyzing percent cancelations. Our results are summarized in Figures 2.3 and 2.4. Analysis of variance showed that the percent no-shows were different in at least one year (p-value=0) and in particular, no-show percent was lower in 2002 and 2005 (p-value=0.05). There was no significant difference in month-to-month no-shows (p-value=0.499).

We also found a weak positive correlation between percent no-shows and demand. The correlation coefficient was 0.281 with a p-value of 0.036. As in the previous section, the true reason why percent no-shows were correlated with demand volume remains unknown. However, the same reasons that resulted in greater cancelations are also likely to cause a greater proportion of no shows.

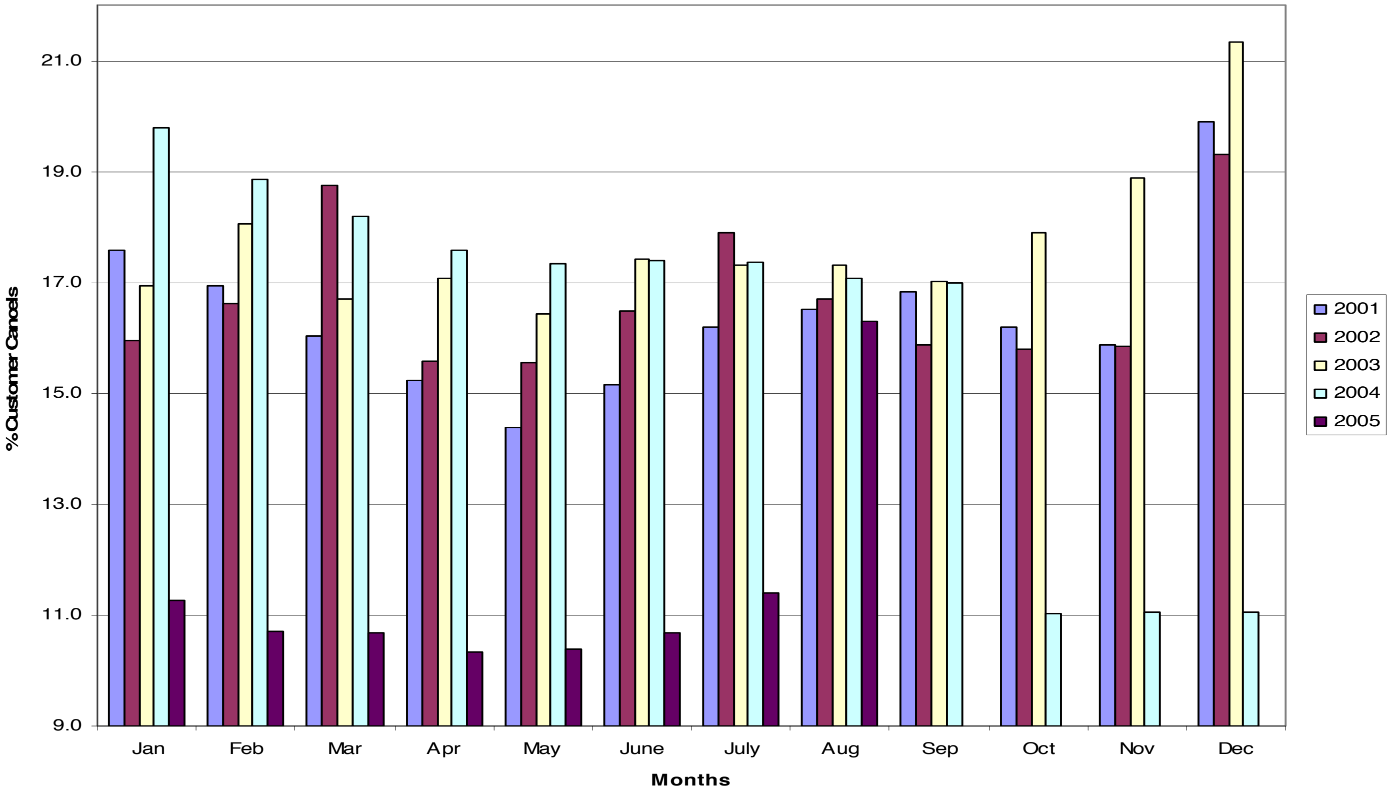


Figure 2.2: Monthly cancelation analysis

2.3 Analysis of Productivity

Metro Mobility measures productivity by the average number of passengers served per revenue hour. We analyzed data for productivity comparisons by year and by month. These results are summarized in Figures 2.5 and 2.6.

Analysis of variance revealed that whereas there was no statistically significant difference in productivity from month to month, productivity was lower in 2005 (p-value=0.05). Because 2005 had lower demand, this observation led us to investigate whether productivity changes could be explained, in part, by demand levels. Plots of the relationship between productivity and demand are shown in Figures 2.7 and 2.8 for Laidlaw and Transit Team data, respectively.

Both contractors' productivity tracked demand. Upon performing a linear regression analysis with demand as the independent variable and productivity as the dependent variable, we discovered a statistically significant relationship, but with a small R square. This led us to conclude that although there was a relationship between productivity and demand, the relationship was likely not linear. We will return to this issue in a later section when we discuss the implication of this relationship on contract parameter selection.

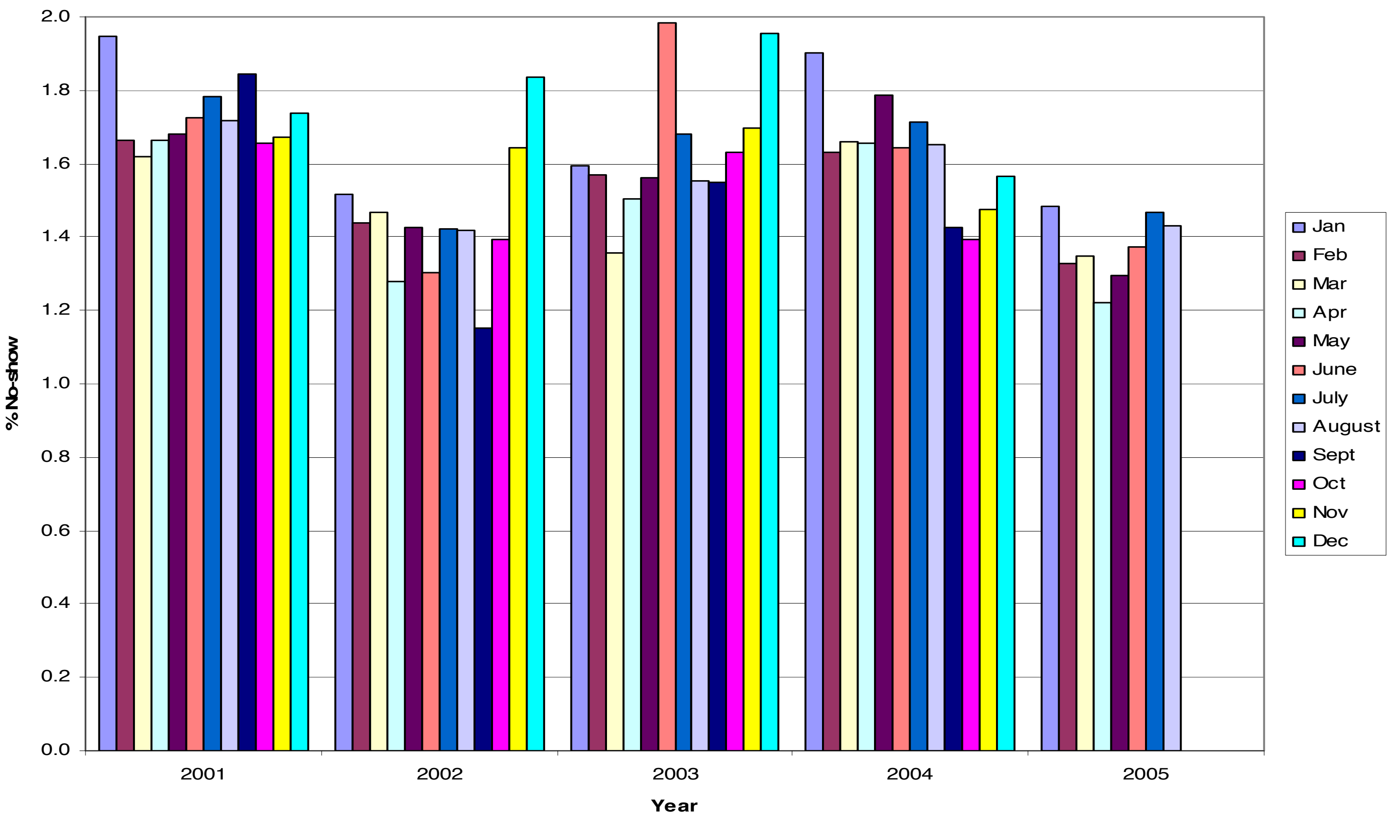


Figure 2.3: Yearly no-shows analysis

2.4 Analysis of Capacity Denials

We studied the variability in capacity denials by the day of the week for each month, and by the month for each weekday. The results of such comparisons are shown graphically in Figures 2.9 and 2.10.

Analysis of variance in which we controlled for weekday variability showed that capacity denials were significantly higher in April and lower in December. When classified by month alone, there was no significant difference in capacity denials by month. However, when controlling for day-of-week variability, capacity denials did vary by month as noted above.

2.5 Contract Design

The fact that contractor productivity was correlated with demand led us to further investigate the appropriateness of incentives used in typical contracts. The main purpose of this analysis was to check if contracts left open the possibility of moral hazard – situations where

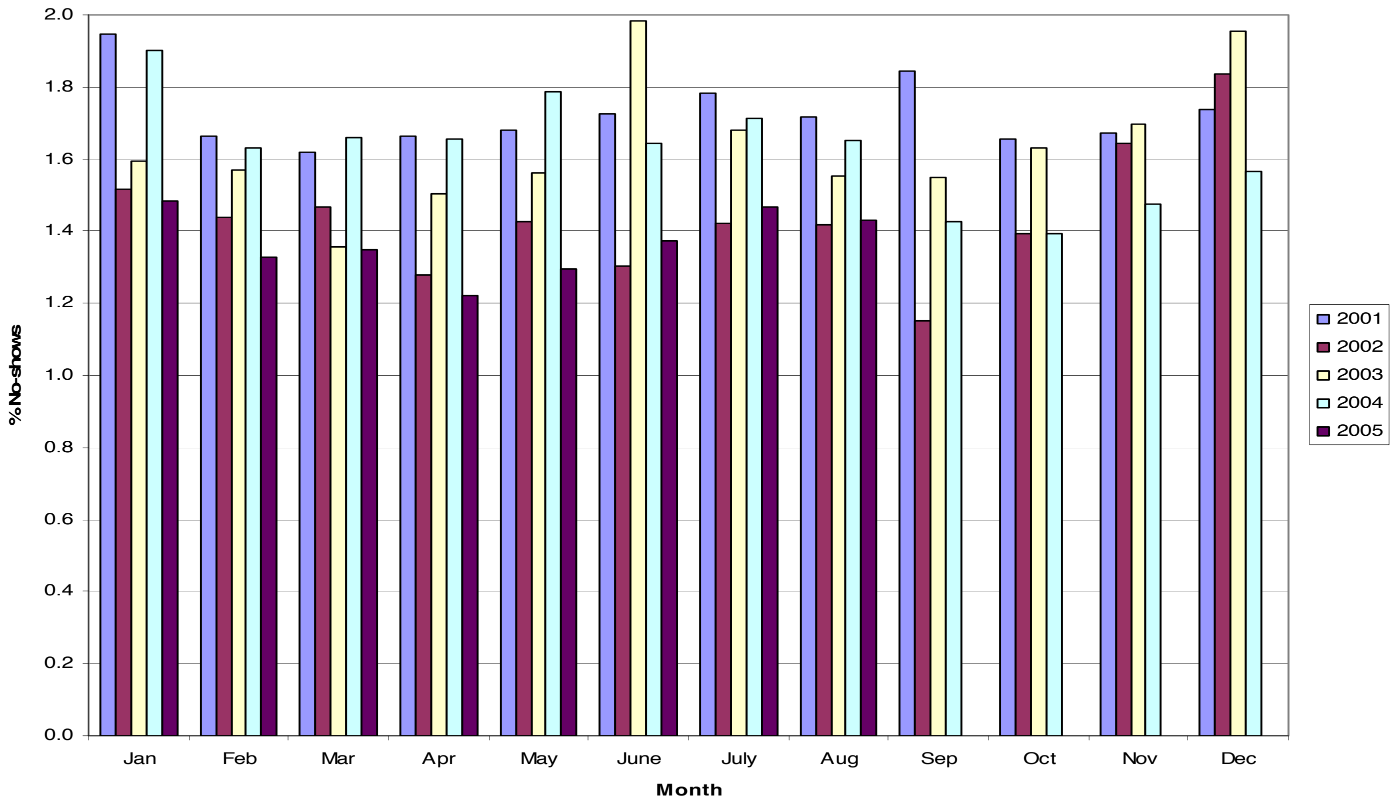


Figure 2.4: Monthly no-shows analysis

the contractor could benefit financially by deliberately choosing non-optimal actions.

For this purpose, we looked at Laidlaw's data from 2004 and first 8 months of 2005. We calculated the payments made to the contractor, including the damages for lower productivity, and compared it with what the contractor would have earned by maintaining a nominal productivity of 1.6 passengers per revenue hour. The latter is the desired productivity level. In each month that a contractor's productivity dropped below 1.6, damages were assessed according to the following rule. If the contractor productivity fell below 1.6 but remained higher than 1.59, a penalty of 0.25% was applied to the total amount billed. This penalty was changed to 0.5% for productivity in the range 1.58 and 1.59, and to 0.75% for productivity below 1.58. Results of our computations are shown in Tables 2.1 and 2.2.

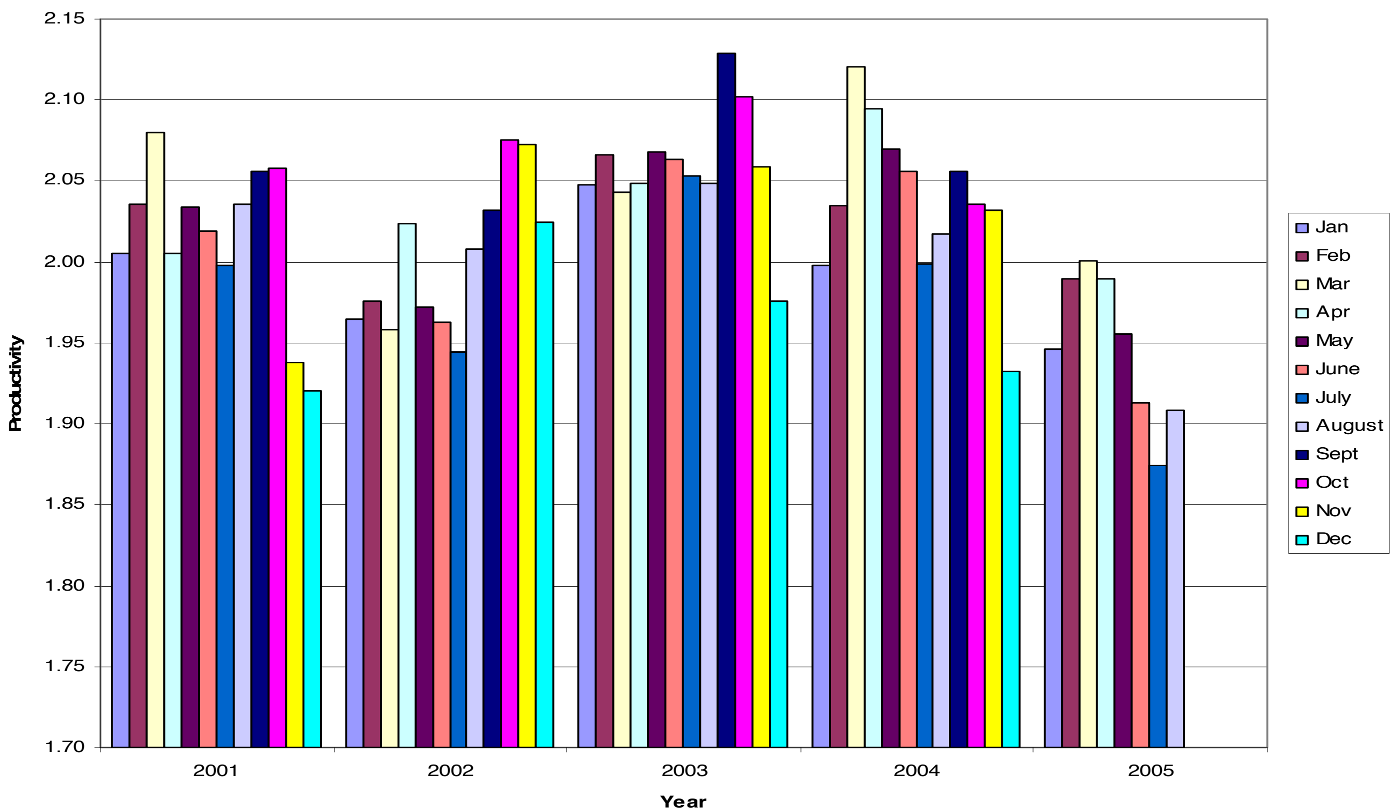


Figure 2.5: Yearly productivity analysis

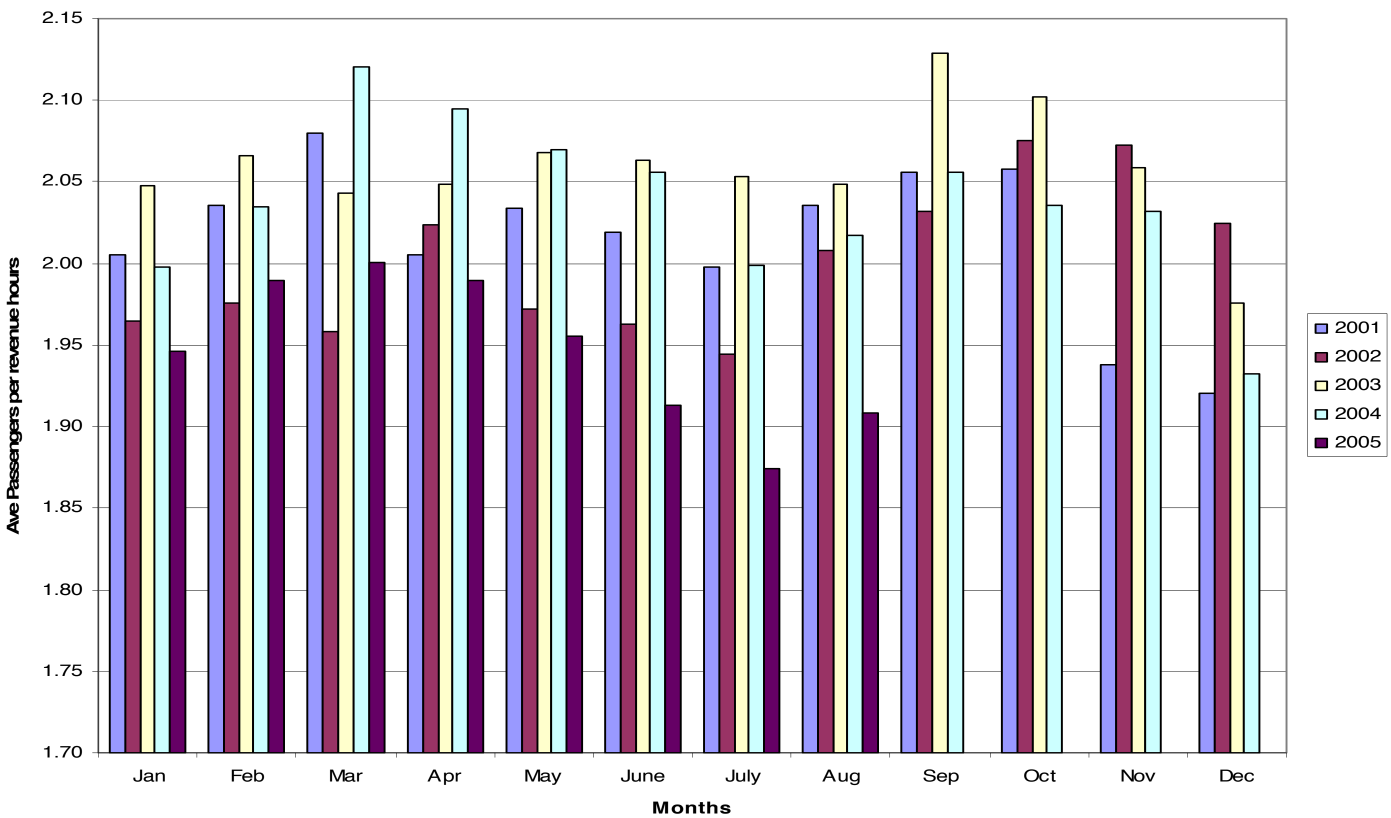


Figure 2.6: Monthly productivity analysis

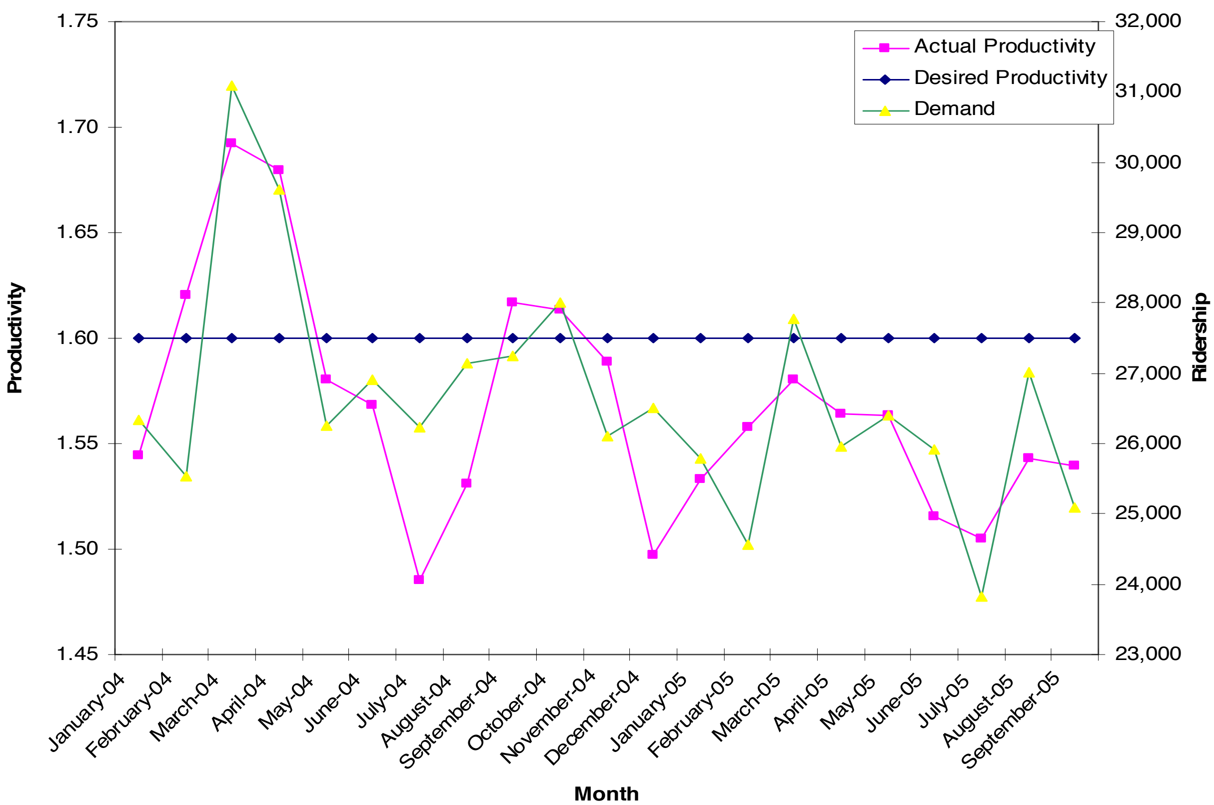


Figure 2.7: Laidlaw's productivity and demand

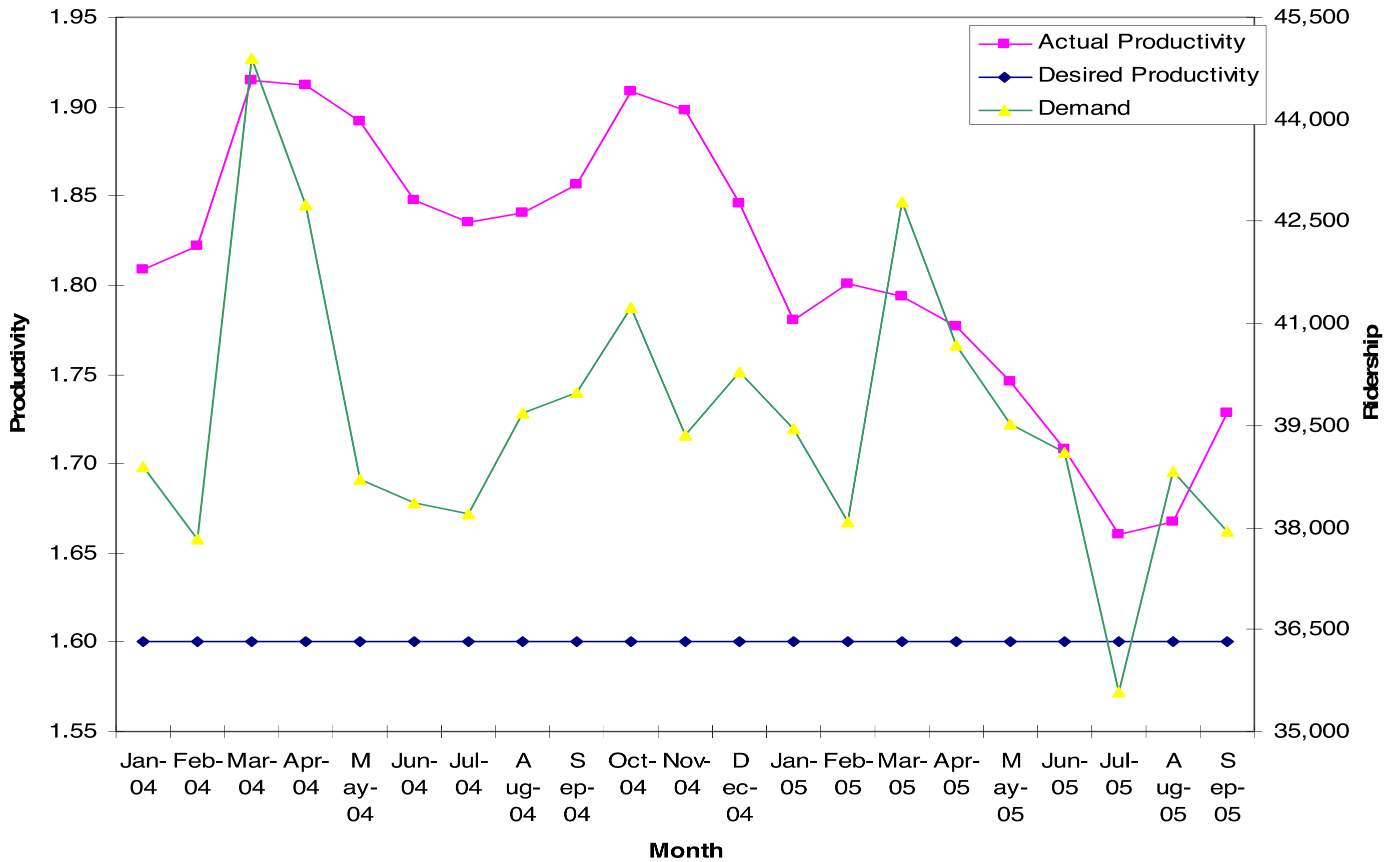


Figure 2.8: Transit Team's productivity and demand

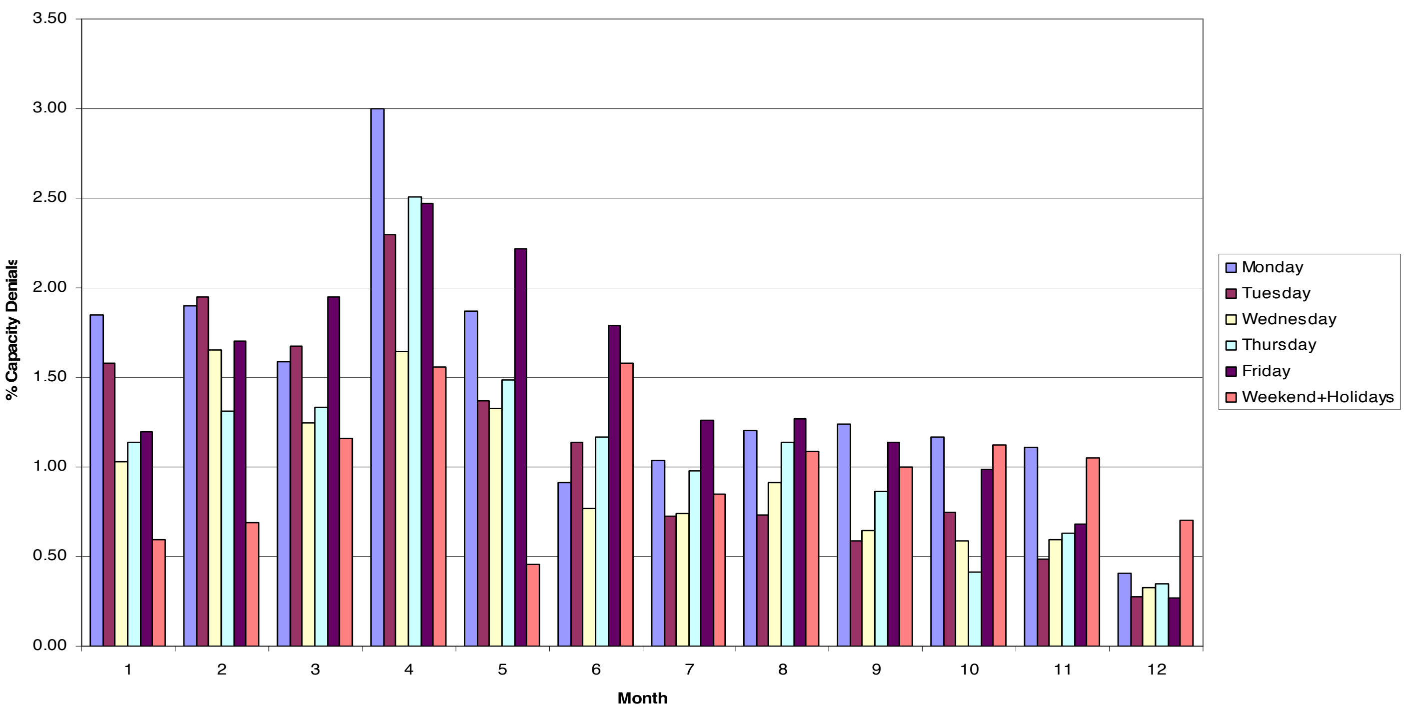


Figure 2.9: Analysis of capacity denials – by month

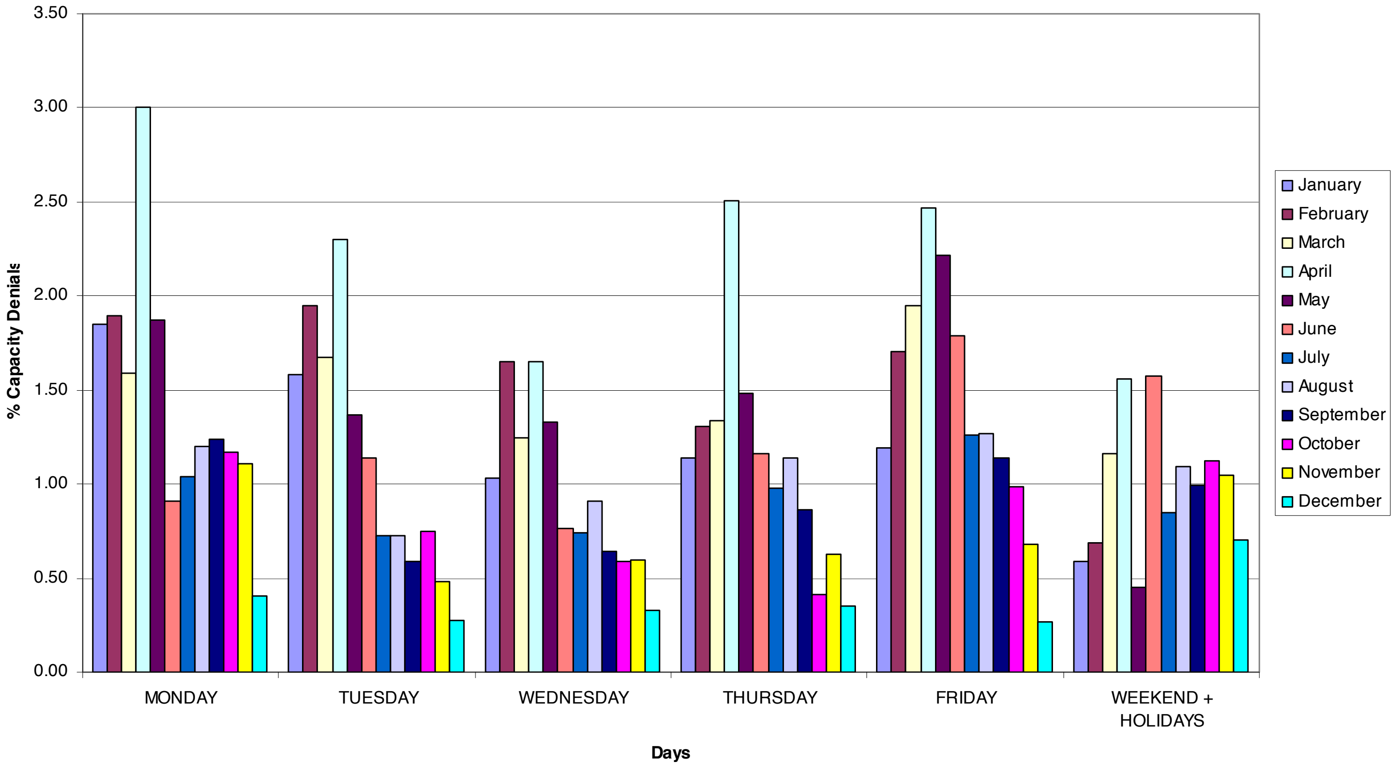


Figure 2.10: Analysis of capacity denials – by day of week

Mon	Demand	Productivity	Rate	Gross Cost	Bonus/ Damages	Total Cost	Allocated Hours	Rev Hrs Used	Nominal Hours	Total Cost (Nominal Hrs)
Jan	26,329	1.54	$38.63	$658,560.38	($4,939.20)	$653,621.18	17,393.54	17,047.90	16,455.63	$635,680.79
Feb	25,532	1.62	$38.63	$608,647.71	$0.00	$608,647.71	16,394.03	15,755.83	15,957.50	$616,438.23
Mar	31,088	1.69	$38.63	$709,642.76	$0.00	$709,642.76	18,326.00	18,370.25	19,430.00	$750,580.90
Apr	29,604	1.68	$38.63	$680,768.38	$0.00	$680,768.38	17,622.79	17,622.79	18,502.50	$714,751.58
May	26,259	1.58	$38.63	$641,997.76	($3,209.99)	$638,787,78	16,896.88	16,619.15	16,411.88	$633,990.73
June	26,916	1.57	$38.63	$663,092.83	($4,973.20)	$658,119.64	17,644.66	17,165.23	16,822.50	$649,853.18
July	26,238	1.48	$40.90	$772,688.69	($5,420.17)	$717,268.52	17,919.56	17,669.65	16,398.75	$670,708.88
Aug	27,143	1.53	$40.90	$724,983.58	($5,437.38)	$719,546.21	17,805.87	17,725.76	16,964.38	$693,842.94
Sept	27,251	1.62	$40.90	$689,451.30	$0.00	$689,451.30	17,199.74	16,857.00	17,031.88	$696,603.69
Oct	28,015	1.61	$40.90	$710,220.32	$0.00	$710,220.32	17,364.80	17,364.80	17,509.38	$716,133.44
Nov	26,111	1.59	$40.90	$672,277.80	($3,361.39)	$668,916.41	16,634.59	16,437.11	16,319.38	$667,462.44
Dec	26,504	1.5	$40.90	$723,939.82	($5,429.55)	$718.510.27	18,315.75	17,700.24	16,565.00	$677,508.50
TOT		Ave=1.58		$8,206,271.33	($32,770.87)	$8,173,500.46	209,518.21	206,335.71	204,368.75	$8,123,555.28
					Cash Fares	$341,216.16				$341,216.16
					Payment	$7,832,284.30				$7,782,339.12
					Difference	$49,945.19				

Table 2.1: Analysis of Laidlaw's Charges for 2004.

Mon	Demand	Productivity	Rate	Gross Cost	Bonus/ Damages	Total Cost	Allocated Hours	Rev Hrs Used	Nominal Hours	Total Cost (Nominal Hrs)
Jan	25,793	1.53	$40.90	$688,095.87	($5,160.72)	$682,935.15	17,216.74	16,823.86	16,120.63	$659,333.56
Feb	24,559	1.56	$40.90	$644,734.51	($4,835.51)	$639,889.00	16,101.76	15,763.68	15,349.38	$627,789.44
Mar	27,770	1.58	$40.90	$718,687.51	($3,593.44)	$715,094.41	18,279.01	17,571.83	17,356.25	$709,870.63
Apr	25,954	1.56	$40.90	$678,644.74	($5,089.99)	$673,574.76	17,067.33	16,593.27	16,221.25	$663,449.13
May	26,398	1.56	$40.90	$690.620.22	($5,179.65)	$685,440.57	17,265.64	16,885.58	16,498.75	$674,798.88
June	25,908	1.52	$40.90	$699,359.73	($5,245.60)	$694,114.54	17,552.37	17,099.26	16,192.50	$662,273.25
July	23,819	1.5	$38.62	$611,304.01	($4,584.78)	$606,719.23	17,079.23	15,828.69	14,886.88	$574,931.11
Aug	27,016	1.54	$38.62	$676,359.78	($5,072.20)	$671,287.09	18,501.94	17,513.20	16,885.00	$652,098.70
TOT		Ave=1.55		$5,407,826.72	($38,761.98)	$5,369,064.74	139,064.02	134,079.37	129,510.63	$5,224,544.69
					Cash Fares	$228,070.91				$228,070.91
					Difference	$144,520.06				

Table 2.2: Analysis of Laidlaw's Charges for 2005.

The column "Nominal Hours" shows the number of hours for which the contractor would have received payment if it maintained a productivity of 1.6. The differences in actual total cost (amount paid) and total cost using nominal hours are the potential savings. The analysis shows that the contractor received a greater payment, even after paying damages, than what it would have received if it had maintained the desired productivity level of 1.6. Specifically, in 2004, Laidlaw was paid about $50,000 more, and between January–August 2005 it was paid about $145,000 more than what it would have received with a productivity of 1.6.

Productivity can be an issue in periods of low demand. During these times, the contractor benefits by charging MM for a greater number of revenue hours at the expense of productivity. Our analysis reveals a potential for improving contract parameter selection, e.g. one remedy might be to calculate revenue hours assuming a productivity of 1.6 or the actual productivity, whichever is higher. In this arrangement, the contractor should also receive a share of the efficiency-related savings when its productivity exceeds 1.6.

2.6 Conclusions

In addition to the performance metrics reported above, the investigators also examined a variety of service quality measures used by Metro Mobility. There were no specific patterns identified. However, for sake of completeness, a copy of the MS Power-Point presentation summarizing the results of this exercise is included in Appendix A. The key finding of this Chapter are as follows.

1. Cancelations and no shows are affected by demand volume and tend to be proportionally higher when demand volumes are higher. Thus, cancelations and no shows increase variability in demand, leading to an inefficient use of capacity at a time when demand is high and greater efficiency would be needed.

   Contractors could adopt a variety of protocols during periods of high demand to mitigate the negative consequences of a higher proportion of cancelations and no shows, e.g. they may maintain a waiting list of passengers whose requests were not be met when they first called. When coupled with the reoptimization routine described in Chapter 3, a waiting list would provide an opportunity to fill freed up capacity from cancelations.

2. The contract terms offered to the contractors do not eliminate moral hazard. In particular, during low demand periods, the contractors can benefit financially from having low productivity even after paying the damages.

   This project did not study contract parameter optimization problem, but that would be a fruitful area of future investigation.

Chapter 3

Route Reoptimization

In this chapter, we answer the question "Is it possible to improve the efficiency of routes without changing the procedures used to schedule trips and form routes?" Our approach must also honor all ADA and state requirements for delivering service.

A high-level summary of our methodology is as follows. Starting with the routes generated by Trapeze, we developed and tested a repeated reoptimization approach that improves upon these routes. Trapeze is the software tool used by Metro Mobility (MM) to form routes based on requests received from the ADA population in a dynamic fashion. Our approach ensures that all planned pick-up and drop-off times fall within the ADA and MM norms. That is, each customer is still picked up and dropped off within the allowable time window from the pick-up and drop-off times originally scheduled by Trapeze. However, we use a rigorous mathematical approach to resequence stops along each route, and to move/swap passengers from one route to another. This leads to more efficient routes and reduces the total number of hours needed to serve all customers.

Numerical tests show that the reoptimization procedure can result in savings of between 5-10% in planned operating costs. Putting this in the context of MM's annual operating cost of \$19.5 million in 2004, this would mean potential annual saving between \$0.975 million to \$1.95 million at the 2004 level of expenditure. As fuel and labor costs continue to rise, these savings are likely to be even greater in the future. We recommend that the proposed reoptimization approach be used at the end of each day of bookings. This will open up more room in the partially scheduled routes each day to accommodate certain requests on subsequent days that would have been denied otherwise.

The remainder of this chapter is organized as follows. We begin by describing the data we received from MM in Section 3.1. Section 3.2 gives the sequence of steps that constitute our reoptimization routine for improving operational efficiency. This routine relies on two mathematical models. Technical details of the mathematical models can be found in Appendix B. We discuss the relevant literature dealing with similar problems in Appendix C. In Section 3.3, we present numerical results that provide evidence of the savings that can be achieved using the reoptimization routine. Finally, in Section 3.4, we describe what is necessary to integrate this approach into Metro Mobility's current operations, and provide a summary of our findings and recommendations.

3.1 Data Description

We obtained historical trip data for one day of operations for each service provider from Metro Mobility. The data come in two Microsoft (MS) Access databases. The first database contained the trip data for each passenger served on that day. The categories of data we received are shown in Table 3.1. The second database contained the distances between every location to every other location visited on that day. To protect passenger privacy, we were not given the distance between the depot to any other location. We also did not know the exact location of any pick-up or drop-off stop.

Table 3.1: Data Fields

Fields	Input
SCHEDATE	The date that the trip took place
RUNNUM	Route numbers
STATUS	Status of each request: performed, no-show, cancel
SUBTYPEABBR	Request type: DEM and REG. DEM is demand and REG is subscription
O	Pick up (PU) or drop-off (DO)
BOOKINGID	A unique number that identifies the trip
ADDRESSID	A unique number that identified the location
DISTANCE	Distance between PU and DO
TRAVELTIME	Time traveled between PU and DO
CREATETIME	Time when the request was created
CREATEDATE	Date when the request was created
REQUESTTIME	Time requested by passenger for a pick up
SCHETIME	Pick up time communicated to the passenger by service provider
APPTIME	Time specified by passenger for a drop off
ACTUALARRIVETIME	Time that the provider claims its vehicle arrived at the location
SPACETYPE	Mobility need of the passenger: AM = Ambulatory, WH = Wheelchair, LF = Lift, SC = Scooter
MAPPAGE	Map page of where the stop is located (page # and grid coordinate in Kings' 2005 street atlas)
CITY	The city where the stop is located
PROVIDERNAME	The name of the provider that completed the trip

The data from Laidlaw pertained to its operations on June 1, 2005. It contained 103 routes serving a total of 1009 passengers. The number of passengers in each route varied from 3 to 20, with 60% of the routes having at most 10 passengers. The shortest route lasted 1 hour and 47 minutes with 3 passengers, the longest route lasted 12 hours and 4 minutes with 19 passengers, and 65% of the routes were completed in less than 8 hours. About 60% of the routes started between 4:00 and 9:00 in the morning, 15% started between 9:01 a.m. and 1:00 p.m. in the afternoon, 20% started between 1:01 and 4:00 in the afternoon, and the rest between 4:01 and 6:30 in the afternoon. The break down for the end times were as follows: 15% between 8:00 a.m. and 12:00 p.m., 35% between 12:01 and 4:00 in the afternoon, 35% between 4:01 and 8:00 in the evening, and the rest after 8:00 p.m.

The Transit Team data pertained to its operations on May 9, 2007. It contained 110 routes serving a total of 1534 passengers. The number of passengers in each route varied from 2 to 26, with 65% of the routes having at most 15 passengers. The shortest route lasted 51 minutes with 2 passengers, the longest route lasted 11 hours and 42 minutes with 15 passengers, and 65% of the routes were completed in less than 11 hours. About 80% of the routes started between 4:00 and 9:00 in the morning, 5% started between 9:01 a.m. and 1:00 p.m., and 15% started between 1:01 and 5:00 in the afternoon. The break down for the end times were as follows: 10% between 8:00 a.m. and 12:00 p.m., 65% between 12:01 and 4:00 in the afternoon, and 25% between 4:01 PM and 11:59 PM.

We were also given several contractor-independent parameters that are determined either by Metro Mobility and or by the ADA requirements. The load and unload times for each passenger type is described in Table 3.2. Passengers may not be in the vehicle for longer than $s_1 = 90$ minutes. Each passenger must be picked up within $s_2 = 30$ minutes of their scheduled pick up time and dropped of no earlier than $s_3 = 45$ minutes before their scheduled appointment time, if one is specified. A feasible route must not drop off a passenger after his/her appointment time. The capacity of each vehicle is $c_v = 7$ passengers. In the computer implementation of our reoptimizaton algorithm, parameters s_1, s_2, s_3 and c_v are part of the input data file so that in different applications of this approach, these parameters can be set to their appropriate values without having to change the computer program. Only the data file needs to be changed.

Table 3.2: Passenger type factors

Passenger Type	Load Time (minutes)	Unload Time (minutes)
AM (Ambulatory)	2	1
WH (Wheelchair)	5	4
LF (Lift)	2	1
SC (Scooter)	5	4

We converted the pairwise distances given to us by Metro Mobility into travel times. This step was required only to protect passenger's private information. Metro Mobility computed vehicle speed by multiplying the nominal speed of 25 mph by two adjustment factors, one for distance and one for time of day. The time factor accounted for slower speeds during rush hours. The time-of-day conversion factors we used are shown in Table 3.3. The distance factor accounted for increased speed as longer distances are traveled. The distance conversion factors we used are shown in Table 3.4. Because our model required travel times to be estimated before the arrival times at nodes are known, we used the time window during which each location could be visited to determine the time factor for travel to and from that location. If the time window overlapped two time segments shown in Table 3.3, then the lower factor was used. For instance, if a location could be visited anytime between 6:15-6:45, then a factor of 80% was used. We also added load and unload times to the travel times; see Table 3.2.

Table 3.3: Time factors

Time	Factor	MPH
0:00 – < 6:30	100%	25
6:30 – < 9:30	80%	20
9:30 – < 15:00	100%	25
15:00 – < 18:30	80%	20
18:30 – < 28:00	100%	25

Table 3.4: Distance factors

Distance	Factor
1 – < 4	90%
4 – < 6	95%
6 – < 8	100%
8 – < 10	110%
10 – < 12	120%
12 – < 14	135%
14 – < 16	150%
16 – < 30	160%
30 – < 100	175%

3.2 Reoptimization Routine

MM uses Trapeze to generate routes dynamically as service requests arrive. We worked on the problem of reoptimizing routes generated by Trapeze. We also used the Trapeze solution as a benchmark to compute cost savings that can come from using our approach. A graphical sketch of our approach can be found in Figure 3.1. This approach employs four computer programs: MS Access, MS Excel, ILOG-OPL Studio (an optimization modeling and solution software), and Matlab. The first five steps in our approach are called First-Pass Analysis, whereas steps six to eight are called Second-Pass Analysis. The difference is that in the first five steps, we focus on one route at a time, whereas in the remaining steps we optimize pairs of routes by moving/swapping customers between routes. The steps in our routine are as follows.

Step 1: Export data from MS Access to Excel.

Step 2: Generate OPL Studio data files using Matlab.
We wrote Matlab code (available on the attached disk) to read data from the Excel file and write data files in the format required by OPL Studio. The conversion of travel distances to travel times is also performed by this Matlab code. One data file was generated for each individual route.

Step 3: Check the feasibility of the Trapeze solutions.
Routes with infeasible Trapeze solutions were removed from our test set. From Laidlaw data we found that 18 out of the 103 Trapeze routes were infeasible, i.e. at least one MM or ADA-mandated constraint was violated by these routes. Our method could find a feasible solution in 7 of these instances. However, because our formulation of the routes met all constraints, they were typically more expensive (took more time to complete) than the solution recorded by Trapeze. We did not include infeasible routes in our analysis.

Within Transit Team data, 17 out of 110 routes had missing data, 16 were infeasible (both Trapeze and re-optimization), and 11 routes are too large to solve by OPL (out of memory). This left 66 routes in our analysis.

Step 4: Solve the Stage-1 model for each individual route.
For each individual route, we solved the Stage-1 model (see Appendix B.1 for details) to find the node sequence that minimizes travel time. The Stage-1 model was solved using OPL Studio and the data files generated in Step 2. OPL Studio output was also stored in a solution file for each route. To limit computational burden, a maximum run time of 3 hours was enforced for solving the Stage-1 model for each route.

Step 5: Solve the Stage-2 model for each individual route.
A second Matlab code was written to read the solution files from Step 4 and solve the Stage-2 model (see Appendix B.2 for details). The final solution was output from Matlab in a table format.

If a route reached the 3-hour time limit in Stage-1 without finding a feasible solution, or if after 3 hours, the Trapeze solution had shorter total time than the best OPL solution, then the Trapeze solution was kept as the best solution. Otherwise, the solution from the Stage-2 model was kept as the best solution for that route.

Step 6: Choose pairs of routes to reoptimize.
For additional gains, we allowed passengers to be reshuffled from their originally scheduled routes. We restricted passenger moves to pairs of routes. We tested several criteria for choosing pairs of routes. These criteria included the amount of idle time within

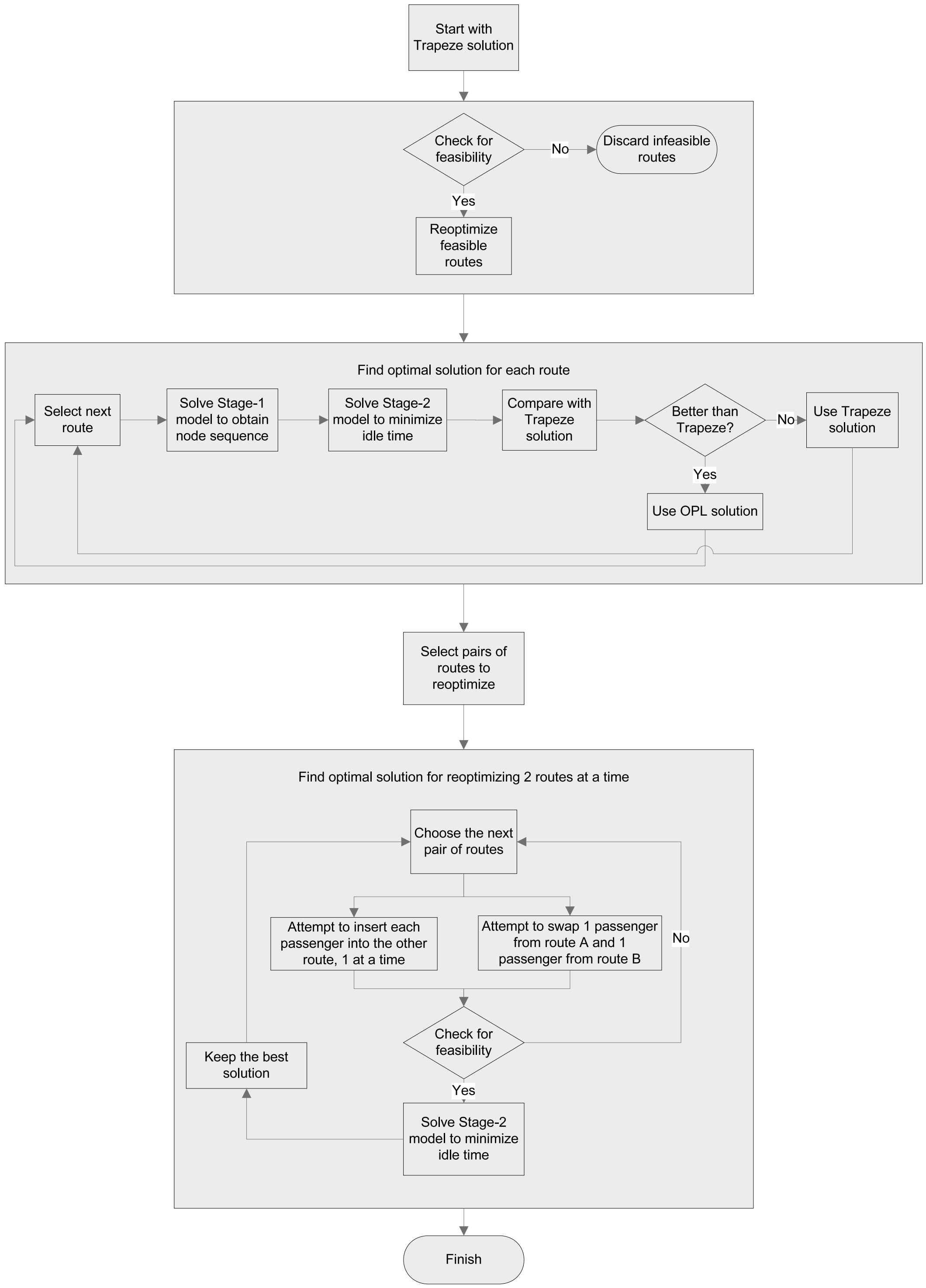


Figure 3.1: Summary of optimization procedure

each route, the start and end times of each route, and the map page numbers visited by each route. There are many more criteria that could be tried. Our goal here was to provide a proof of concept.

For each pair of routes we considered, we used the Matlab code from Step 2 to generate new data files.

Step 7: Solve the Stage-1 models for pairs of routes.
For each pair of routes A and B chosen in Step 6, we started with the best solutions found in Step 5. Note, this is not necessary for our approach to work. One may start with any feasible solution. We solved the Stage-1 model several times after performing several move/swap operations. First, we attempted to swap each pair of passengers from route A and route B and solved the Stage-1 model for the resulting passenger sets. Then we tried moving each passenger from route A to route B, and solved the Stage-1 model for route A with that passenger removed, and route B with that passenger inserted. The same is done for moving passengers from route B to route A. Every time a swap or insertion/deletion leads to a feasible solution for both routes A and B, OPL records the solution into the solution file.

Step 8: Solve the Stage-2 models for pairs of routes.
The Matlab code, described in Step 5, was run on each feasible solution output from Step 7. The solution that had the shortest total time for the sum of routes A and B was kept as the best solution for that pair.

3.3 Numerical Results

In this section we present the numerical results we obtained using the reoptimization routine presented in Section 3.2. This section is divided into two parts. In the first part, we calculate the reduction in total route times that can be achieved by reoptimizing individual routes generated by Trapeze. In the second part, we calculate the amount of additional reductions can be achieved by allowing passengers to be moved to different routes.

3.3.1 First-Pass: Reoptimizing Individual Routes

As described in the previous section, we can apply the Stage-1 and Stage-2 models to find the sequence of locations and corresponding arrival times to reduce travel and idle times for each route (Steps 4-5 of the reoptimization routine). We applied this routine to a test set of 85 Laidlaw routes and 66 Transit Team routes. A summary of the results is given in Tables 3.5 – 3.10. The routes in Tables 3.5–3.7 correspond to Laidlaw data and Tables 3.8–3.10 correspond to Transit Team data. Specifically, Table 3.5 contains of 2 to 12 passenger routes (5 to 25 nodes), Table 3.6 has 13 to 16 passenger routes (27 to 33 nodes), and Table 3.7 has 17 to 21 passenger routes (35 to 43 nodes) for Transit Team. Similarly, Table 3.8 contains of 3 to 7 passenger routes (7 to 15 nodes), Table 3.9 has 8 to 10 passenger routes (17 to 21 nodes), and Table 3.10 has 11 to 16 passenger routes (23 to 33 nodes) for Laidlaw.

The tables include the following columns of data: `Number of Passengers` for each route, the `Travel Time` required to visit the nodes in the original Trapeze solution and the reoptimized solution, the `Total Time` for both solutions (travel time plus idle time), the `% Total Time Improvement` of the reoptimized solution versus the original Trapeze solution, and the `Computation Time` required to obtain the reoptimized solution. The percent total time improvement is calculated as follows:

$$\% \text{ Total Time Improvement} = \frac{\text{Trapeze Total Time} - \text{OPL Total Time}}{\text{Trapeze Total Time}} \times 100\%$$

Next, we describe how the reoptimization procedure changed an example route from Laidlaw and an example route from Transit Team. Table 3.11 shows the node sequence and arrival times for Laidlaw's Route 554 for both the original Trapeze solution and the reoptimized solution. The reoptimized solution has travel time 23 minutes less than the original solution, and the total time is reduced by 44 minutes. The reduction in total time is due to the reoptimized route beginning 30 minutes later and finishing 14 minutes earlier than the original route, picking up the first passenger 30 minutes after their originally scheduled pickup time and dropping off the last customer 14 minutes earlier. The travel time reduction comes from resequencing six nodes in the middle of the route, shown in bold, while still meeting all time window requirements.

Similarly, Table 3.12 shows the node sequence and arrival times for Transit Team's Route 702 for both the original Trapeze solution and the reoptimized solution. The reoptimized solution reduces both travel time and total time by 19 minutes. The reduction in time is due to the reoptimized route beginning 23 minutes later and finishing 4 minutes later than the original route, picking up the first passenger 23 minutes after their originally scheduled pickup time and dropping off the last customer 4 minutes later. Resequencing eight nodes in the middle of the route, shown in bold, also contributes the time reduction.

The percent reduction in total revenue hours is slightly smaller for Transit Team as compared to Laidlaw. This can be explained by observing differences in the initial routes generated by Trapeze for the two service providers – see Figure 3.2. The route sizes are considerably larger for Transit Team. Since the gains from reoptimization come from the ability to move the last drop off in a route, the number of opportunities are reduced when there are fewer routes (and therefore fewer last drop offs that could be moved). Moreover, with longer routes, one needs to find proportionally greater total savings in time to realize a similar percent savings. Data suggest that this is not the case, i.e., larger routes do not offer proportionally greater savings from reoptimizing pick ups and drop offs. This is illustrated in Figure 3.3 in which we show the percent improvement (reduction) in the time to complete the route as a function of route size for the two service providers.

We also analyzed, for several ranges of percent improvement, what portion of the routes achieved an improvement in that range. Figure 3.4 illustrates this analysis. The highest relative frequency is in the range of 4%-6% improvement.

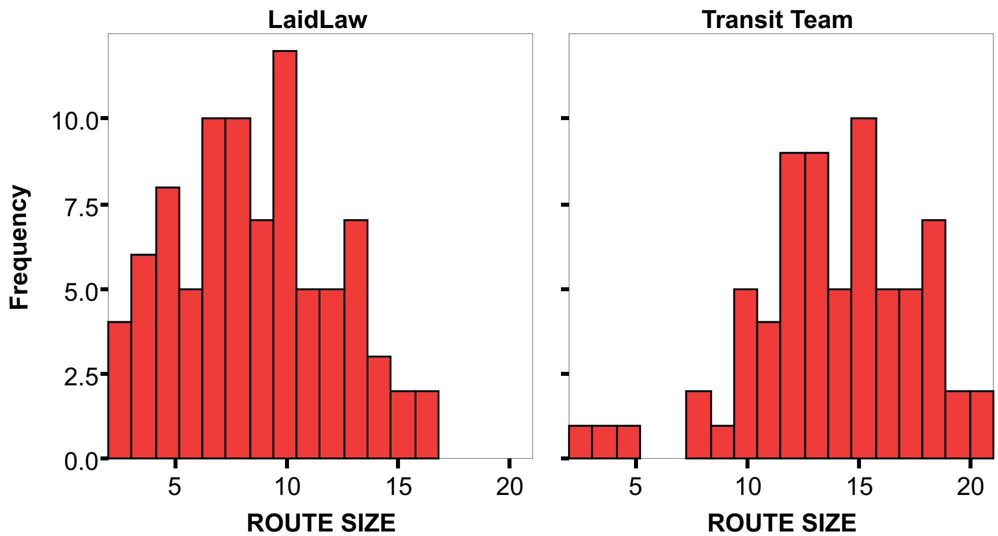


Figure 3.2: Sizes of Trapeze Routes for Laidlaw and Transit Team

3.3.2 Second-Pass: Reoptimizing Pairs of Routes

We next present numerical results from reoptimizing two routes at a time, as described in Steps 6-8 of the reoptimization routine. In this preliminary study, we tested 21 pairs of Laidlaw routes. These pairs were chosen as representatives of several classes of selection criteria. The criteria and the route pairs are shown in Table 3.14.

The numerical results of these tests are shown in Table 3.15. All travel times and total times are summed over both routes in each pair. The percent improvement shown is the improvement over the reoptimized solution in the previous section. The highest improvement was 14.69% with 16 passengers and the lowest was 0%, which indicates that the nodes sequence found by OPL do not improve the total time. The average improvement from re-solving all 21 pairs of routes was 5.64%. The improvements obtained do not appear to be strongly correlated to the selection criteria used to choose each pair. We observed that the greatest improvements came from inserting the last passenger of one route into the other route in the pair, which allowed the first route to finish earlier, often without extending the time of the second route. Tests were not performed to check if moving more than one passenger from one route to another would improve efficiency even more, but this could be done in future studies.

The results of a limited number of experiments with Second-Pass algorithm over Transit Team data are shown in the Table 3.16. We could not perform a more detailed analysis since Transit Team has many large routes (with 15 or more passengers per route) and their reoptimization in pairs takes too much computer time. In fact, we were not able to solve problems involving two routes at a time for a number of pairs that we tried with Transit Team data. Consequently, we selected routes that were relatively small and could not fully evaluate the performance of the Second-Pass algorithm. However, the limited results do show that the Second Pass can further reduce total revenue hours.

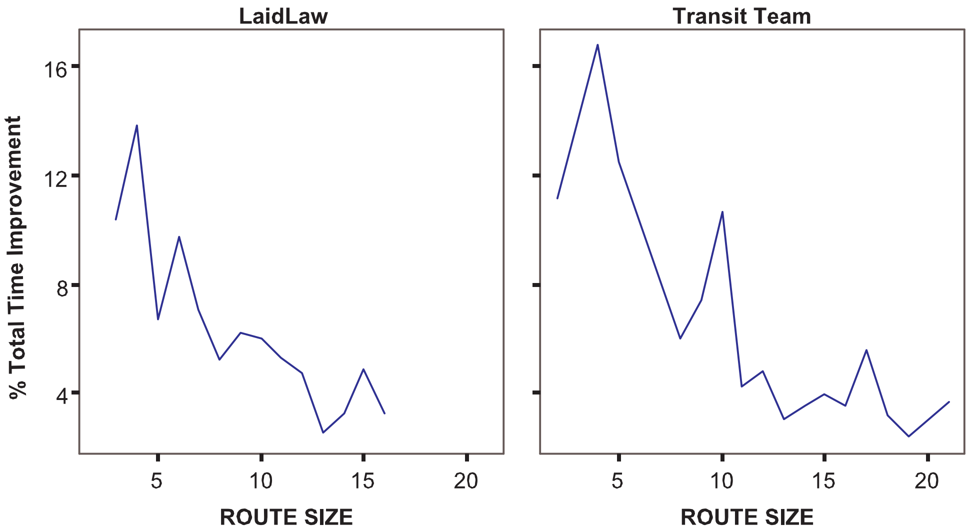


Figure 3.3: Percent Improvement as a Function of Route Size

Figure 3.5 presents the relative frequency of the percent improvement of the pairs reoptimized with our routine. Figure 3.6 shows the total percent improvement when reoptimizing the two routes of each pair individually, represented by the bottom bar, and the additional improvement obtained by reoptimizing the pair together, represented by the top bar. The horizontal axis category labels are the route-pair labels in Table 3.15.

3.4 Conclusions

The Transit Team's data contained 110 routes (scheduled on May 9, 2007) and Laidlaw's data contained 103 routes (scheduled on June 1,2005). Of these, we used 85 Laidlaw and 66 Transit-Team routes for the comparison because the other routes either had missing data or the Trapeze solution was not feasible with respect to one or more constraints. Upon applying the First-Pass algorithm to these routes we realized a total reduction in revenue hours of 30 hours (6.5%) for Laidlaw and 25.4 hours (5.0%) for Transit Team. This translates to a saving of about 21 minutes per route for Laidlaw and 23 minutes per route for Transit Team. The results, reported in Table 3.13 show that First-Pass reoptimization can also increase productivity (the number of customers served per hour). For Laidlaw, the productivity increases from 1.39 customers per hour to 1.47 customers per hour. For Transit Team, it increases from 1.6 customers per hour to 1.68 customers per hour.

Second pass analysis, i.e. reoptimizing two routes at a time can improve efficiency by an additional 5%. However, this step is computationally demanding and requires experimentation to find the best pairs of routes to analyze. The size of problem is too large to evaluate

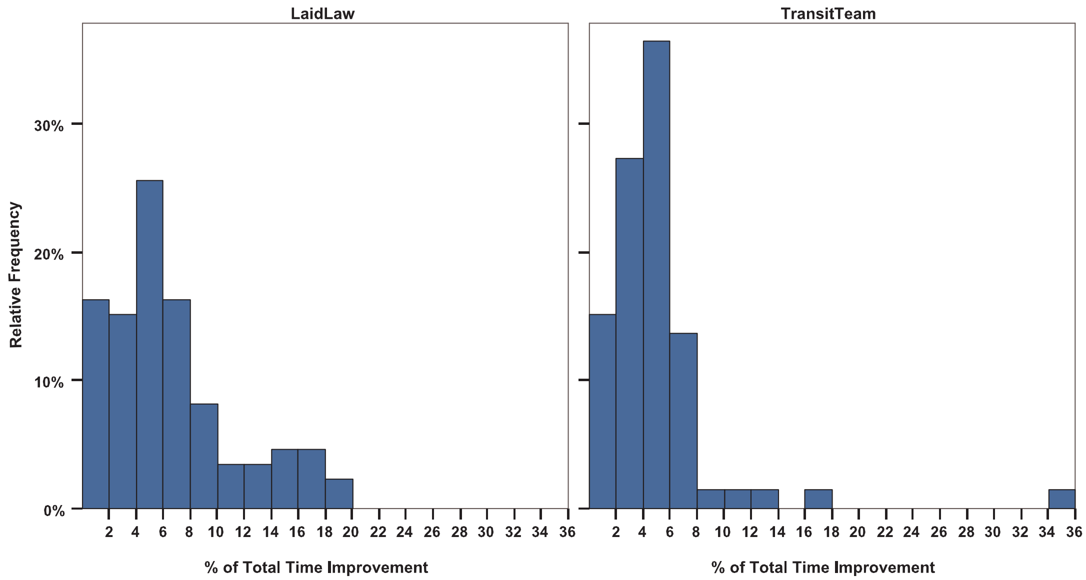


Figure 3.4: Relative frequency of percent total time improvement

all pairs of routes. In contrast, the first-pass analysis can be carried out overnight, that is, during a period when the trip reservation system is shut down.

Key findings of this Chapter are as follows.

1. It is possible to provide the same level of service at a lower cost. The estimated savings are of the order of 5-10% of annual operating costs, or \$975,000-\$1,950,000 annually at the 2004 level of expenditures. Of this, it is realistic to expect that MM can achieve a 5% level of savings by employing only the single-route reoptimization procedure.

2. When our reoptimization routine is applied repeatedly at the end of each day, it will lead to more efficient trip schedules. This will allow MM to schedule more customer requests, some of which may have been denied with Trapeze's original schedule. Thus, reoptimization can also improve service level.

Our computer codes (available on a computer disk) provide proof-of-concept that reoptimization can provide significant savings and better customer service. They also serve as prototypes for commercial-grade software that can be integrated into Metro Mobility's current system. We created our codes with a modular structure so that the data management and application of routines can be automated by a programmer with knowledge of Metro Mobility's current system. If Metro Mobility decides to implement a reoptimization routine, it will need to automate the following three steps: (1) export passenger data and current route/trip information to data files readable by OPL Studio (or another optimization solver),

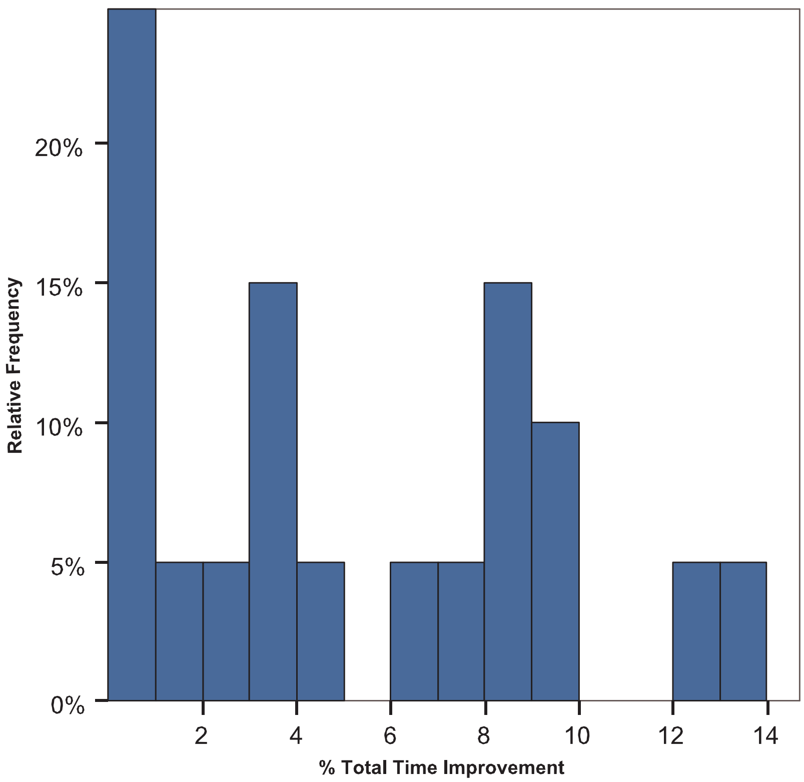


Figure 3.5: Relative frequency of percent total time improvement for reoptimizing 2 routes at a time

or write OPL Studio code to directly read data from system databases; (2) import the reoptimized solution back into Trapeze; and (3) build capability to change certain parameters manually in the reoptimization routine. Examples of the latter include the maximum computation time allowed after which an effort to reoptimize a particular route is terminated, and the criterion for selecting which two routes should be paired to realize savings from insertion and swapping of customers.

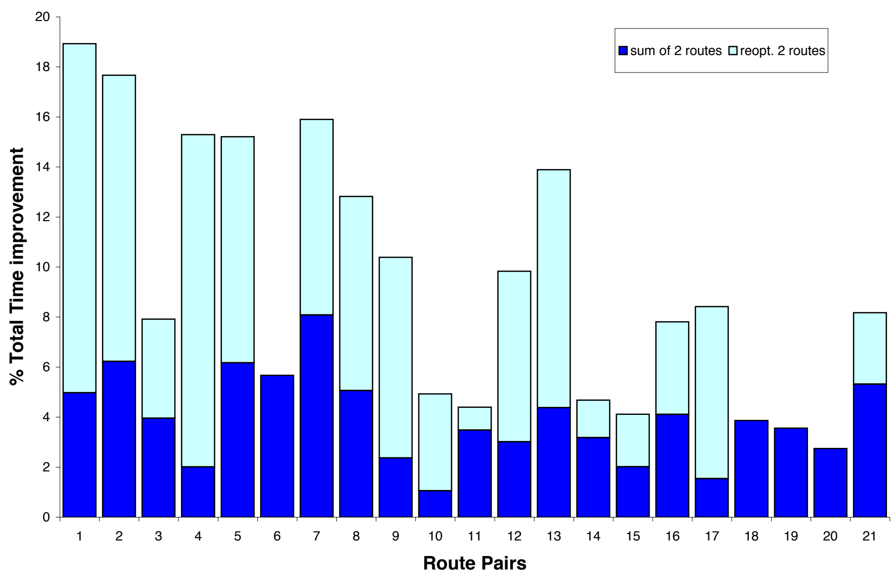


Figure 3.6: Percent total time improvement of reoptimizing the routes individually and reoptimizing 2 routes at a time

Table 3.5: Reoptimizing individual routes with 2 to 12 passengers (Transit Team)

Number of Passengers	Travel Time: Trapeze Solution	Travel Time: Reoptimized Solution	Total Time: Trapeze Solution	Total Time: Reoptimized Solution	% Total Time Improvement	Computation Time (seconds)
2	1:12	1:12	1:21	1:12	11.11	0.0
4	1:53	1:53	2:59	2:29	16.76	0.0
5	2:55	2:31	4:00	3:30	12.50	0.0
8	4:50	4:25	5:56	5:26	8.43	0.3
8	4:24	4:10	5:57	5:44	3.64	1.0
9	5:17	5:17	6:32	6:03	7.40	3.8
10	6:44	6:25	8:25	8:06	3.76	0.5
10	5:54	5:54	6:24	6:18	1.56	0.4
10	4:50	4:07	7:35	4:51	36.04	58.2
10	5:53	5:19	9:06	8:42	4.40	0.7
10	5:03	4:48	6:35	6:06	7.34	13.4
11	7:09	6:27	7:40	7:24	3.48	0.3
11	4:16	4:01	4:44	4:27	5.99	12.8
11	7:03	7:03	7:47	7:40	1.50	1.0
11	7:00	6:53	8:28	7:58	5.91	0.5
12	6:22	6:03	6:22	6:03	4.97	143.4
12	6:21	5:54	8:05	7:53	2.47	1.0
12	6:47	6:15	9:11	8:41	5.44	0.8
12	6:40	6:11	8:18	7:58	4.02	0.4
12	7:33	6:37	7:33	6:59	7.51	5.4
12	6:25	6:22	8:29	7:59	5.89	2.8
12	7:03	6:49	8:52	8:25	5.08	3.0
12	6:59	7:10	8:35	8:19	3.11	48.8
12	5:34	5:35	6:41	6:22	4.74	32.9

Table 3.6: Reoptimizing individual routes with 13 to 16 passengers (Transit Team)

Number of Passengers	Travel Time		Total Time		% Total Time Improvement	Computation Time (seconds)
	Trapeze Solution	Reoptimized Solution	Trapeze Solution	Reoptimized Solution		
13	5:43	5:37	8:50	8:47	0.57	189.1
13	7:28	7:03	10:15	10:00	2.44	1.7
13	7:21	6:27	9:16	8:50	4.68	9.6
13	7:15	7:09	8:08	7:38	6.15	7315.4
13	7:42	7:23	8:56	8:42	2.61	1.4
13	6:52	6:52	9:06	8:36	5.49	4.8
13	6:03	5:55	9:02	8:42	3.69	0.6
13	8:02	7:42	10:09	10:07	0.33	4.2
13	6:45	6:26	7:24	7:17	1.58	406.8
14	7:01	6:54	9:07	8:44	4.20	3.3
14	6:41	6:26	9:01	8:38	4.25	2.4
14	6:20	6:13	7:27	7:15	2.68	72.1
14	7:32	7:09	9:37	9:22	2.60	22.8
14	6:35	6:15	9:07	8:45	4.02	3.8
15	6:31	6:03	7:11	6:48	5.34	17973.0
15	8:14	7:35	9:40	9:11	5.00	12.6
15	8:16	7:57	10:05	10:02	0.50	402.0
15	7:28	7:04	8:33	8:21	2.34	2291.4
15	6:34	6:14	9:03	8:30	6.08	713.5
15	7:23	6:42	9:30	8:52	6.67	5.3
15	8:01	7:32	10:31	10:12	3.01	13.6
15	6:44	6:13	8:19	7:55	4.81	208.5
15	6:42	6:09	7:54	7:44	2.11	8.0
16	9:29	9:02	10:16	9:39	6.01	393.9
16	7:49	7:25	9:27	9:13	2.47	269.8
16	6:33	5:45	9:19	8:55	4.29	3.0
16	8:51	8:34	9:09	9:02	1.28	18.0

Table 3.7: Reoptimizing individual routes with 17 to 21 passengers (Transit Team)

Number of Passengers	Travel Time		Total Time		% Total Time Improvement	Computation Time (seconds)
	Trapeze Solution	Reoptimized Solution	Trapeze Solution	Reoptimized Solution		
17	8:10	7:44	9:49	9:10	6.62	2031.7
17	8:35	7:50	9:23	8:44	6.93	2449.2
17	7:27	7:22	9:35	9:11	4.17	442.6
17	7:42	7:03	10:04	9:33	5.13	257.6
17	9:17	8:51	10:33	10:02	4.90	172.5
18	8:11	7:54	9:58	9:39	3.18	4284.2
18	7:01	6:24	9:49	9:25	4.07	10800.0
18	9:09	9:02	10:16	9:52	3.90	110.6
18	8:20	8:12	10:18	9:50	4.53	18.3
18	9:52	9:15	10:23	10:18	0.80	742.9
18	8:49	8:24	9:17	9:03	2.51	10800.0
19	9:20	8:33	10:39	10:32	1.10	437.9
19	8:20	6:46	10:25	10:02	3.68	10374.1
21	9:49	9:36	10:24	10:14	1.60	10800.0
21	6:32	6:16	8:07	7:39	5.75	10800.0

Table 3.8: Reoptimizing individual routes with 3 to 7 passengers (Laidlaw)

Number of Passengers	Travel Time: Trapeze Solution	Travel Time: Reoptimized Solution	Total Time: Trapeze Solution	Total Time: Reoptimized Solution	% Total Time Improvement	Computation Time (seconds)
3	2:20	2:20	2:20	2:20	0.00	0.0
3	1:28	1:28	1:39	1:28	11.11	0.0
3	1:40	1:37	2:56	2:34	12.50	0.0
3	1:01	1:01	2:47	2:17	17.96	0.0
4	2:05	1:53	3:04	2:34	16.30	0.0
4	2:08	1:30	2:11	1:53	13.74	0.0
4	2:31	2:31	2:37	2:31	3.82	0.0
4	1:42	1:41	2:04	1:41	18.55	0.1
4	3:06	2:40	3:40	3:08	14.55	0.0
4	1:56	1:54	2:39	2:14	15.72	0.0
5	2:27	2:27	3:01	2:33	15.47	0.0
5	2:36	2:32	3:08	2:50	9.57	0.1
5	2:39	2:19	3:06	2:36	16.13	0.0
5	2:56	2:51	5:12	5:12	0.00	0.0
5	2:40	2:30	2:49	2:30	11.24	0.0
5	2:06	2:38	2:17	2:17	0.00	0.5
5	2:18	2:00	2:18	2:18	0.00	0.1
5	2:13	1:53	4:42	4:38	1.42	0.0
6	2:06	2:32	3:22	2:44	18.81	32.3
6	3:49	3:48	5:41	5:17	7.04	0.1
6	2:49	2:44	2:50	2:44	3.53	0.5
6	2:34	2:34	2:51	2:34	9.94	0.1
6	2:37	2:29	3:35	3:15	9.30	0.1
7	3:47	3:32	3:47	3:32	6.61	0.4
7	4:33	4:29	7:50	7:24	5.53	0.1
7	3:05	2:46	5:40	5:10	8.82	0.3
7	3:24	3:13	7:31	7:01	6.65	0.0
7	2:38	2:36	3:26	2:58	13.59	0.2
7	2:40	2:40	2:52	2:40	6.98	0.8
7	3:03	2:48	3:09	2:48	11.11	0.3
7	4:22	4:12	6:56	6:26	7.21	0.1
7	3:18	3:18	3:18	3:18	0.00	0.9
7	3:58	3:56	7:48	7:28	4.27	0.1

Table 3.9: Reoptimizing individual routes with 8 to 10 passengers (Laidlaw)

Number of Passengers	Travel Time		Total Time		% Total Time Improvement	Computation Time (seconds)
	Trapeze Solution	Reoptimized Solution	Trapeze Solution	Reoptimized Solution		
8	3:10	3:01	3:13	3:11	1.04	0.5
8	4:34	4:02	7:19	7:08	2.51	0.1
8	4:55	4:42	7:28	7:20	1.79	0.1
8	3:48	3:35	7:23	6:53	6.77	0.0
8	4:38	4:31	7:26	6:56	6.73	0.1
8	4:20	4:20	7:35	7:07	6.15	0.1
8	4:52	4:32	6:49	6:21	6.85	0.0
8	4:50	4:25	8:58	8:30	5.20	0.1
8	3:25	3:19	6:00	5:30	8.33	0.2
8	3:39	3:18	6:36	6:08	7.07	0.1
9	4:24	4:11	7:05	6:47	4.24	0.3
9	4:59	4:51	7:16	7:07	2.06	1.9
9	4:55	4:44	7:36	7:18	3.95	0.2
9	5:16	4:15	9:02	8:31	5.72	0.1
9	5:31	4:52	6:46	6:30	3.94	0.2
9	3:42	3:19	4:07	3:23	17.81	0.9
9	5:43	5:02	8:20	7:50	6.00	0.1
10	5:09	5:06	6:42	6:12	7.46	0.7
10	5:05	4:31	7:31	7:06	5.54	0.7
10	4:20	4:10	5:40	5:10	8.82	0.4
10	5:17	5:02	8:06	7:45	4.32	0.5
10	5:57	4:56	7:43	7:22	4.54	0.2
10	3:58	3:38	7:57	6:45	15.09	0.4
10	5:58	5:28	6:30	6:30	0.00	0.5
10	4:07	3:45	6:51	6:27	5.84	3.2
10	5:44	5:42	8:03	7:46	3.52	0.2
10	5:49	5:53	10:10	9:58	1.97	0.4
10	4:47	4:22	6:24	5:53	8.07	4.8
10	4:51	4:47	7:17	6:47	6.86	0.1

Table 3.10: Reoptimizing individual routes with 11 to 16 passengers (Laidlaw)

Number of Passengers	Travel Time: Trapeze Solution	Travel Time: Reoptimized Solution	Total Time: Trapeze Solution	Total Time: Reoptimized Solution	% Total Time Improvement	Computation Time (seconds)
11	4:44	4:29	7:44	7:13	6.68	0.8
11	5:32	5:32	9:55	9:25	5.04	0.6
11	4:49	4:43	7:07	6:37	7.03	2.9
11	6:04	6:01	7:36	7:20	3.51	0.4
11	5:47	5:40	7:29	7:10	4.23	6.5
12	5:43	5:21	9:36	9:11	4.34	5.6
12	6:15	5:46	8:20	7:53	5.40	2.6
12	6:02	5:20	7:50	7:22	5.96	12.0
12	6:01	5:32	9:12	8:43	5.25	1.3
12	6:51	6:36	9:44	9:29	2.57	0.8
13	7:35	7:33	9:00	8:51	1.67	0.7
13	5:32	5:30	8:03	8:03	0.00	43.4
13	6:29	6:20	10:01	9:47	2.33	1.6
13	6:17	6:06	6:35	6:17	4.56	6.1
13	5:24	5:14	9:29	9:01	4.92	31.9
13	4:39	4:41	7:24	7:06	4.05	10.9
13	7:13	7:16	9:17	9:15	0.36	3.1
14	7:09	6:38	9:07	9:02	0.91	1.2
14	7:19	6:57	9:36	9:09	4.69	14.0
14	7:10	7:05	10:12	9:46	4.25	0.9
15	7:56	7:13	8:45	8:15	5.71	6.9
15	6:08	5:54	7:07	6:50	3.98	7.6
16	6:08	5:48	8:15	7:58	3.43	74.4
16	9:55	9:15	11:06	10:45	3.15	62.6

Table 3.11: Route 554 with 9 passengers (Laidlaw)

Trapeze Solution		Reoptimized Solution	
Node	Arrival Time	Node	Arrival Time
Nodes	Arrival Times	Nodes	Arrival Times
17	14:00	17	14:30
18	14:07	18	14:37
16	14:45	16	14:50
13	15:08	13	15:13
11	15:19	11	15:24
12	15:24	12	15:30
1	15:48	1	15:54
3	16:07	3	16:13
4	16:11	4	16:17
5	16:11	5	16:17
8	**16:32**	**7**	**16:31**
7	**16:49**	**9**	**16:40**
9	**16:58**	**10**	**16:52**
10	**17:09**	**8**	**17:01**
6	17:27	6	17:13
2	17:30	2	17:16
14	17:58	14	17:44
15	18:07	15	17:53
Total Time	4:07		3:23
Total Travel Time	3:42		3:19

Table 3.12: Route 702 with 12 passengers (Transit Team)

Trapeze Solution		Reoptimized Solution	
Node	Arrival Time	Node	Arrival Time
Nodes	Arrival Times	Nodes	Arrival Times
9	5:25	9	5:48
15	**5:56**	**17**	**6:15**
13	6:01	13	6:30
22	**6:25**	**15**	**6:38**
17	**6:35**	**21**	**6:50**
21	**6:51**	**18**	**7:01**
18	**7:02**	**20**	**7:05**
20	**7:06**	**24**	**7:24**
24	**7:25**	**19**	**7:38**
19	**7:39**	**22**	**7:52**
16	8:15	16	8:22
10	8:34	10	8:41
7	8:54	7	9:01
4	9:02	4	9:09
2	9:13	2	9:20
3	9:20	3	9:26
8	9:34	8	9:41
14	9:57	14	10:04
1	10:24	1	10:31
6	10:51	5	10:54
5	10:56	6	11:07
11	11:15	11	11:20
12	11:25	12	11:30
23	11:47	23	11:51
Total Time	6:22		6:03
Total Travel Time	6:22		6:03

Table 3.13: Summary of Productivity Improvements from First-Pass Analysis

	LaidLaw	**Transit Team**
Date of data extract	June 1, 2005	May 9,2007
Total number of routes	103	110
Number of routes compared	85	66
Original Total Time (hours)	539.5	559.0
Optimized Total Time (hours)	509.5	533.6
Time reduced (hours)	30.0	25.4
Average Time reduced per route (minutes)	21.1	23
Total number of customers	749	897
Average number of customers per route	8.8	13.6
Productivity - Trapeze Solution	1.39	1.60
Productivity - After reoptimization	1.47	1.68

Table 3.14: Criteria for selecting route pairs

Route Pairs	Criteria		Legend
525 & 535	B, D		A: both routes have more than 2.5 hours idle time
525 & 513	F		B: both routes have less than 2.5 hours idle time
551 & 460	A, E		C: one route has more than and one has less than 2.5 hours idle time
551 & 426	F		D: one route's end time is in the middle of the other route's trip
400 & 691	C, D		E: the routes' start times differ by more than 3 hours
400 & 507	F		F: the second route has maximum map page overlap with the first
400 & 406	F		
408 & 513	A, D		
408 & 403	F		
412 & 430	C, E		
412 & 530	F		
533 & 443	C		
533 & 502	F		
450 & 690	A		
450 & 412	F		
517 & 406	B		
517 & 532	F		
403 & 559	B, E		
403 & 560	F		
500 & 540	B, D		
500 & 530	F		

Table 3.15: Two Routes at a Time with 15 to 26 passengers (Laidlaw)

Pair Label	Routes	Total # of Pass	Original		Reoptimized		% Total Time Impr.	% Page Overlap	Comp Time (Seconds)
			Travel Time	Total Time	Travel Time	Total Time			
1	525 & 535	16	9:30	13:03	9:44	11:08	14.69	20.00	108.45
2	525 & 513	15	8:03	13:48	7:43	12:07	12.20	54.55	101.72
3	551 & 460	18	9:06	16:59	8:27	16:17	4.12	25.00	131.33
4	551 & 426	20	10:44	17:05	10:03	14:46	13.56	55.56	206.37
5	400 & 691	18	9:31	14:42	9:35	13:17	9.64	46.15	156.12
6	400 & 507	25	11:00	13:02	-	-	0.00	81.82	309.21
7	400 & 406	20	9:16	11:22	9:28	10:24	8.50	63.64	187.15
8	408 & 513	18	8:26	15:18	8:44	14:03	8.17	30.77	139.88
9	408 & 403	23	10:49	15:49	10:57	14:31	8.22	63.64	242.92
10	412 & 430	18	10:24	14:03	10:03	13:30	3.91	25.00	194.05
11	412 & 530	25	13.05	17:34	12:47	17:24	0.95	63.64	312.72
12	533 & 443	19	10:03	14:28	8:35	13:27	7.03	55.56	181.95
13	533 & 502	22	11:33	16:45	11:00	15:05	9.95	62.50	245.53
14	450 & 690	20	10:16	17:15	10:18	16:59	1.55	33.33	203.75
15	450 & 412	26	13:53	18:38	13:32	18:14	2.15	50.00	293.9
16	517 & 406	20	9:38	11:40	9:33	11:13	3.86	38.46	211.82
17	517 & 532	22	12:04	15:59	12:01	14:52	6.99	58.33	204.94
18	403 & 559	24	11:00	14:55	-	-	0.00	53.85	319.35
19	403 & 560	24	11:25	15:49	-	-	0.00	80.00	295.87
20	500 & 540	25	13:02	17:08	-	-	0.00	15.38	336.94
21	500 & 530	24	11:18	16:36	11:07	16:06	3.01	40.00	313.39

Table 3.16: Two Routes at a Time with 18 to 21 passengers (Transit Team)

Pair Label	Routes	Total # of Pass	Original		Reoptimized		% Total Time Impr.	% Page Overlap	Comp Time (Seconds)
			Travel Time	Total Time	Travel Time	Total Time			
1	722 & 768	18	10:49	13:30	11:01	12:40	6.17	33.15	103.37
2	722 & 772	20	12:18	14:22	12:18	13:43	4.52	28.35	141.72
3	722 & 729	21	12:51	15:28	13:14	15:05	2.48	26.42	201.31

Chapter 4

The Non-Dedicated Option

Taxis are on average more expensive per passenger trip than dedicated vehicles. Still, using taxis selectively for a few passengers can help lower costs. For example, suppose the last pick up and drop off on a route adds one hour to the length of that route. This costs approximately $42 according to the current payment schedule. However, if this customer's trip would cost approximately $20 by taxi, then it will be worthwhile to offer a taxi option to the passenger who is scheduled last on the route and save $22 in case the passenger accepts this offer. The main challenge for evaluating the benefit of using nondedicated vehicles is therefore that of finding a way to identify which customers should be offered a taxi service and develop realistic estimates of the cost savings that will result after accounting for passenger preferences. In this Chapter, we focus on developing and testing an algorithm for choosing which passengers should be offered taxi services and the cost impact of such decisions.

We used our algorithm for selecting candidate passengers for taxi service on reoptimized routes of both providers (after First-Pass in Chapter 3) under different values of the proportion of customers who agree to take the taxi service. Note, the taxi service is offered at no additional cost to the passenger. The reason for offering taxi service is to reduce overall cost to the system. It is not to be confused with premium taxi service which is a pilot program being tested by MM to accommodate customers whose requests cannot be met. Premium service costs more to the customers.

We found that Metro Mobility can realize further savings by selectively using taxi services. Specifically, if half of the passengers who are offered a taxi service result in a successful match, Laidlaw's daily savings average $113.4 when taxis cannot handle wheelchairs, and $170.7 when they can. Similar daily savings for Transit Team are $124.7 and $182.4. We also computed the time of day when most of the taxi rides are needed and found that although more taxis are used during rush hours, the total number is still quite small.

The algorithm for finding candidate passengers for nondedicated vehicles can be applied just as easily to the solution generated by Trapeze. It is not necessary to perform the reoptimization step described in Chapter 3. This algorithm is simple to use and computation times are modest. Therefore, we recommend testing the use of nondedicated vehicles in a pilot if the potential savings identified in this Chapter are deemed attractive. A pilot program will provide critical data on (1) the proportion of passengers who accept an offer to travel by taxi, (2) the availability of taxis at the time needed, and (3) the availability of taxis that can take passengers with special needs such as wheelchairs. Pilot should also monitor program costs and benefits during the test period since possible increase in cost from administering nondedicated vehicle services and possible fuel-cost savings from reduced miles driven by dedicated vehicles are not included in our analysis for lack of appropriate data.

4.1 The Algorithm

Our approach has two steps. First, we develop a procedure for identifying the preferred passenger who would be called and offered the option to travel by taxi. The choice of the next preferred passenger may depend on the outcome of the offer made to the previous preferred passenger. Then, we test this procedure in a computer simulation where the probability that a passenger will accept the offer of riding in the taxi can be changed. The purpose of this exercise is to develop realistic estimates of savings from using taxis and to identify the number and time-of-day of the requests for taxi rides.

We start by developing a criterion to qualify routes from which customers are offered taxi service. Qualified routes are those that take either between 4 and 6 hours to complete or more than 8 hours. Routes that take less than 4 hours or between 6 and 8 hours are not included in the analysis. The rationale for this choice is that due to union rules, it appears that drivers need to be paid for at least 4 hours on any shift. Drivers who operate dedicated vehicles for between 6 and 8 hours on any given day are assumed to receive credit for a full 8-hour shift. The procedure for qualifying routes is not necessary to implement our algorithm. Also, the thresholds we used were selected based on what we think best describes the Metro Mobility situation. These thresholds can be changed easily.

After identifying qualified routes, we developed a heuristic approach that focuses only on the last passengers in each qualified route and rank orders them on the basis of potential savings. The ranking procedure works as follows. We simulate removing the last customers one by one from each qualified route. After each removal, we rerun the second part of the First-Pass algorithm of Task 2. This algorithm calculates the optimal pick up and drop off times that meet all constraints and minimize the total time to complete the route. Then, we calculate the net savings resulting from the removal of the customer by subtracting the taxi cost from savings generated from reducing total time. At each decision epoch, the passenger who results in maximum savings is called the preferred customer. Ties are broken arbitrarily.

We allow for the possibility that some customers may prefer to use dedicated services or that a taxi may not be available at the time requested. In some other cases, passengers may require extra care that cannot be provided by non-dedicated vehicle providers. For example, customers who use lift (LF) or scooter (SC) generally cannot travel by taxi. These factors are combined into a successful match probability that a preferred customer who is offered the option to travel by taxi will result in the use of taxi service.

We discard those routes on which the savings from removing the last customer are smaller than a prespecified cutoff. This cutoff was set to $10 in our experiments described in the next section. However, the algorithm accommodates any reasonable value of the cutoff. The value of the cutoff reflects the cost of administering each taxi service request. A complete schematic of the logic behind our algorithm is shown in Figure 4.1. If a customer turns down the offer to travel by taxi, then the route to which the customer belongs is removed from the list of qualified routes. If both the taxi provider and the customer accept the taxi option, we remove the last customer from the route in question, treat the second last customer as the last customer and recompute the savings from removing this customer according to the procedure described above. The algorithm stops when the savings from the preferred customer at a decision epoch fall below the cutoff or when there are no qualified routes, whichever occurs first.

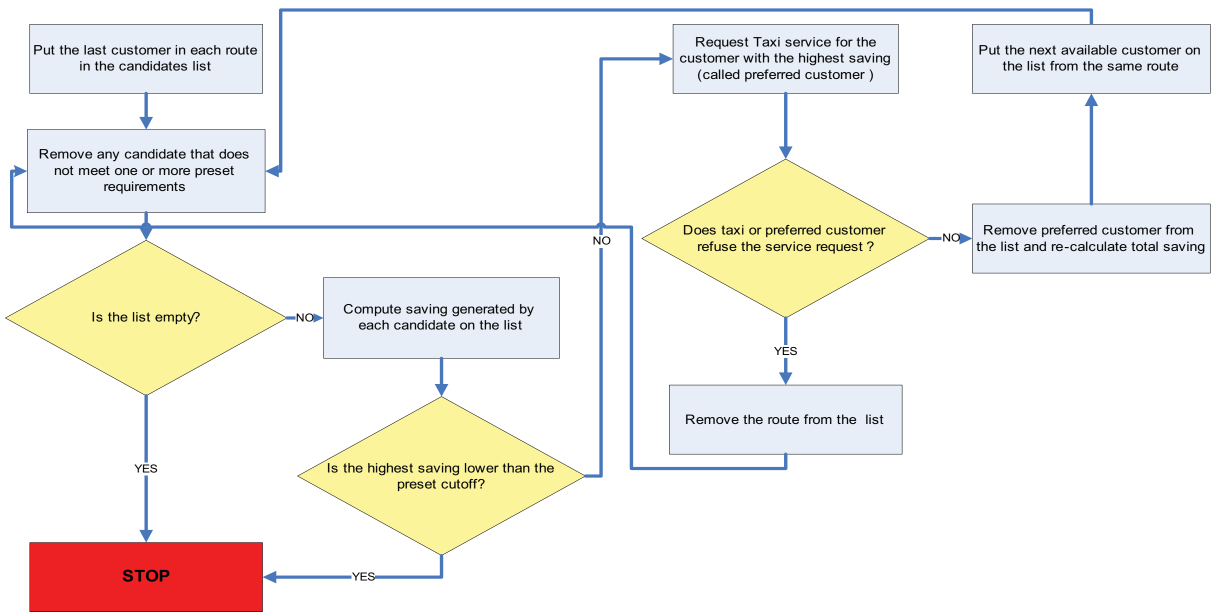


Figure 4.1: A Schematic of the Algorithm

4.2 Results From Computer Simulation Experiments

As mentioned earlier, some customers may not accept the option to travel by taxi even if it does not cost them more to take the taxi option. Moreover, during rush hours, there might not be enough taxis to provide the desired service. Therefore, we carried out computer simulation experiments to evaluate the impact of using nondedicated services.

In these experiments, we started with routes generated by the First-Pass reoptimization algorithm and selected only those routes that take between 4 and 6 hours to complete or more than 8 hours. We assumed that the cost of a revenue hour is $42 and the cost of taxi service is calculated as follows: taxi charges = distance $\times$ 1.05 $\times$ 1.9 + 2.5. This is the same formula that is used by Metro Mobility for calculating premium taxi service charges. Figure 4.2 shows simulation results for Laidlaw when the probability that the offered taxi service option is successful ranges from 0.1 to 1.0 in steps of 0.1. We use dark-shaded bars to indicate experiments in which we assumed that taxis cannot handle customers with wheelchairs (WC) and light-shaded bar to indicate situations in which they can. The error bars indicate 95% confidence intervals.

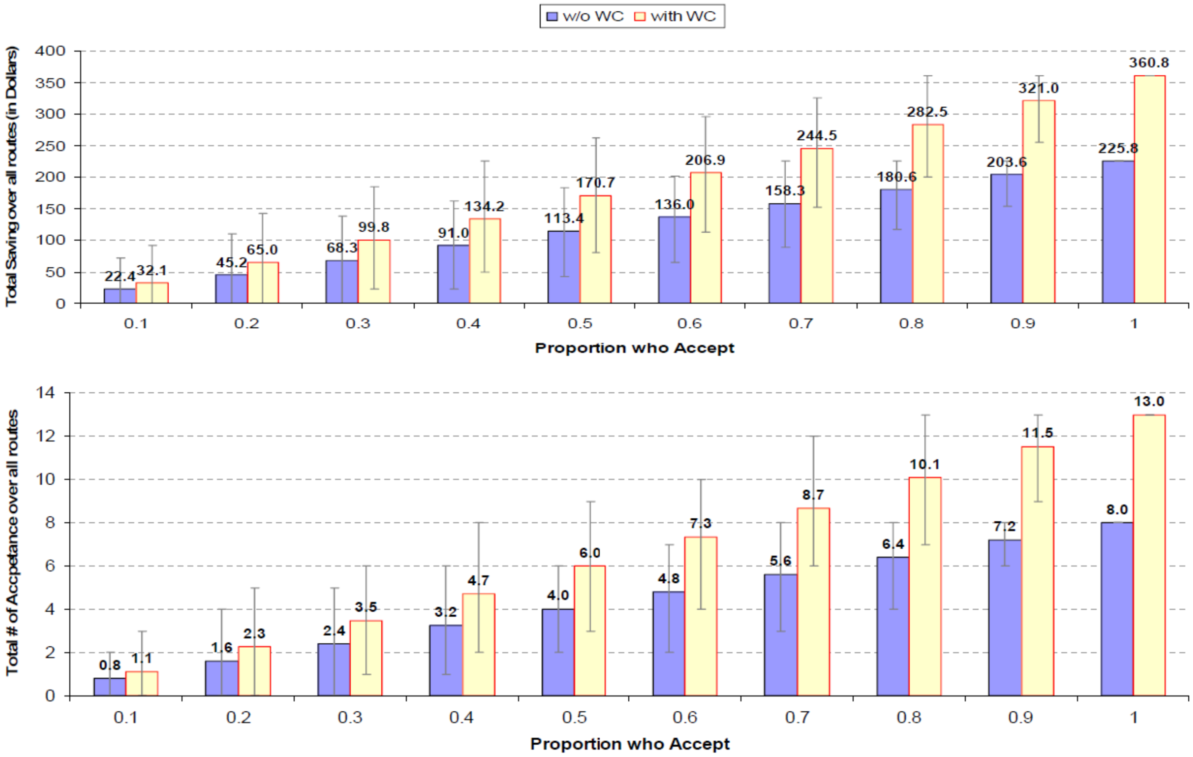


Figure 4.2: Simulation Results for Laidlaw

Simulation results for Transit Team are shown in the Figure 4.3. Here too, selective use of nondedicated vehicles can save money. However, Metro Mobility will end up having more customers from Transit Team service area being served by taxis. This is in part due to the fact that Transit Team has longer routes, which often exceed 8 hours, and that Transit Team has more customer trips.

The availability of taxis can also play a role in the success of using the nondedicated services option. Figure 4.4 shows maximum taxi requirements that arise in simulation experiments at different times of the day. Note that many more requests are made in the period from 3 PM to 5 PM. Since this is the time when demand for taxis is usually high, additional MM demand during this period could pose a problem. However, the number of taxis required is relatively small, and appears to be quite manageable for a metropolitan area of the size of the Twin Cities. In Figure 4.4, the taxi requirements are calculated after assuming that every customer who is offered a taxi option accepts this offer. Thus, the reported numbers are upper bounds on the number of taxis that would be required on a typical day of operations.

Figure 4.5 shows the impact on productivity of performing reoptimization (first pass

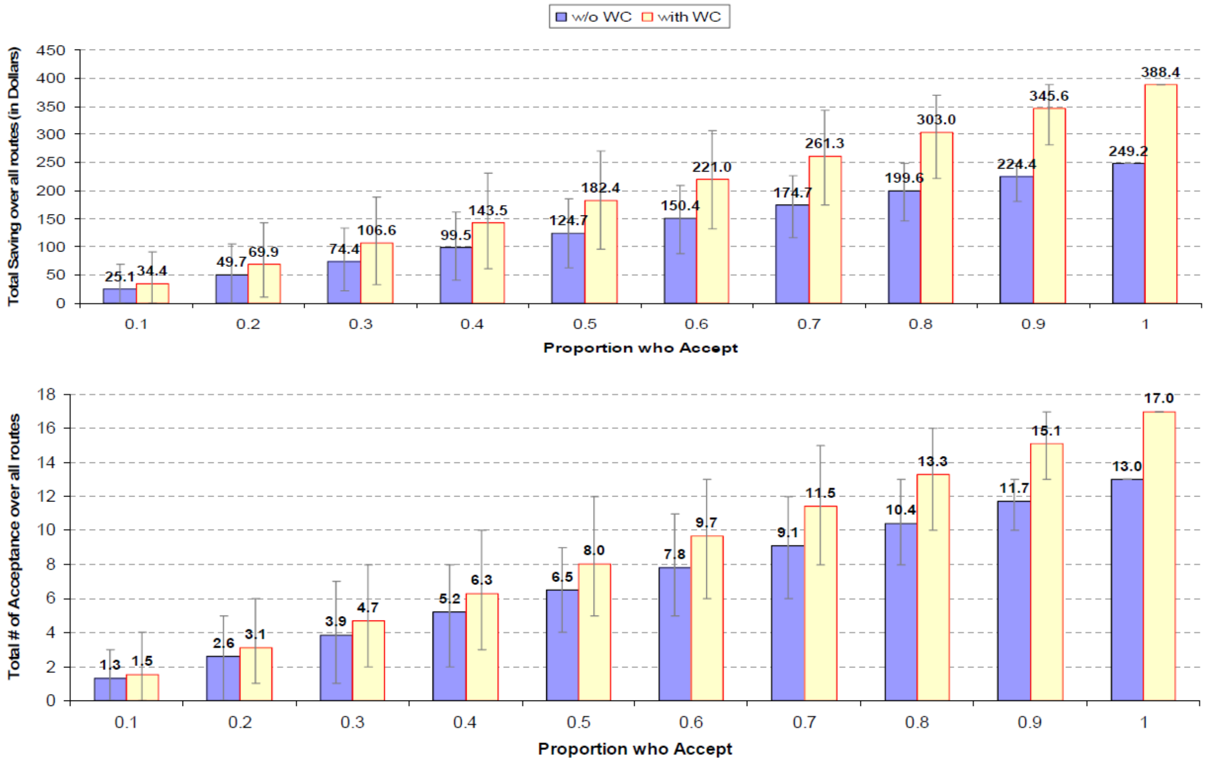


Figure 4.3: Simulation Results for Transit Team

only) and selectively using taxis. For Transit Team, the productivity increases from 1.60 to 1.68 after First-Pass and approaches 1.71 when non-dedicated services are factored in. For Laidlaw, the productivity increases from 1.39 to 1.47 after First-Pass reoptimization and approaches 1.5 when non-dedicated services are factored in. In other words, both route re-optimization and use of nondedicated services can increase productivity and improve service quality.

4.3 Conclusions

The use of nondedicated vehicles (in particular taxis) can result in average daily savings over both contractors of \$118.5 when taxis do not accept WC customers, and \$176.6 when they do, assuming 50% of the offers to travel by taxi result in a successful match. More taxis are needed during 3-5 PM, but these requirements are not too large.

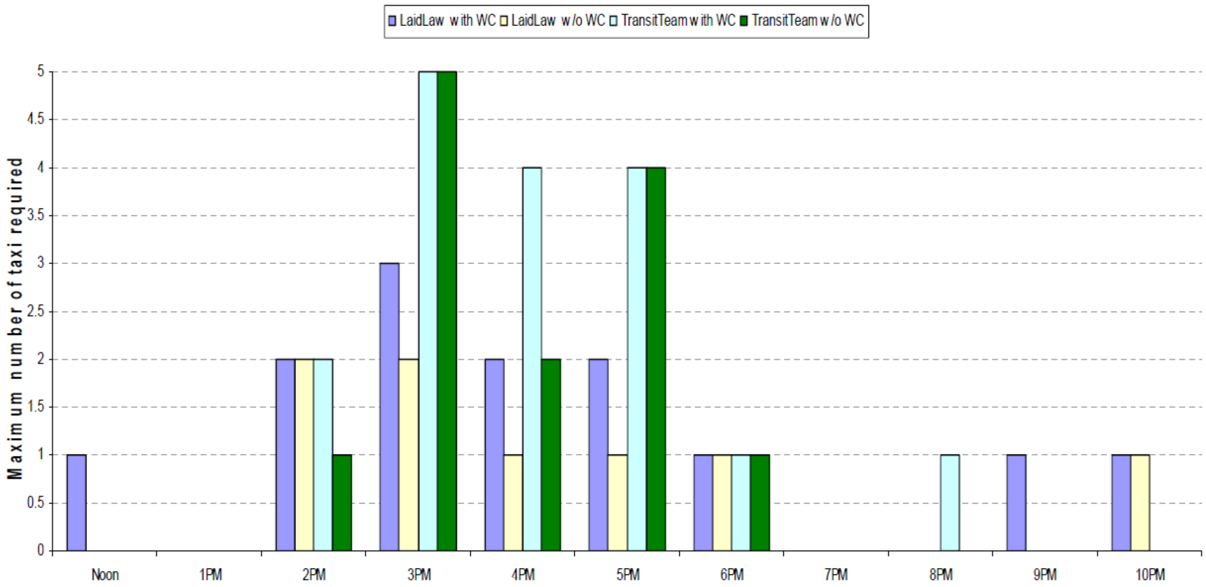


Figure 4.4: Taxi Requirements

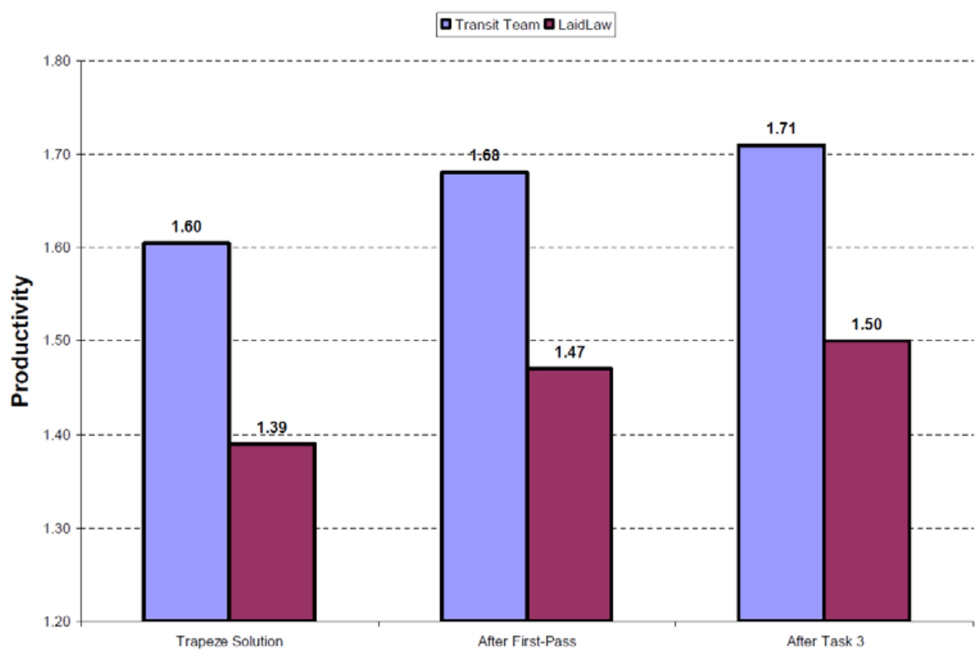


Figure 4.5: Productivity Comparison

Chapter 5

Concluding Remarks

The remarks in this section both summarize our findings and present a set of recommendations that can help improve the efficiency of Metro Mobility's operations. These remarks are based on MM's data and operational characteristics. However, other similar service providers can benefit from using techniques described in this report after adjusting the analysis to suit their particular environments.

In Chapter 2, we found a weak positive correlation between certain performance metrics, e.g. percent cancelations and no shows, and demand. This underscores the difficult challenge that a service provider faces when attempting to match capacity with demand because greater cancelations and no shows decrease the efficiency of capacity utilization precisely when capacity is tight. This Chapter also identifies potential moral hazard situations for contractors. That is, the selection of contract parameters by Metro Mobility does not prevent contractors from earning more by lowering productivity during low demand periods even after paying damages.

In Chapter 3, we reported a reoptimization approach that can improve efficiency of route formation by up to 5% when routes are studied one at a time and by an additional 5% when pairs of routes are reoptimized. In the latter scheme, passengers can be moved (or swapped) from one route to the other. This procedure does not require contractors to change their current booking and route formation procedures. The reoptimization can be performed every night when the trip booking process is shut down.

In Chapter 4, we developed an algorithm for selecting passengers who could be offered a taxi service at no extra cost to them. Although taxis are in general more expensive than dedicated vehicles, our algorithm identifies those passengers that can be served more economically by taxis. The savings from the non-dedicated (taxi) depend on the proportion of passengers who would accept this offer by the service provider. Using available data and reasonable assumptions, we estimated the savings to Metro Mobility to be in hundreds of dollars per day. The number of taxis needed during peak demand hours is quite manageable. Use of non-dedicated option does not require the use of reoptimization routine and our methodology is relatively easy to implement.

Main recommendations of this study are as follows.

1. Metro Mobility, and paratransit service providers in general, need to examine terms of

contracts they offer and attempt to eliminate moral hazard.

2. We recommend that MM implement a reoptimization routine. This approach can be implemented without changing the current operating environment. Although the implementation requires additional programming work, the savings far outweigh the potential costs of developing a functional system. Reoptimization can be repeated at the end of each day of booking. It maintains the pick up and drop off times within the ADA and MM prescribed time-windows of their originally scheduled values. Thus, there is no need to contact the customers after running the reoptimization routine.

3. We recommend a pilot program to ascertain the proportion of passengers who accept an offer to travel by taxi, availability of taxis at the time needed and ability of taxis to carry passengers with different needs. We did not factor the increased administrative burden of providing taxi service in our calculations. We also did not include additional fuel-cost savings that may come from fewer miles traveled because of lack of reliable data. A pilot should monitor program costs and benefits carefully to ascertain if the savings match those observed in the computer simulation experiments.

Use of nondedicated services does not require making any change to the current trip booking process. Our approach can be used starting with any existing routes' schedule. Therefore, it seems reasonable to test this option first if the magnitude of savings reported here is deemed attractive.

In Chapter 3, our methodology takes the routes generated by Trapeze as given. Therefore, its benefit comes primarily from improving the sequence of stops along a route and from moving/swapping passengers from one route to another. Significant additional savings may be possible from considering dynamic optimization algorithms for scheduling trips as requests are received. Therefore, we recommend a critical evaluation of Trapeze software in comparison to other commercially available products for dynamic route optimization. An in-house Operations Research analyst can help in this task as well as develop MM-specific refinements to the commercial software currently used.

References

L. Bodin, B. Golden, A. Assad, and M. Ball. Routing and scheduling of vehicles and crews, the state of art. *Computers and Operations Research*, 10:69–211, 1983.

J.-F. Cordeau. A branch-and-cut algorithm for the dial-a-ride problem. *Operations Research*, 54(3):573–586, May –June 2006.

J.-F. Cordeau and G. Laporte. A tabu search heuristic for the static multi-vehicle dial-a-ride problem. *Transportation Research Part B*, 37:579–594, 2003.

J.-F. Cordeau, G. Laporte, and A. Mercier. A unified tabu search heuristic for vehicle routing problems with time windows. *J. of the Operational Research Society*, 52:928–936, 2001.

U. Derigs and G. Grabenbauer. Intime - a new heuristic approach to the vehicle routing problem with time windows, with a bakery fleet case. *American Journal of Mathematical and Management Sciences*, 13:249–266, 1993.

M. Desrochers and F. Soumis. A generalized permanent labeling algorithm for the shortest path problems with time windows. *INFOR*, 26:191–212, 1988.

M. Desrochers, J. Desrosiers, and M. Solomon. A new optimization algorithm for the vehicle routing problem with time windows. *Operations Research*, 40(2):342–354, Mar – Apr 1992.

J. Desrosiers, M. Sauve, and F. Soumis. Lagrangian relaxation methods for solving the minimum fleet size multiple traveling salesman problem with time windows. *Management Science*, 34(8):1005–1022, August 1988.

J. Desrosiers, Y. Dumas, and F. Soumis. Time constrained routing and scheduling. *In: Handbooks of operations research and management sciences: networks, North Holland, Amsterdam*, 1993.

M. L. Fisher, K. O. Jornsten, and O. B. G. Madsen. Vehicle routing with time windows: two optimization algorithms. *Operations Research*, 45(3):488–492, 1997.

L. Fu. Scheduling dial-a-ride paratransit under time-varying, stochastic congestion. *Transportation Research*, 36:485–506, 2002.

M. Gendreau, A. Hertz, and G. Laporte. New insertion and postoptimization procedures for the traveling salesman problem. *Operations Research*, 40:1086–1094, 1992.

M. Gendreau, A. Hertz, G. Laporte, and M. Stan. A generalized insertion heuristic for the traveling salesman problem with time windows. *Operations Research*, 46(3):330–335, 1998.

N. Kohl and O. B. G. Madsen. An optimization algorithm for the vehicle routing problem with time windows based on lagrangian relaxation. *Operations Research*, 45(3):395–406, May – Jun 1997.

Metropolitan Council. Options, alternatives, and strategies for future Metro Mobility/ ADA paratransit services. Technical report, Report prepared at the direction of the 2000 Minnesota Legislature, January 2001. available at `www.metrocouncil.org`.

J. Renaud, F. F. Boctor, and G. Laporte. A fast composite heuristic for the symmetric traveling salesman problem. *INFORMS Journal on Computing*, 8(2):134–143, 1996.

J. Renaud, F. F. Boctor, and J. Ouennhiche. A heuristic for the pickup and delivery traveling salesman problem. *Computers & Operations Research*, 27:905–916, 2000.

M. W. P. Savelsbergh and M. Sol. The general pickup and delivery problem. *Transportation Science*, 29:17–29, 1995.

M. Solomon. Algorithms for the vehicle routing and scheduling problem with time window constraints. *Operations Research*, 35:254–265, 1987.

M. M. Solomon and J. Desrosiers. Time window constrained routing and scheduling problems. *Transportation Science*, 22:1–13, 1988.

S. R. Thangiah, I. H. Osman, R. Vinayagamoorthy, and T. Sun. Algorithms for the vehicle routing problems with time deadlines. *American Journal of Mathematical and Management Sciences*, 13:323–355, 1993.

P. Toth and D. Vigo. Heuristic algorithms for the handicapped persons transportation problem. *Transportation Science*, 31(1):60–71, 1997.

N. Wilson and N. Colvin. Computer control of the rochester dial-a-ride system. Technical Report R77-31, Department of Civil Engineering, MIT, Cambridge, MA, 1977.

N. Wilson and H. Weissberg. Advanced dial-a-ride algorithms research project: final report. Technical Report R76-20, Department of Civil Engineering, MIT, Cambridge, MA, 1976.

N. Wilson, J. Sussman, H. Wong, and T. Higonnet. Scheduling algorithms for a dial-a-ride system. Technical report, Massachusetts Institute of Technology, Cambridge, MA, 1971.

Appendix A

Service Quality Measures

This section contains a summary of the analysis of Metro Mobility data pertaining to service quality measures. These results were presented to the Technical Advisory Panel for this project.

Service Quality Measures of contractors from 2001-Sept 2005

1. Percent on-time pick-ups
2. Percent missed trips
3. Percent on-time drop-offs
4. Percent acceptable ride times
5. Percent excessive ride times

1

Service provider's actual performance on February 2005

- Actual pick-up delay is calculated by subtracting the actual arrival time by the scheduled pick-up time
- Actual drop-off delay is calculated by subtracting the actual arrival time by the appointment time of the passengers
- Actual ride time delay is calculated by subtracting the actual arrival time of drop-off by the actual arrival time of pick-up

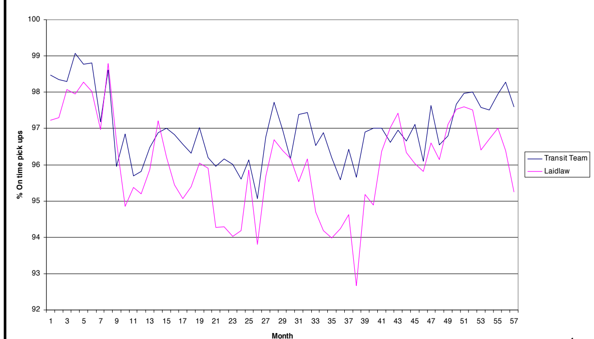
2

Service Quality Measure:

Percent on-time pick-ups

3

Percent on-time pick ups



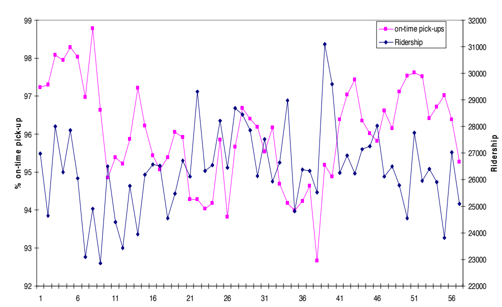
4

Percent on-time pick up and demand correlation

Laidlaw and Transit Team

5

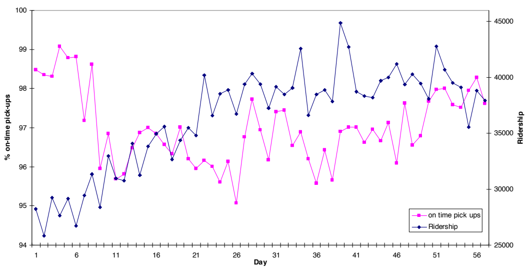
Laidlaw



•Correlation coefficient is -0.157 with p-value of 0.243 for Laidlaw

6

Transit Team



•Correlation coefficient is -0.284 with p-value of 0.032 for Transit Team

•There is a difference of .968 between TT's and Laidlaw's mean on-time performance with p-value of zero

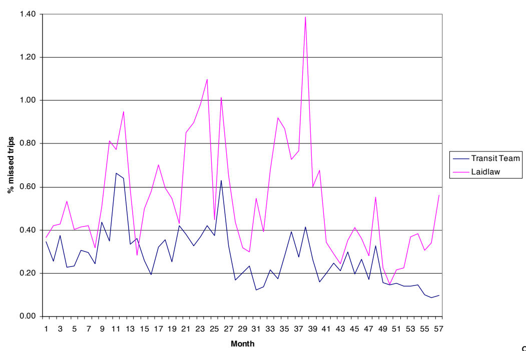
7

Service Quality Measure:

Percent missed trips

8

Percent missed trips



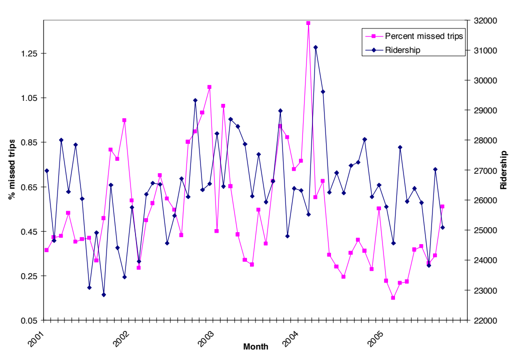
9

Percent missed trips and demand correlation

Laidlaw and Transit Team

10

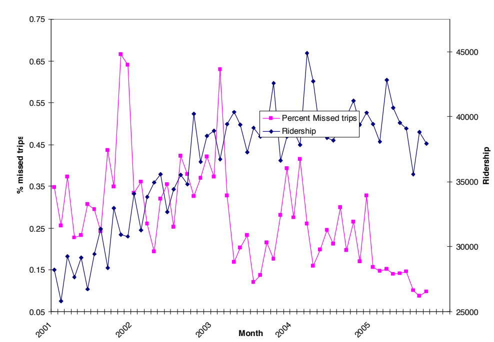
Laidlaw



•Correlation coefficient is 0.055 with p-value of 0.683

11

Transit Team



•Correlation coefficient is -0.400 with p-value of 0.002

•There is a difference of -0.26 between Transit Team's and Laidlaw's mean missed trips with p-value of zero

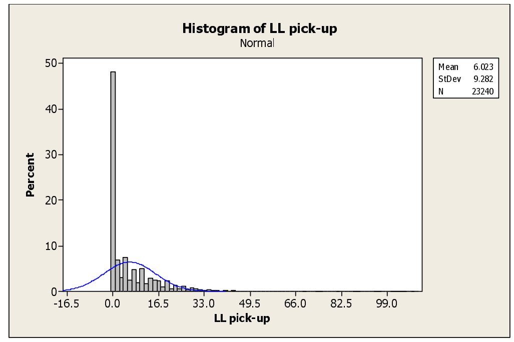
12

Actual performance

Pick-up delay

13

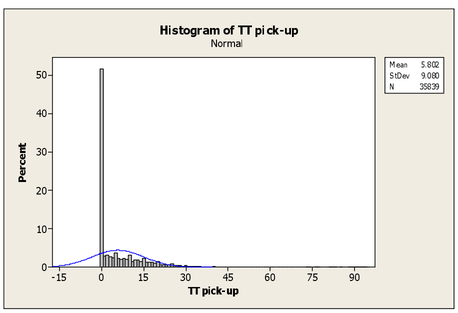
Laidlaw



•Zero minute contains early pick-up and delay of zero minute
•The histogram represents actual pick-up delay in minutes performed by Laidlaw for the month of February 2005 with its relative frequency

14

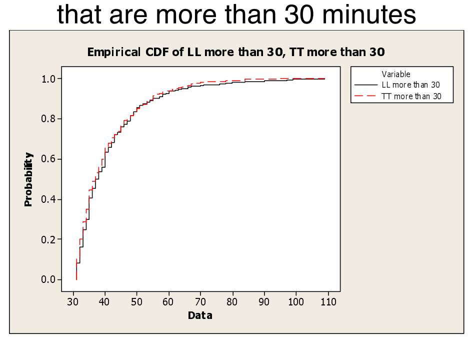
Transit Team



•The histogram represents actual pick-up delay in minutes performed by Transit Team for the month of February 2005 with its relative frequency

15

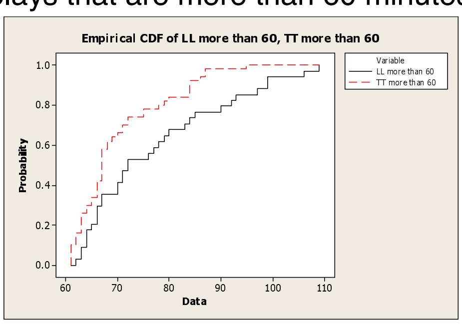
Cumulative Distribution Function of delays that are more than 30 minutes



•Laidlaw and Transit Team are close. Neither dominates the other.

16

Cumulative Distribution Function of delays that are more than 60 minutes



•Laidlaw has stochastically larger delays. Transit Team performs better.

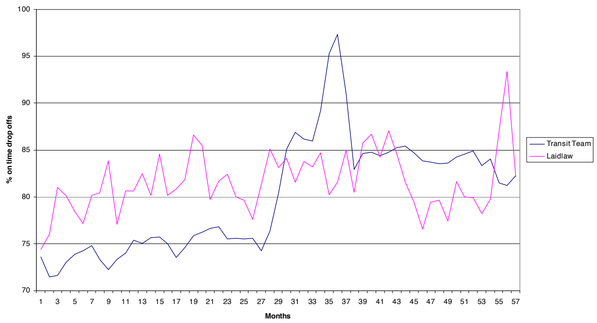
17

Service Quality Measure:

Percent on-time drop-offs

18

Percent on-time drop-offs



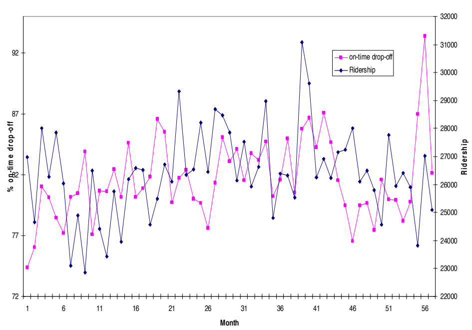
19

Percent on-time drop-offs and demand correlation

Laidlaw and Transit Team

20

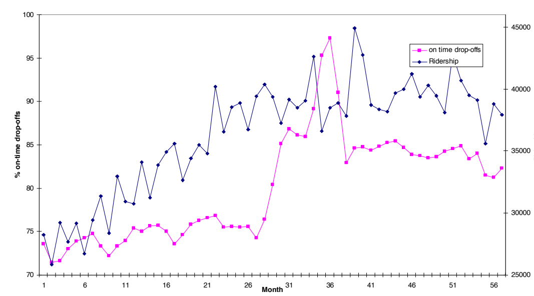
% on-time drop-offs and demand correlation



•Correlation coefficient is 0.138 with p-value of 0.307 for Laidlaw

21

Transit Team



•Correlation coefficient is 0.684 with p-value of 0 for Transit Team

•There is a difference of -1.550 between TT's and LL's mean on time drop-offs with p-value of 0.063

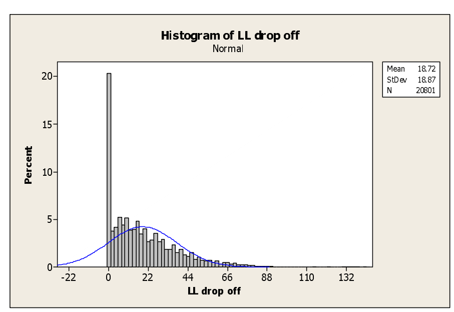
22

Actual Performance

Drop-off delay

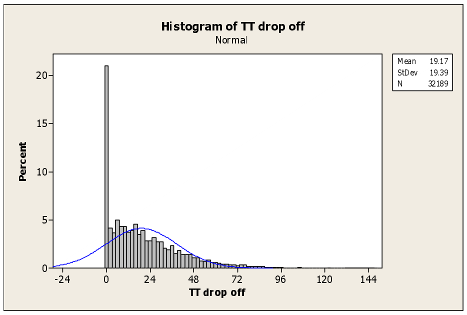
23

Laidlaw



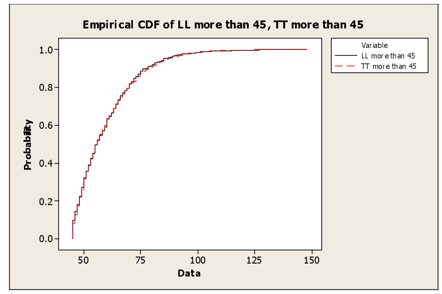
24

Transit Team



25

Cumulative Distribution Function of delays that are 45 minutes and longer



It is hard to tell which provider performs better by dropping off their passengers in a timely fashion.

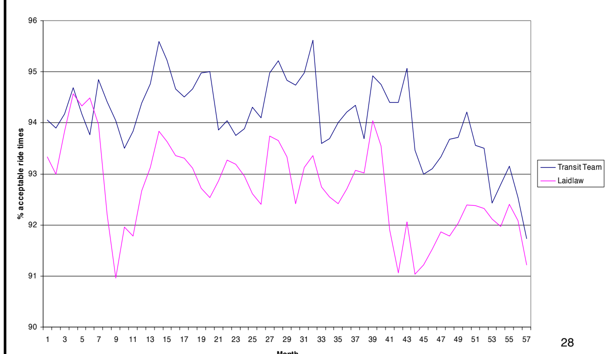
26

Service Quality Measure:

Percent acceptable ride times

27

Percent acceptable ride times



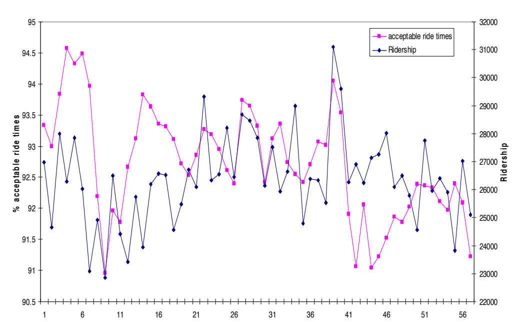
28

Percent acceptable ride times and demand correlation

Laidlaw and Transit Team

29

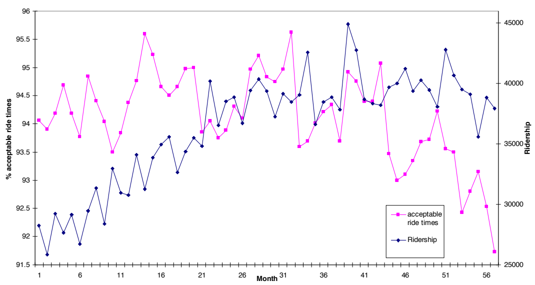
Laidlaw



•Correlation coefficient is 0.230 with p-value of 0.086 for Laidlaw

30

Transit Team



•Correlation coefficient is -0.170 with p-value of 0.206 for Transit Team

•There is a difference of 1.398% between TT's and LL's mean acceptable ride times with p-value of zero

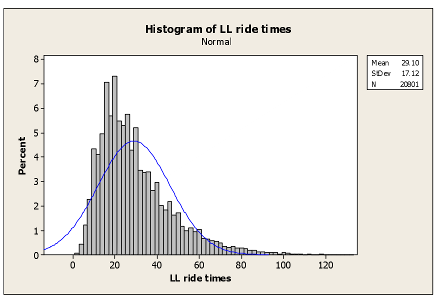
31

Service Quality Measure:

Ride times
more than 60 minutes

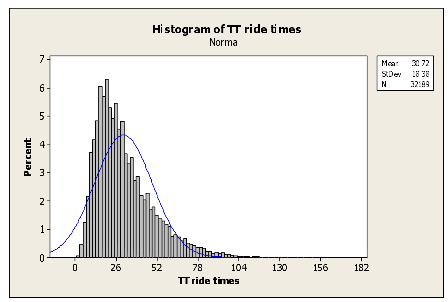
32

Laidlaw



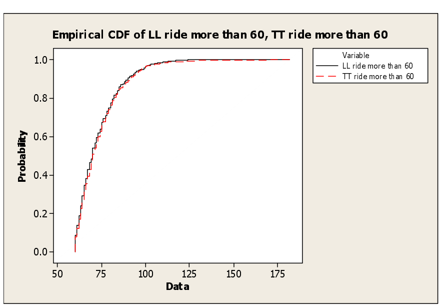
33

Transit Team



34

Cumulative Distribution Function of ride times that are 60 minutes and longer



•It is hard to see which provider has better performance in its ride time

•Looking at the histograms, mean delay of Laidlaw is smaller and Transit Team has a few longer ride times

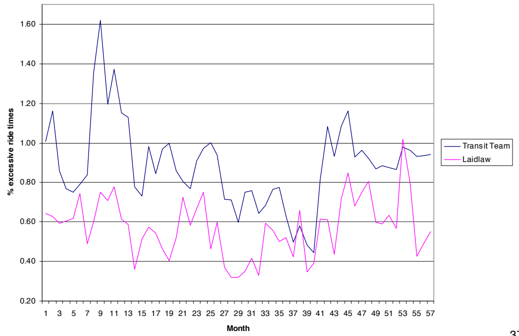
35

Service Quality measure:

Percent excessive ride times

36

Percent excessive ride times



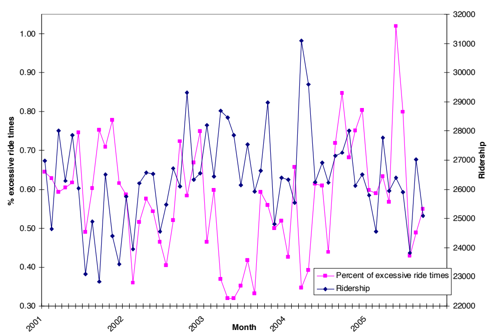
37

Percent excessive ride times and demand correlation

Laidlaw and Transit Team

38

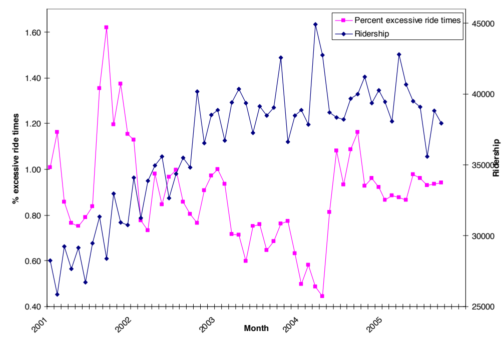
Laidlaw



•Correlation coefficient is -0.195 with p-value of 0.145

39

Transit Team



•Correlation coefficient is -0.350 with p-value of 0.008

•There is a difference of 0.315 between Transit Team's and Laidlaw's mean excessive ride times with p-value of zero

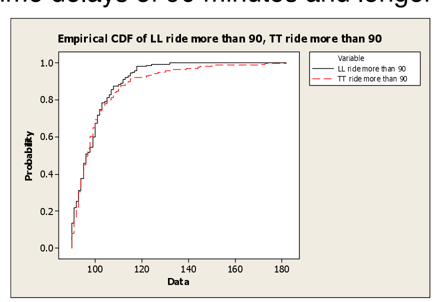
40

Actual Performance

Ride times more than 90 minutes

41

Cumulative Distribution Function for ride time delays of 90 minutes and longer



There is no clear stochastic dominance.

42

Appendix B

Mathematical Models for Chapter 3

As described in Section 3.2, our approach solves the reoptimization problem in two stages. In the first stage, we find the sequence of locations, or nodes, that each vehicle should visit in order to *minimize the total travel time* required to serve all passengers. In the second stage, we *minimize the total service time* (sum of travel time and vehicle idle time) using the sequence of locations determined in the first stage.

We use this two stage process for two reasons. First, lower fuel costs are incurred when a vehicle is idling versus when it is traveling, so reducing travel times is our primary objective. Second, this two-stage approach reduces the computational time needed to find solutions, which is necessary for the repeated application of the reoptimization approach. The two-stage procedure is applied first to individual routes and then to pairs of routes as described in Chapter 3.

B.1 Stage-1: Minimizing Travel Time

Let V be the set of dedicated vehicles such that for each vehicle $v \in V$, the maximum passenger capacity c_v is known. Let K be the set of passenger requests. For each passenger request $k \in K$, there is an associated demand of size n_k that needs to be transported from an origin, or pick up, node o_k to a destination, or drop off, node d_k. The demand size n_k is greater than 1 when the passenger has a personal care attendant traveling with him/her. Let $N = 2K + 1$ be the set of nodes that the vehicles must visit (origin and destination nodes and the depot). The direct travel time from node i to node j is indicated by t_{ij}. This travel time includes loading time at node i if node i is a pickup node and unloading time at node j if node j is a drop-off node.

ADA and MM dictate several requirements with respect to when passengers are picked up and dropped off. First, no passenger may be in a vehicle for more than s_1 minutes. Secondly, if passenger k was originally assigned a pick up time of e_k, then that passenger must be picked up between times e_k and $e_k + s_2$, where s_2 is the length of the allowable pick up time window. Finally, if passenger k specified an appointment time l_k by which they must arrive at their destination, then that passenger must be dropped off at their destination during the time window from $l_k - s_3$ to l_k. We use variable names s_1, s_2 and s_3 in our formulation because our approach can accommodate any value of these parameters. Actual parameter values used in the numerical study can be found in Section 3.1.

We use the following notational conventions in our model. Note $i = 0$ is the depot. Parameters $o(i, k)$ [respectively $d(i, k)$] equal 1, if node i is a pick up node [resp. drop off node] for passenger k, and 0 otherwise. M denotes a sufficiently large constant. It is used in the formulation to enforce certain constraints. In the numerical experiments reported in Section 3.3, M was set equal to 1,000.

We introduce one set of binary variables and two sets of continuous variables as decision variables for this model. They are $\{x_{ijv}\,(i, j \in N, v \in V)\}$, $\{a_i\,(i \in N)\}$, and $\{y_i\,(i \in N)\}$. Each x_{ijv} equals 1 if vehicle v travels directly from i to j and 0 otherwise. Similarly, a_i specifies the arrival time of a vehicle at node i and y_i specifies the number of passengers in a vehicle when it arrives at node i, before the vehicle picks up or drops off any passengers at that node. All vehicles start the shift at the depot, so $a_0 = 0$, and there are no passengers on board the vehicles at the depot, so $y_0 = 0$.

The objective function, given below, minimizes the total travel time of all vehicles.

$$\text{Minimize} \sum_{v=1}^{V} \sum_{j=0}^{N} \sum_{i=0}^{N} x_{ijv} t_{ij} \tag{1}$$

subject to the following constraints:

$$\sum_{j=1}^{N} x_{0jv} = 1 \qquad \text{for each } v \in V \tag{2}$$

$$\sum_{i=1}^{N} x_{i0v} = 1 \qquad \text{for each } v \in V \tag{3}$$

$$x_{iiv} = 0 \qquad \text{for each } i \in N \text{ and } v \in V \tag{4}$$

$$\sum_{v=1}^{V} \sum_{j=0}^{N} x_{ijv} = 1 \qquad \text{for } i = 1, \ldots, N \tag{5}$$

$$\sum_{v=1}^{V} \sum_{j=0}^{N} x_{jiv} = 1 \qquad \text{for } i = 1, \ldots, N \tag{6}$$

$$\sum_{j=0}^{N} x_{ijv} = \sum_{l=0}^{N} x_{liv} \qquad \text{for } i = 1, \ldots, N, \text{and } v = 1, \ldots, V \tag{7}$$

$$\sum_{i=1}^{N} x_{id_kv} = \sum_{j=1}^{N} x_{o_kjv} \qquad \text{for } v = 1, \ldots, V, \text{and } k = 1, ..., K. \tag{8}$$

Constraint (2-3) ensure that the trip for each vehicle starts and ends at the depot. Constraint (4) guarantees that the vehicle always leaves its current location. The requirement that each node is visited exactly once is imposed by constraints (5)-(6). Constraint (7) ensures vehicle flow balance; i.e., if a vehicle enters a node, then it must also exit that node. If the set V is a singleton, i.e., there is only one vehicle, constraint (7) can be omitted, as it is implied by (5)-(6). Constraint (8) ensures that the origin and destination of each passenger are served by the same vehicle. In addition, we need the following constraints to compute arrival times $a(i)$ and vehicle passenger loads $y(i)$.

$$y_i + \sum_{k=1}^{K} o(i,k)n_k \geq y_j - M(1 - \sum_{v=1}^{V} x_{ijv}) \qquad \text{for } i = 1,\ldots,N, j = 0,\ldots,N, \text{ when } i \text{ is origin node} \tag{9}$$

$$y_i + \sum_{k=1}^{K} o(i,k)n_k \leq y_j + M(1 - \sum_{v=1}^{V} x_{ijv}) \qquad \text{for } i = 1,\ldots,N, j = 0,\ldots,N \text{ when } i \text{ is origin node} \tag{10}$$

$$y_i - \sum_{k=1}^{K} d(i,k)n_k \geq y_j - M(1 - \sum_{v=1}^{V} x_{ijv}) \qquad \text{for } i = 1,\ldots,N, j = 0,\ldots,N \text{ when } i \text{ is destination node} \tag{11}$$

$$y_i - \sum_{k=1}^{K} d(i,k)n_k \leq y_j + M(1 - \sum_{v=1}^{V} x_{ijv}) \qquad \text{for } i = 1,\ldots,N, j = 0,\ldots,N \text{ when } i \text{ is destination node} \tag{12}$$

$$y_j \leq \sum_{v=1}^{V} \sum_{i=1}^{N} x_{ijv} c_v \qquad \text{for } j = 1,\ldots,N. \tag{13}$$

Constraints (9)–(12) ensure that the number of passengers are properly accounted for at the origin and the destination nodes. These constraints are only enforced for node pairs (i, j) when a vehicle travels directly from node i to node j ($x_{ij} = 1$).

If node i is a pickup node, then constraints (9)–(10) are included in the model. In this case, $\sum_{v=1}^{V} x_{ijv} = 1$, since each pick up node is visited by exactly one vehicle as implied in constraints (5)-(6). Therefore, the value of $(1 - \sum_{v=1}^{V} x_{ijv})$ is 0, and the right hand sides of constraints (9) and (10) become y_j. Now, constraints (9) and (10) together imply $y_i + \sum_{k=1}^{K} o(i,k)n_k = y_j$.

Similarly, when node i is a drop off node, constraints (11) and (12) are included in the model and together imply that the number of passengers on board the vehicle when it arrives at node j is the number on board when it arrived at the predecessor node i minus the number of passengers dropped off at node i.

The requirement that the number of passengers in the vehicle at each node do not exceed the capacity of the vehicle is guaranteed by constraint (13). Finally, we need the following constraints to ensure all variables take values in the correct ranges.

$$a_{d_k} \geq a_{o_k} \qquad \text{for each } k \in K \tag{14}$$

$$a_{d_k} \leq a_{o_k} + s_1 \qquad \text{for each } k \in K \tag{15}$$

$$e_k \leq a_{o_k} \qquad \text{for each } k \in K \tag{16}$$

$$a_{o_k} \leq e_k + s_2 \qquad \text{for each } k \in K \tag{17}$$

$$a_{d_k} \geq l_k - s_3 \qquad \text{for each } k \in K \tag{18}$$

$$a_{d_k} \leq l_k \qquad \text{for } k \in K \tag{19}$$

$$a_i + t_{ij} - M(1 - \sum_{v=1}^{V} x_{ijv}) \leq a_j \qquad \text{for } i = 1, \ldots, N, \text{and } j = 1, \ldots, N \tag{20}$$

$$x_{ijv} \in \{0, 1\} \qquad \text{for } i, j \in N, \text{and } v \in V \tag{21}$$

$$a_i \geq 0 \qquad \text{for } i \in N \tag{22}$$

$$y_i \geq 0 \qquad \text{for } i \in N. \tag{23}$$

Constraints (14) act as precedence constraints to ensure that each passenger's origin is visited before his destination. The requirement that the passenger can stay in the vehicle for no more than s_1 minutes is ensured by constraints (15). The pick up and drop off time window constraints are ensured by constraints (16)–(19). Constraints (18) and (19) are enforced only when a passenger specifies an appointment time (i.e. l_k exists). Constraints (20) guarantee that the arrival time is properly accounted for when a vehicle travels directly from node i to node j.

B.2 Stage-2: Minimizing Total Time

The Stage-1 model finds the sequence of nodes for each vehicle that minimizes the total travel time. In any operational plan, some vehicle idling may be necessary while the vehicle waits for the scheduled earliest pick up time of its next passenger. However, because the objective of the Stage-1 model does not consider idle time, its solution may include unnecessary idle time. Therefore, we developed a Stage-2 model to find the arrival times that minimize idle time (and, therefore, total time) for a given sequence of nodes.

This Stage-2 model is applied independently for each vehicle/route v. For a fixed vehicle v, $first_v$ and $last_v$ denote the first and last nodes visited by that vehicle. K_v is the set of passenger requests to be served by vehicle v, according to the Stage-1 solution. I_v is the set of nodes to be visited by vehicle v. It includes the depot and the origins and destinations of all passengers $k \in K_v$.

The Stage-2 model is:

$$\text{Minimize } a_{last_v} - a_{first_v}$$

subject to the constraints

$$
\begin{aligned}
e_k &\leq a_{o_k} && \text{for each } k \in K_v && (24)\\
a_{o_k} &\leq e_k + s_2 && \text{for each } k \in K_v && (25)\\
a_{d_k} &\geq l_k - s_3 && \text{for each } k \in K_v && (26)\\
a_{d_k} &\leq l_k && \text{for } k \in K_v && (27)\\
a_i + t_{ij} &\leq a_j && \text{for } i, j \in I_v \text{ such that } x_{ijv} = 1 && (28)\\
a_i &\geq 0 && \text{for } i \in I_v && (29)
\end{aligned}
$$

Note that constraints (24)–(29) are a subset of constraints (16)–(20) and (22) from the Stage-1 model. However, in Stage-2, the $x'_{ijv}s$ have been fixed by the Stage-1 solution, and so the only decision variables are the arrival times a_i.

By implementing this two stage process, we eliminate all unnecessary idle time while keeping computational effort manageable. More importantly, we obtain a solution that minimizes total travel time and subsequently the idle time. This approach makes sense from a practical viewpoint since MM incurs different costs for travel and idle times. In fact, in 2004, fuel accounted for 8% of MM's operational costs. It may be possible to construct routes with shorter total times than the routes output from our models by eliminating the two-stage process. However, the resulting model would be significantly more complex and require excessive time to solve.

We developed these models so they can be used in several ways. The raw passenger request data for a single day could be passed to the Stage-1 model, V would represent the complete set of available vehicles. The model would then perform two tasks: assigning passengers to vehicles and ordering the nodes visited by each vehicle. However, given the size of the data (approximately 100 routes per provider and 1,000 passengers per day), that approach is computationally intractable.

An alternative is to decompose the problem by using the assignments of passengers to routes determined by Trapeze. The Stage-1 model can then be solved sequentially for each route, or for small subsets of routes. Although some flexibility is lost by pre-assigning passengers to routes, the gain in computational efficiency makes this a practical alternative.

Appendix C

Literature Review

The Dial-a-Ride Problem (DARP) has been the focal point of demand-responsive paratransit service research since it was first introduced. Wilson et al. (1971), Wilson and Weissberg (1976), and Wilson and Colvin (1977) were among the first to study the DARP with specific interest in developing real-time algorithms for paratransit systems. Due to high operating costs, most of the dial-a-ride systems turned into reservation-based operations after the late 1970s. Research on DARP shifted to developing heuristic algorithms for solving the static DARP. Cordeau and Laporte (2003) present a tabu search heuristic for the static multi-vehicle DARP. Recent surveys on the DARP and, more generally, the vehicle routing and scheduling problems are presented by Bodin et al. (1983), Desrosiers et al. (1993), and Savelsbergh and Sol (1995). Cordeau (2006) presents a branch-and-cut algorithm for the DARP by introducing new valid inequalities, which can solve small to medium-size instances. Another formulation of DARP is presented by Fu (2002) that incorporates time-dependent and stochastic travel times.

MM's problem can be viewed as the dynamic DARP. However, the focus of this project was to improve upon the Trapeze solution by reoptimizing routes. Therefore it was appropriate for us to consider the static Vehicle Routing Problem with Time Windows (VRPTW) and precedence constraints. Surveys of the VRPTW have been presented in Solomon and Desrosiers (1988) and Desrochers and Soumis (1988).

The VRPTW can be formulated in different ways. The first is the Pickup and Delivery Problem (PDP), where vehicles must transport loads from origins to destinations without any transshipment at other locations. The paper by Savelsbergh and Sol (1995) presents a survey of the PDP types, such as the Static PDP, the static 1-PDP with time windows, etc., and solution methods found in the literature. A generalization of the PDP with time windows is called the Handicapped persons' Transportation Problem (HTP). Toth and Vigo (1997) present a fast and effective parallel insertion heuristic algorithm for the HTP that can determine good solutions for real-world problem instances. They also present a Tabu thresholding procedure that can be used to improve the solution obtained by the insertion algorithm.

The VRPTW can be formulated as a Traveling Salesman Problem (TSP), if there is no origin or destination and no time windows or precedence constraint. Renaud et al. (1996) propose a fast composite heuristic that consists of three phases: construction of an initial tour, insertion of the remaining vertices, and the improvement procedure for the symmetric TSP. A variation of TSP is the TSP with time windows (TSPTW). Gendreau et al. (1998) present a generalized insertion heuristic for the TSPTW with the objective of minimizing the travel times. The heuristic constructs a route by inserting a vertex in its neighborhood (based on distance) while performing a local optimization of the tour. This method is known as GENI, which is proposed by Gendreau et al. (1992). Desrosiers et al. (1988) propose a Lagrangian relaxation method for solving the minimum fleet size multiple TSPTW, and this proposed method is able to optimally solve practical school bus scheduling problems. The next case is the pickup and delivery TSP (PDTSP). Renaud et al. (2000) present a proposed heuristic for the PDTSP that is composed of two phases: a construction phase that includes a local optimization and a deletion and re-insertion improvement phase.

The VRPTW itself is NP-hard (see Derigs and Grabenbauer (1993) and Cordeau et al. (2001)), which means that the time it takes to find an exact solution to the problem increases exponentially as the size of the problem grows. That is why many researchers use a mathematical approximation for small to medium instances, and heuristics for large instances.

The work by Desrochers et al. (1992) proposes an LP relaxation of the set partitioning formulation of the VRPTW, which is then solved by column generation. The authors assume a homogeneous fleet, but the fleet size is not fixed a priori. This algorithm has solved problems with up to 100 customers from the Solomon data set (Solomon (1987)), including some problems that other methods were not able to solve. Fisher et al. (1997) describe two methods, which are able to solve problems of the same size to optimality.

The two methods proposed by Fisher et al. (1997) are Lagrangian relaxation or variable splitting approach and a K-tree relaxation method. Their formulation is an extension of the formulation given by Solomon (1987), which we use in developing our formulation. Kohl and Madsen (1997) propose a similar method, which is based on a Lagrangian relaxation of the constraint set that requires each customer to be serviced.

The work by Desrochers et al. (1992) is believed to be the best known algorithm for solving the VRPTW, and it has solved problems with up to 100 customers. The papers mentioned above find the set of routes to satisfy the transportation requests. However, in our situation, we already have the set of routes, and we want to reoptimize these routes to improve upon them. Thus, we decompose the VRPTW to reoptimize one vehicle/route at a time.

For our second approach, we propose to reoptimize two routes at the same time, and since we are dealing with much larger number of passengers, we decide to use heuristics to avoid computational burden. Heuristics start with an initialization or construction phase followed by the improvement phase. There are many papers that focused on solving VRPTWs with heuristics, we present a brief review of this literature below.

The paper by Derigs and Grabenbauer (1993) is motivated by a real world application of distribution planning for a bakery in Bayreuth, Germany. The authors implement a two-phase heuristic procedure. The first phase is the tour construction phase, which focuses on the problem of choosing customers to be added to the tour. The tour improvement phase is done by exchanging the nodes, known as k-best-insert-algorithm. The node-exchange procedure is done in two phases, deterministic exchange phase followed by a stochastic exchange phase with each phase having a predetermined number of iterations.

The paper by Thangiah et al. (1993) proposes three heuristics: deadline sweep, push-forward insertion, and genetic sectoring. The solutions obtained from these heuristics are further improved using a local optimization procedure. The Deadline Sweep heuristic (DSH) clusters customers and routes the vehicles within the clusters and is time oriented. The Push-Forward Insertion heuristic (PFIH) is an insertion heuristic where customers are inserted into a current route until one of the constraints is violated. The Genetic Sectoring Heuristic (GSH) is a cluster-first-route-second heuristic that uses a Genetic Algorithm (GA) to cluster the customers and uses cheapest insertion procedure to route the vehicles. When a solution is obtained, a local post-optimization procedure is used to improve the solution by shifting and exchanging customers between the routes.

Cordeau et al. (2001) present two generalizations of VRPTW: the Periodic Vehicle Routing Problem with Time Windows (PVRPTW) and the Multi-Depot Vehicle Routing Problem with Time Windows (MDVRPTW). The PVRPTW considers a planning horizon of t days and each customer specifies a service frequency and a set of allowable combination of visit days. The problem selects an allowable combination of visit days for each customer and generates vehicle routes for each day of the planning horizon at the same time.

The MDVRPTW considers several depots at different locations instead of only one location. The problem is to assign each customer to a depot and construct routes for each depot simultaneously using VRPTW rules. Furthermore, MDVRPTW can be seen as a special case of PVRPTW by associating depots with days of the planning horizon. The same methodology can then be used for both problems provided that the distances and travel times that vary day to day reflect the different locations of the depots. The algorithm that they use is called the tabu search algorithm. It is a local search heuristic that explores the solution space by moving at each iteration from the current solution s to the best solution in its neighborhood $N(s)$. The post optimization procedure used is the TSPTW heuristic developed by Gendreau et al. (1998).

In our situation, we only perform the improvement phase. We do not need to construct the route because we have the reoptimized routes from the first approach. In this approach, we insert each passenger (pick up and drop off nodes of each passenger) of one route to another route and vice versa, and we also allow exchange of passengers, one pair at a time.