UTCM

Improving the Quality of Life by Enhancing Mobility

University Transportation Center for Mobility™

DOT Grant No. DTRT06-G-0044

Multi-vehicle Mobility Allowance Shuttle Transit (MAST) System: An Analytical Model to Select the Fleet Size and a Scheduling Heuristic

Final Report

Luca Quadrifoglio and Wei Lu

Performing Organization
University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System
College Station, TX

Sponsoring Agency
Department of Transportation
Research and Innovative Technology Administration
Washington, DC

Texas Transportation Institute

UTCM Project #11-47-77
June 2012

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UTCM 11-47-77		

4. Title and Subtitle
Multi-vehicle Mobility Allowance Shuttle Transit (MAST) System - An Analytical Model to Select the Fleet Size and a Scheduling Heuristic

5. Report Date
June 2012

6. Performing Organization Code
Texas Transportation Institute

7. Author(s)
Luca Quadrifoglio and Wei Lu

8. Performing Organization Report No.
UTCM 11-47-77

9. Performing Organization Name and Address
University Transportation Center for Mobility™
Texas Transportation Institute
The Texas A&M University System
3135 TAMU
College Station, TX 77843-3135

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTRT06-G-0044

12. Sponsoring Agency Name and Address
Department of Transportation
Research and Innovative Technology Administration
400 7th Street, SW
Washington, DC 20590

13. Type of Report and Period Covered
Final Report
May 2011 – May 2012

14. Sponsoring Agency Code

15. Supplementary Notes
Supported by a grant from the US Department of Transportation, University Transportation Centers Program

16. Abstract

The mobility allowance shuttle transit (MAST) system is a hybrid transit system in which vehicles are allowed to deviate from a fixed route to serve flexible demand. A mixed integer programming (MIP) formulation for the static scheduling problem of a multi-vehicle Mobility Allowance Shuttle Transit (MAST) system is proposed in this thesis. Based on the MIP formulation, we analyzed the impacts of time headways between consecutive transit vehicles on the performance of a two-vehicle MAST system. An analytical framework is then developed to model the performance of both one-vehicle and two-vehicle MAST systems, which is used to identify the critical demand level at which an increase of the fleet size from one to two vehicles would be appropriate. Finally, a sensitivity analysis is conducted to find out the impact of a key modeling parameter, w_1, the weight of operations cost on the critical demand.

In this research, we developed an insertion heuristic for a multi-vehicle MAST system, which has never been addressed in the literature. The proposed heuristic is validated and evaluated by a set of simulations performed at different demand levels and with different control parameters. By comparing its performance versus the optimal solutions, the effectiveness of the heuristic is confirmed. Compared to its single-vehicle counterpart, the multiple-vehicle MAST prevails in terms of rejection rate, passenger waiting time and overall objective function, among other performance indices.

17. Key Word	18. Distribution Statement
Hybrid transit; Scheduling; Mixed-integer programming; Heuristics	Public distribution

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	n/a

Multi-vehicle Mobility Allowance Shuttle Transit (MAST) System - An Analytical Model to Select the Fleet Size and a Scheduling Heuristic

by
Luca Quadrifoglio, Ph.D.
Zachry Department of Civil Engineering
Texas A&M Univeristy

and

Wei Lu
Zachry Department of Civil Engineering
Texas A&M University

Final Report UTCM 11-47-77

University Transportation Center for Mobility^{TM}
Texas Transportation Institute
The Texas A&M University System
3135 TAMU
College Station, TX 77843-3135

June 2012

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

Support for this research was provided in part by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the University Transportation Center for Mobility^{TM} (DTRT06-G-0044) at the Texas Transportation Institute and the Texas Department of Transportation.

Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

Transit service is an inalienable and important part of transportation systems. Nowadays, transit agencies are facing challenges due to urban sprawl. Urban sprawl is a phenomenon that as the population exploding, the population density of cities is actually dropping. As urban sprawl occurs, cities begin losing their traditional centralized functions, and people spread into the suburbs to seek more space and a lower cost of living. As a result, the urban areas are becoming more and more car-dependent, and it appears that new highway construction may not satisfy the demand. We have a few archetypes of urban sprawl such as the Los Angeles metro area, Houston, and Atlanta, among others. In these areas, traditional transit services are struggling because they are not able to provide a satisfying service in such a low-density context. To face the challenges of urban sprawl, on one hand, urban planning agencies are proposing policies to regulate unlimited urban sprawl. On the other hand, transit agencies are actively seeking innovative public transit solutions that are attractive enough to serve people in the low-density urban areas.

In terms of their flexibility, public transit services can be divided into two broad categories: fixed-route transit and a more flexible option called the demand-responsive transit. The fixed-route transit systems include the common buses, subway systems, and school shuttles, etc. They are considered to be cost-efficient because of their ride-sharing attribute and sufficiently large loading capacity. Fixed-route transit works well in traditional cities that have a high density, but it is considered to be inconvenient since the fixed stops and schedule cannot meet individual passenger's needs. This lack of flexibility is the most significant constraint of fixed-route transit and prevents it from being effective when used in the urban sprawl context. On the other hand, the demand-responsive transit (DRT) systems are much more flexible due to their door-to-door pickup and drop-off services. DRT has been operated in numerous cities and works as an effective type of flexible transit service especially within low-density residential areas.

Since both fixed-route transit and demand-responsive transit have their advantages and disadvantages, a possible improvement is to be eclectic. The idea is combining the cost-efficient operability of traditional fixed-route transit with the flexibility of demand-responsive systems. This new concept is called the mobility allowance shuttle transit (MAST). It is a hybrid transit system in which vehicles are allowed to deviate from a fixed route to serve a flexible demand.

In the first part of the research, the researchers investigated the complexity of the scheduling problem for multiple-vehicle MAST and gave the formal proof for its computational intractability. A mathematical formulation to solve the problem exactly is proposed. Based on the formulation, the research team analyzed the impacts of time headways between consecutive transit vehicles on the performance of a two-vehicle MAST system. An analytical framework is then developed to model the performance of both one-vehicle and two-vehicle MAST systems. The framework is then used to identify the critical demand level at which an increase of the fleet size from one to two vehicles would be appropriate. After that, researchers conducted a sensitivity analysis to find out the impact of a key modeling parameter, w_1, the weight of operations cost on the critical demand.

As shown in the first part, the problem is NP-hard, meaning the time needed to solve the problem grows exponentially as the problem size increases. As a result, the researchers developed an insertion heuristic (an approximation algorithm) for a multi-vehicle MAST system, which has never been addressed in the literature. The proposed heuristic is validated and evaluated by a set of simulations performed at different demand levels and with different control parameters. By comparing its performance versus the optimal solutions, the effectiveness of the heuristic is confirmed. Compared to its single-vehicle counterpart, the multiple-vehicle MAST prevails in terms of rejection rate, passenger waiting time and overall objective function, among other performance indices.

CHAPTER 1 INTRODUCTION

1.1 Background

Public transit systems are gaining more concerns due to urban sprawl and the heavy traffic congestion in urban areas. Transit systems are more cost-efficient than personal vehicles. Thus, with the economic crisis and the increase of fuel prices, transit systems are a better choice for the public. However, the financial support for the whole transportation system has decreased, so it's critical to find a more cost-efficient transit type.

Public transit services are divided into two broad categories: fixed-route transit (FRT) and demand responsive transit (DRT). The FRT systems are thought to be cost-efficient because of their ride-sharing attribute and sufficient loading capacity, but they are considered by the general public to be inconvenient since the fixed stops and schedule are not always convenient. This inherent lack of flexibility is the most significant constraint of fixed-route transit. The DRT systems are much more flexible since they offer door-to-door pickup and drop-off services. They operate in numerous cities and work as an effective type of flexible transit service, especially within low-density residential areas such as examples in Denver (CO), Raleigh (NC), Akron (OH), Tacoma (WA), Sarasota (FL), Portland (OR) and Winnipeg (Canada) [1]. However, the associated high cost prevents the DRT from being deployed as a general transit service. As a result they are largely limited to specialized operations such as shuttle service, cab and Dial-a-Ride services, which are mandated under the Americans with Disabilities Act. Thus, transit agencies are faced with increasing demand for improved and extended DRT service. Thus, a combination of these two types of transit systems is needed to provide a relatively cost-efficient and flexible transit type.

The mobility allowance shuttle transit (MAST) is an innovative concept that combines the cost-efficient operability of traditional FRT with the flexibility of DRT systems. It allows transit vehicles to deviate from a fixed route consisting of a few mandatory checkpoints to serve on-demand customers within a predetermined service area, and thus can be both affordable and convenient enough to attract the general public. For the MAST system, the fixed route can be either a loop or a line between two terminals. The checkpoints are usually located at major transfer stops or high demand zones and are relatively far from each other. A hard constraint of the MAST system is the scheduled departure time from checkpoints. Such a service already exists in Los Angeles County with MTA Line 646 serving as a nighttime bus line transporting mostly night-shift employees of local firms. They developed the insertion heuristic scheduling of a single vehicle MAST system [2], but an advanced system can be performed with multiple vehicles, and the scheduling problem will be more complicated.

1.2 Literature Review

The design and operations of the MAST system have attracted considerable attention in recent years. Quadrifoglio et al. [3] evaluated the performance of MAST systems in terms of serving capability and longitudinal velocity. Their results indicate that some basic parameters are helpful in designing the MAST system such as slack time and headway. Quadrifoglio et al. later developed an insertion heuristic scheduling algorithm to address a large amount of demand dynamically[2]. Quadrifoglio and Dessouky [4] carried out a set of simulations to show the sensitivity analysis for the performance of the insertion heuristic algorithm and the capability of the system over different shapes of service area. In 2008, Zhao and Dessouky[5] studied the optimal service capacity for the MAST system. Although these studies investigated the design and operations of the MAST system from various aspects, they are all for the single-vehicle MAST system.

Since the MAST system is a special case of the pickup and delivery problem (PDP, see [6] for a complete review), it can be modeled as a mixed integer program (MIP). The PDP has been extensively studied, and many of the exact algorithms are based on integer programming techniques. Sexton and Bodin [7] reported a formulation and an exact algorithm using Bender's decomposition. Cordeau introduced an MIP formulation of the multi-vehicle Dial-a-Ride Problem (DARP) [8], which is a variant of PDP. He proposed a branch-and-cut algorithm using new valid inequalities for DARP. This multi-vehicle DARP MIP formulation is a good reference for the multi-vehicle MAST MIP formulation. Cordeau and Laporte gave a comprehensive review on PDP, in which different mathematical formulations and solution approaches were examined and compared [9]. Lu and Dessouky [10] formulated the multi-vehicle PDP as an MIP and developed an exact branch-and-cut algorithm using new valid inequalities to optimally solve multi-vehicle PDP of up to 5 vehicles and 17 customers without clusters and 5 vehicles and 25 customers with clusters within a reasonable time. In [11], Cortes et al. proposed an MIP formulation for the PDP with transfers. Very recently, Ropke and Cordeau [12] combined the techniques of row generation and column generation and proposed a branch-cut-and-price algorithm to solve PDP with time windows (PDPTW). In their algorithm, the lower bounds are computed by solving the linear relaxation of a set partitioning problem through column generation, and the pricing subproblems are the shortest path problems. Berbeglia et al. reviewed the most recent literature on dynamic PDPs and provided a general framework for dynamic one-to-one PDPs [13]. Quadrifoglio et al. proposed an MIP formulation for the static scheduling problem of a single-vehicle MAST system and solved the problem by strengthening the formulation with logic cuts [14]. Other exact algorithms include dynamic programming. Psaraftis used dynamic programming to solve the single-vehicle DARP [15] and its variant with time windows [16]. Both algorithms have a time complexity of $O(N^2 3^N)$ (N for customers) and can solve an instance of N up to 20 in a meaningful time. Very recently, Fortini et al. [17] proposed a new heuristic for TSP based on computing compatible tours instead of TSP tours. They proved that the best compatible tour has a worst-case cost ratio of $5/3$ to that of the optimal TSP tour. A branch-and-cut algorithm was developed to compute the best compatible tour, and Teodorovic and Radivojevic developed a fuzzy logic approach for the DAR problem [18].

Since the optimization problem of PDP is known to be strongly $\mathcal{NP}$-hard [19], researchers have been studying heuristic approaches to solve PDP with large instances in a reasonable (polynomial) time, while not compromising the quality of solution too much. Along these approaches, insertion heuristics are the most popular because they can quickly provide meaningfully good results and are capable of handling problems with large instances. Another reason that justifies insertion heuristics in practice is that they can be easily implemented in dynamic environments [20]. Some other efforts in insertion heuristics include research by Lu and Dessouky's [21]. A major disadvantage of the insertion heuristics is that usually it's hard to bound its performance. Another disadvantage is its myopic and greedy approach for current optimum at each time step without having an overview of all the requests. The insertion heuristic controlled by "usable slack time" resolved this issue efficiently [2]. To evaluate the performance of the proposed heuristics, worst-case analysis can be found for PDP and its fundamental or related problems such as traveling salesman problem (TSP) and vehicle routing problem (VRP). Savelsbergh and Sol [6] gave a complete review on the pickup and delivery problem and discussed several variants of the problem in terms of different optimization objectives, time-constraints and fleet sizes. Both exact algorithms based on mathematical modeling and heuristics were reviewed. Christofides [22] proposed a new heuristic of ratio $3/2$ for metric-TSP based on constructing minimum spanning tree and Euler tour. Rosenkrantz et al. [23] analyzed the approximation ratio of several heuristics, including the cheapest insertion heuristic for TSP. Archetti et al. [24] studied the re-optimization version of TSP, which arises when a new node is added to an optimal solution or when a node is removed. They proved that although the cheapest insertion heuristic has a tight worst-case ratio of 2 [23], the ratio decreases to $3/2$ when applied to the re-optimization TSP problem. So far the best results on TSP is Arora's polynomial time approximation scheme for Euclidean TSP [25].

1.3 Report Overview

This report consists of five chapters. In Chapter 2, the optimization problem of scheduling multiple-vehicle MAST (m-MAST) is formally defined and a $\mathcal{NP}$-hardness proof through reduction from m-PDP is given. The problem is modeled as a mixed-integer program (MIP), and the model is explained in detail.

In Chapter 3, the researchers provide an analytical modeling framework to help MAST operators with their system planning and to identify the critical transit demand, which is used to decide when to switch from the one-vehicle MAST system (1-MAST) to two-vehicle MAST (2-MAST) system. A series of experiments are conducted to verify the analytical model by comparing its results with those obtained by solving the MIP. The utility function values generated by the one-vehicle MAST system and the two-vehicle MAST system are compared. A sensitivity analysis is then conducted to find out the impact of a key modeling parameter on the critical demand.

Chapter 4 first develops an insertion heuristic for scheduling m-MAST based on the previous work of Quadrifoglio et al. [2]. The core idea is reserving the crucial resource of slack time used by each insertion for future use, thus resolving the inherit "myopia" of insertion-based heuristics. Due to the existence of complicated time constraints and weighted objective function, the researchers resort to experiments to evaluate the algorithm. Three series of experiments are conducted: 1. control parameter tuning, 2. comparing the performance of 1-MAST and 2-MAST and 3. comparing heuristic results with optimal solution obtained by solving MIP.

Chapter 5 summarizes the findings and contributions of this research. Concluding remarks on future research are also provided.

CHAPTER 2 M-MAST PROBLEM AND MIP FORMULATION

2.1 Problem Description

The multi-vehicle MAST system considered consists of a set of vehicles with predefined schedules along a fixed-route of C checkpoints ($i = 1, 2, \ldots, C$). These checkpoints include two terminals ($i = 1$ and $i = C$) and the remaining $C-2$ intermediate checkpoints. A rectangular service area is considered in this study as shown in Fig. 2.1 [14], where L is the distance between the two terminals, and W/2 is the maximum allowable deviation distance on each side of the fixed-route. Vehicles perform R trips back and forth between the terminals.

In this study, the transit demand is defined by a set of requests. Each request consists of pickup/drop-off locations and a ready time for pickup. There are four possible types of customer requests, which are shown below:

- PD (Regular): pickup and dropoff at a checkpoint
- PND (Hybrid): pickup at a checkpoint and dropoff at a random point
- NPD (Hybrid): pickup at a random point and dropoff at a checkpoint
- NPND (Random): pickup and dropoff at random points

Vehicles need to deviate from the fixed route defined by checkpoints to serve PND, NPD and NPND customers at their non-checkpoints (pickup stops and drop-off stops, see Fig. 2.2), while conforming to the time constraints associated with the checkpoints.

We consider the following two assumptions in formulating the multi-vehicle MAST problem: 1) the scenario is static and deterministic where the transit demand is known in advance; and 2) each

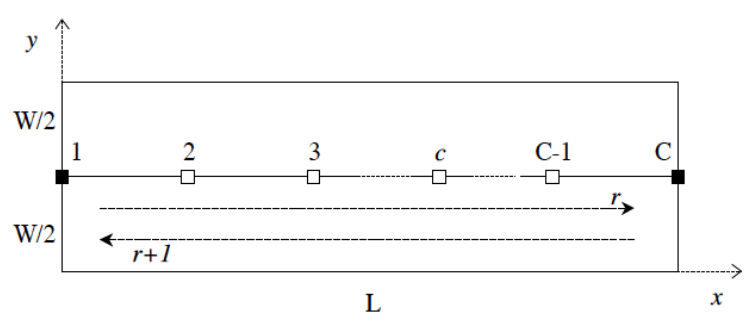

Figure 2.1: MAST system (Quadrifoglio et al., 2008)

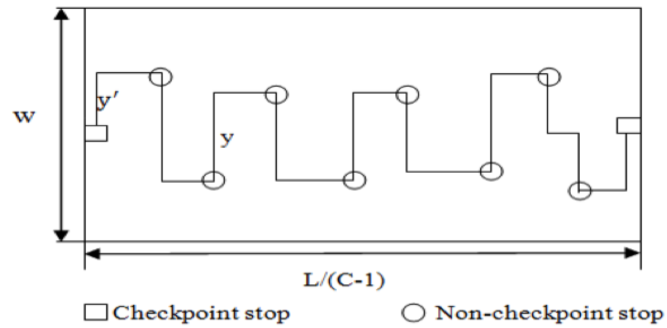

Figure 2.2: Illustration for bus route visiting checkpoints and non-checkpoints

request only has one customer and there is no capacity constraint for transit vehicles. The following presents the notations for the multi-vehicle MAST system:

Sets of Requests:

- $K_{PD}/K_{PND}/K_{NPD}/K_{NPND}$ = set of PD/PND/NPD/NPND requests
- $K_{HYB} = K_{PND} \cup K_{NPD}$ = set of hybrid requests (PND and NPD types)
- $K = K_{PD} \cup K_{HYB} \cup K_{NPND}$ = set of all requests
- $ps(k) \in N$ = pickup of k, $\forall k \in K \backslash K_{PND}$
- $ds(k) \in N$ = drop-off of k, $\forall k \in K \backslash K_{NPD}$
- $pc(k,r,v) \in N_0$ = collections of all the occurrences in the schedule (for each $r \in RD$ and each $v \in V$) of the pickup checkpoint of k, $\forall k \in K_{PND}$
- $dc(k,r,v) \in N_0$ = collections of all the occurrences in the schedule (for each $r \in RD$ and each $v \in V$) of the drop-off checkpoint of k, $\forall k \in K_{NPD}$

Sets of Nodes:

- N_0 = checkpoints
- N_n = non-checkpoint stops
- $N = N_0 \cup N_n$

Sets of Arcs:

- A = all arcs

Sets of Trips:

- $RD = \{1, ..., R\}$ = set of trips
- $HYBR(k) \subset RD$ = feasible trips of k, $\forall k \in K_{HYB}$

To introduce the integer programming model of m-MAST, we need to further define some variables and parameters.

Parameters:

- $R =$ number of trips
- $C =$ number of checkpoints
- $V_e =$ number of vehicles
- $V =$ set of vehicles
- $TC = [(C-1) \times R + 1] \times V_e =$ total number of stops at checkpoints in the schedule
- $TC_0 = (C-1) \times R + 1 =$ number of checkpoint stops of one vehicle
- $TS = TC + |K_{PND}| + |K_{NPD}| + 2 \times |K_{NPND}| =$ total number of stops
- $\theta_i =$ scheduled departure time of checkpoint stop i, $\forall i \in N_0, (\theta_1 = 0)$
- $\tau_k =$ ready time of request k, $\forall k \in K$
- $\delta_{i,j} =$ rectilinear travel time between i and j, $\forall i, j \in N$
- $b_i =$ service time for boarding and disembarking at stop i
- $w_1/w_2/w_3 =$ objective function weights

Variables:

- $x^v_{i,j} = \{0,1\}, \forall (i,j) \in A, \forall v \in V =$ binary variables indicating if an arc (i,j) is used by vehicle v $(x^v_{i,j} = 1)$ or not $(x^v_{i,j} = 0)$
- $t_i =$ departure time from stop i, $\forall i \in N$
- $\bar{t}_i =$ arrival time at stop i, $\forall i \in N \backslash \{1\}$
- $p_k =$ pickup time of request k, $\forall k \in K$
- $d_k =$ drop-off time of request k, $\forall k \in K$
- $z^v_{k,r} = \{0,1\}, \forall k \in K_{HYB} =$ binary variable indicating whether the checkpoint stop of the hybrid request k (a pick-up if $k \in K_{PND}$ or a drop-off if $k \in K_{NPD}$) is scheduled in trip r of vehicle v, $\forall r \in RD, \forall v \in V$

To give an example, we consider $k = 4$ customers (see Table 2.1) with their corresponding pickup and drop-off stops according to the network in Fig. 2.3, describing a simple single-vehicle MAST (1-MAST) system with $TC_0 = 3$ checkpoints in $N_0 = \{1^0, 2^0, 3^0\}$, two pickup stops in $N_{n^+} = \{4^+, 5^+\}$ and two drop-off requests in $N_{n^-} = \{6^-, 7^-\}$.

Table 2.1: Sample Set of Requests

k	$ps(k)$	$ds(k)$
1	4^+	6^-
2	1^0	7^-
3	5^+	2^0
4	1^0	3^0

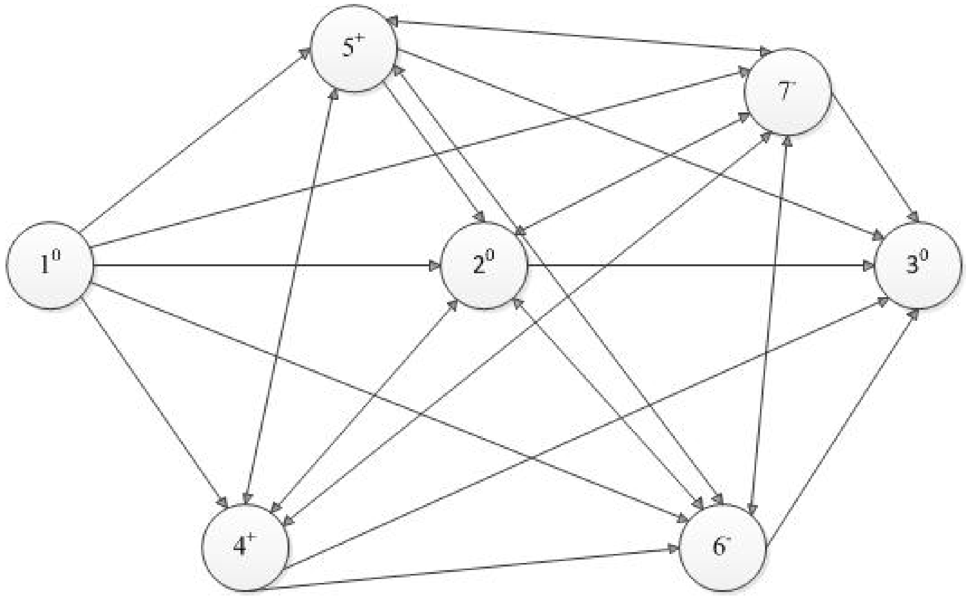

Figure 2.3: Sample 1-MAST network

The network is almost a complete graph, excluding the arcs violating the conditions described above, namely $(2,1)$, $(3,2)$, $(3,1)$ and $(1,3)$ that violate the predetermined sequence of checkpoints $(1 \to 2 \to 3)$ and $(6,4)$, $(2,5)$, $(7,1)$ that violate the pickup before drop-off precedence for each request. In addition, since checkpoints 1 and 3 represent the beginning and the end of the service, there are no arcs to 1 and no arcs from 3.

When another vehicle is added into the system, the network becomes more complicated, the arcs nearly doubled, and even the request set remains the same (see Fig. 2.4). Note that the solid arcs are legal arcs for vehicle 1, while dash arcs are for vehicle 2. The nodes 1^0 & 4^0 (so as 2^0 & 5^0 and 3^0 & 6^0) are essentially the same node geographically, but in the perspective of scheduling they're not, since they are visited by different vehicles at different times. For the graph, the same aforementioned precedence and time constraints still apply in an m-MAST system. No arcs between nodes representing checkpoints visited by different vehicles (such as $(1 \to 4)$ and $(1 \to 5)$) are allowed.

To formally introduce the multiple-vehicle MAST problem, we first present some definitions.

Definition 1 (m-MAST route)**.** *An m-MAST route Rt_v for vehicle v is a directed route through a subset $N_v \subset N$ such that:*

1. *Rt_v starts in* $1+(v-1)\times TC_0$
2. $\{1+(v-1)\times TC_0,...,v\times TC_0\} \subset N_v$
3. $(ps(k)\cup ds(k))\cap N_v = \varnothing$ *or* $(ps(k)\cup ds(k))\cap N_v = ps(k)\cup ds(k)$ *for all* $k \in K$
4. *If* $ps(k)\cup ds(k) \subseteq N_v$*, then* $ps(k)$ *is visited before* $ds(k)$
5. *Vehicle v visits each location in N_v exactly once.*

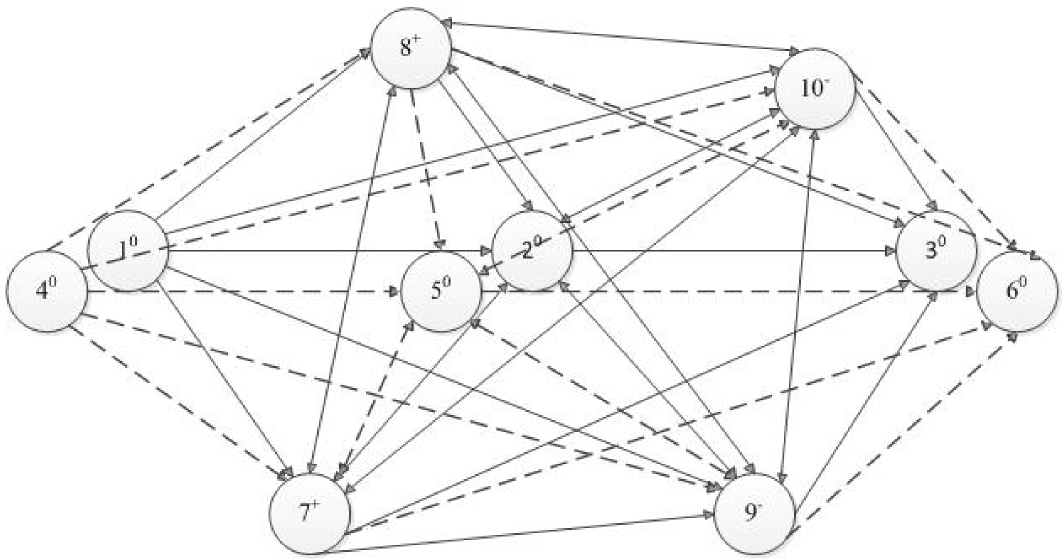

Figure 2.4: Sample 2-MAST network

6. *Precedence constraint of pickup and drop-off is not violated.*

7. *Departure times at checkpoints* $\{1+(v-1)\times TC_0,...,v\times TC_0\}$ *are complied with*

8. Rt_v *ends in* $v\times TC_0$

Definition 2 (m-MAST plan)**.** *An m-MAST plan is a set of routes* $\mathcal{RT}=\{Rt_v|v\in V\}$ *such that:*

1. Rt_v *is an m-MAST route for vehicle v, for each* $v\in V$*.*

2. $\{N_v|v\in V\}$ *is a partition of N.*

Define $f(\mathcal{RT})$ as the price of plan $\mathcal{RT}$ corresponding to a certain objective function f. Then we define the m-MAST problem as:

$$min\{f(\mathcal{RT}|\mathcal{RT} \text{ is an m-MAST plan})\}$$

Particularly in this paper f is a combination of operation cost and dissatisfaction of customers, defined by:

$$w_1\times M/v+w_2\times RT\times|K|+w_3\times WT\times|K|$$

where w_1, w_2 and w_3 are the weights, and M represents the total miles driven by the vehicles, RT is the average ride time per customer and WT the average waiting time per customer from the ready time to the pick-up time. This definition of the f allows optimizing in terms of both the vehicle variable cost (first term) and the service level (the last two terms); modifying the weights accordingly, we can emphasize one factor over the others as needed.

Definition 3 (m-MAST problem)**.** *An optimization problem m-MAST is a 4-tuple* $<I_Q,S_Q,f_Q,opt_Q>$*, where:*

- *I_Q: the set of all MAST graphs G*
- *S_Q: the set of all m-MAST plans of the graph G*
- *f_Q: $f(\mathcal{RT})$ is the price of m-MAST plan $\mathcal{RT}$ of G*
- *opt_Q: min.*

Theorem 1. *m-MAST problem is $\mathcal{NP}$-hard in the strong sense.*

Proof. We prove by showing that pickup and delivery problem (PDP) [6], which is known to be strongly $\mathcal{NP}$-hard [19], is reducible to m-MAST. Given an instance of PDP, we can construct an instance of m-MAST by relaxing the constraints on departure times at checkpoints, i.e., setting the departure times to be a time window $[0,\infty]$. In this way a solution to the constructed m-MAST corresponds to the original PDP. So solving PDP is no harder than solving m-MAST. Since the reduction can certainly be done in polynomial time $O(|N|)$, we have proven that m-MAST is strongly $\mathcal{NP}$-hard. □

2.2 MIP Model

The m-MAST scheduling problem is formulated as the following mixed integer program (MIP):

$$\min \quad z = w_1 \sum_{v\in V} \sum_{(i,j)\in A} \delta_{ij} x_{i,j}^v + w_2 \sum_{k\in K} (d_k - p_k) + w_3 \sum_{k\in K} (p_k - \tau_k) \tag{2.1}$$

Subject to

$$\sum_{v\in V} \sum_{i} x_{i,j}^v = 1 \quad \forall j \in N \backslash \{1, TC_0+1, 2TC_0+1, ..., TC\} \tag{2.2}$$

$$\sum_{v\in V} \sum_{j} x_{i,j}^v = 1 \quad \forall j \in N \backslash \{TC_0, 2TC_0, ..., TC\} \tag{2.3}$$

$$\sum_{j} x_{i,j}^v = \sum_{i} x_{j,i}^v \quad \forall j \in N \backslash \{1, TC_0, TC_0+1, 2TC_0, ..., TC\}; v \in V \tag{2.4}$$

$$t_i = \theta_i \quad \forall i \in N_0 \tag{2.5}$$

$$p_k = t_{ps(k)} \quad \forall k \in K \backslash K_{PND} \tag{2.6}$$

$$d_k = \bar{t}_{ds(k)} \quad \forall k \in K \backslash K_{NPD} \tag{2.7}$$

$$\sum_{v\in V} \sum_{r\in HYBR(K)} z_{k,r}^v = 1 \quad \forall k \in K \backslash K_{HYB} \tag{2.8}$$

$$p_k \geq t_{pc(k,r,v)} - M(1 - z^v_{k,r}), \forall k \in K_{PND}, r \in RD, v \in V \tag{2.9}$$

$$p_k \leq t_{pc(k,r,v)} + M(1 - z^v_{k,r}), \forall k \in K_{PND}, r \in RD, v \in V \tag{2.10}$$

$$d_k \geq \bar{t}_{dc(k,r,v)} - M(1 - z^v_{k,r}), \forall k \in K_{NPD}, r \in RD, v \in V \tag{2.11}$$

$$d_k \leq \bar{t}_{dc(k,r,v)} + M(1 - z^v_{k,r}), \forall k \in K_{NPD}, r \in RD, v \in V \tag{2.12}$$

$$p_k \geq \tau_k \quad \forall k \in K \tag{2.13}$$

$$d_k \geq p_k \quad \forall k \in K \tag{2.14}$$

$$\bar{t}_j \geq t_i + \sum_{v \in V} x^v_{i,j} \delta_{i,j} - M(1 - \sum_{v \in V} x^v_{i,j}) \quad \forall (i,j) \in A \tag{2.15}$$

$$t_i \geq \bar{t}_i + b_i \quad \forall i \in N \backslash \{1, TC_0 + 1, ..., Ve \times TC_0 + 1\} \tag{2.16}$$

$$\sum_j x^v_{ps(k),j} - \sum_j x^v_{j,ds(k)} = 0 \quad \forall v \in V; k \in K_{PD} \cup K_{NPND} \tag{2.17}$$

$$\begin{aligned} &\sum_{r \in HYBR(k)} \sum_j x^v_{pc(k,r,v),j} - \sum_j x^v_{j,ds(k)} = 0, \\ &\forall v \in V; k \in K_{PND} \end{aligned} \tag{2.18}$$

$$\begin{aligned} &\sum_j x^v_{ps(k),j} - \sum_{r \in HYBR(k)} \sum_j x^v_{j,dc(k,r,v)} = 0, \\ &\forall v \in V; k \in K_{NPD} \end{aligned} \tag{2.19}$$

The objective function (2.1) minimizes the weighted sum of three different factors, namely the total vehicle time traveled, the total travel time of all passengers and the total waiting time of all passengers. Here waiting time is the time gap between the passengers ready time and the actual pickup time. Network constraints (2.2), (2.3) and (2.4) allow each stop (except for the starting and ending nodes of each vehicle) to have exactly one incoming arc and one outgoing arc, which guarantee that each stop will be visited exactly once by the same vehicle. Constraint (2.5) forces the departure times from checkpoints to be fixed since they are pre-scheduled. Constraints (2.6) and (2.7) make the pickup time of each request (except for the PND) and the drop-off time of each request (except for the NPD) equal to the departure time and the arrival time of its corresponding node, respectively. Constraint (2.8) allows exactly one z variable to be equal to 1 for each hybrid request, assuring that a unique ride of a unique vehicle will be selected for its pickup or drop-off checkpoint. Constraints (2.9) and (2.10) fix the value of for each request depending on the z variable. Similarly, constraints (11) and (12) fix the value of the variable for each request. Constraints (2.13) and (2.14) guarantee that the pick-up time of each passenger is no earlier than her/his ready time and is also no later than the corresponding drop-off time. Constraint (2.15) is an aggregate form of sub-tour elimination constraint similar to the Miller-Tucker-Zemlin (MTZ) constraint. Constraint (2.16) assures that at each node the departure time is no earlier than the arrival time plus the service time. Constraints (2.17), (2.18) and (2.19) are the key constraints in multi-vehicle MAST MIP formulation, which assure that the pickup and drop-off stop of each request are served by the same vehicle.

CHAPTER 3 AN ANALYTICAL FLEET SIZING MODEL

In this section, we derive the critical demand to identify the switch point between the single-vehicle MAST system and the multi-vehicle MAST system. The number of trips R and the number of checkpoints C are fixed for both MAST systems. The total demand (including all types of requests) is considered to be deterministic during the whole service period of the MAST system. All the requests are assumed uniformly distributed in space and time, thus the non-checkpoint stops (NP and ND) are uniformly distributed in the service area. For simplicity, the time intervals between the departure times of two consecutive checkpoints are assumed to be uniform.

We define the following notation in this section:

- s_0 = service time at an inserted stop
- w = allowed deviation on the y-axis
- v = bus speed
- t = time interval between departure times of two consecutive checkpoints
- t_v = time headway between two consecutive vehicles
- $E(T_{rd}^{PD})$ = expected value of ride time of a PD passenger
- $E(T_{rd}^{PND})$ = expected value of ride time of a PND passenger
- $E(T_{rd}^{NPD})$ = expected value of ride time of an NPD passenger
- $E(T_{rd}^{NPND})$ = expected value of ride time of an NPND passenger
- $E(M)$ = expected value of travel miles of a vehicle
- $E(T_{rd})$ = expected value of ride time of all the customers
- $E(T_{wt})$ = expected value of waiting time of all the customers
- $\alpha, \beta, \gamma, \delta$ = portion of PD, PND, NPD, NPND requests respectively, and $\alpha+\beta+\gamma+\delta=1$

3.1 Performance Measures and Utility Function

$E(M), E(T_{wt}) and E(T_{rd})$ are the performance measures for the MAST system with associated weights. The weight assignment would change in different circumstances. A sensitivity experiment for w_1 will be conducted later. We assume that the weight assignment is fixed for various cases here. The total utility value U is defined as follows:

$$U = w_1 \times \frac{E(M)}{v} + w_2 \times E(T_{wt}) + w_3 \times E(T_{rd}) \tag{3.1}$$

This utility function is consistent with the objective function formulated in Chapter 2. It is obvious that lower values of total utility U indicate better performance of the MAST system. In the next sections we will discuss the analytical computation of U in the one-vehicle case and the two-vehicle case, respectively, for the MAST operating policy. To calculate the expected values of the performance measures, we assume a static situation in which all the requests have been scheduled through a feasible and optimal procedure. This static situation can reflect an expected performance of the MAST system.

3.2 Analytical Modeling for the One-Vehicle Case

Since NP/ND customers are uniformly distributed within the whole service area, a service area delimited by any pair of consecutive checkpoints is defined as a basic unit. As depicted in Fig. 3.1, denote y as the vertical distance between any pair of NP/ND requests within the basic unit of service area, and we have the expected value of y: $E(y) = w/3$. Denote y' as the vertical distance between one of the two consecutive checkpoints (both located at w/2 on the y-axis) and its closest NP/ND stop within a basic unit of service area, and we have the expected value of y': $E(y') = w/4$. Then the formulation for three performance measures will be discussed.

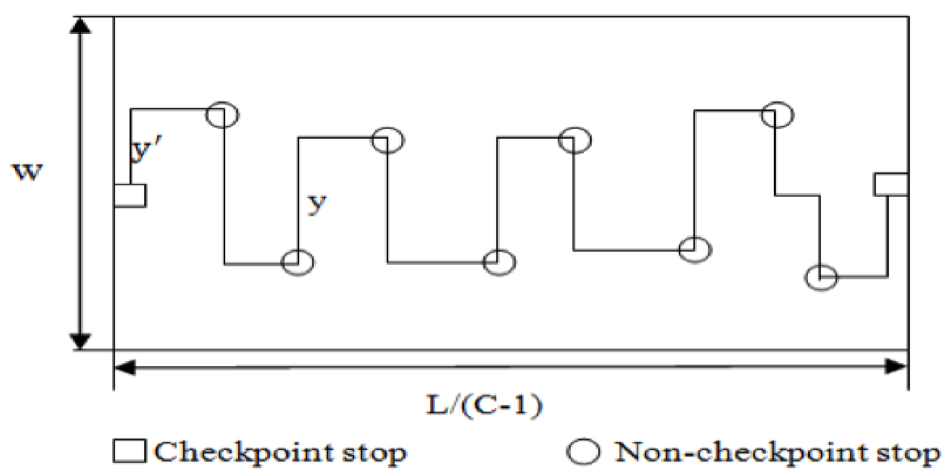

Figure 3.1: Illustration for bus route within a basic unit area

3.2.1 Ride Time

Denote E_0^{PD} as the expected ride time of a PD customer within a basic unit of service area, n_0 as the demand density, meaning the average number of NP/ND stops that need to be inserted between two consecutive checkpoints in one trip, n' as the total number of NP/ND stops that need to be inserted into the schedule and N as the total number of customers. The following equations for n_0, n' and N hold:

$$n_0 = n'/[R(C-1)] \tag{3.2}$$
$$n' = |NPD| + |PND| + 2 \times |NPND| \tag{3.3}$$
$$N = |PD| + |NPD| + |PND| + |NPND| \tag{3.4}$$

Where an NPD (PND) request has one NP (ND) stop to be inserted, and an NPND request has two stops (one NP and one ND) to be inserted into the schedule, then the formulation of E_0^{PD} is as follows:

$$E_0^{PD} = \frac{L}{(C-1)v} + \frac{w}{v}[\frac{1}{4} \times 2 + \frac{1}{3}(n_0 - 1)] + s_0 \times n_0 \tag{3.5}$$

Where the first term is the travel time for horizontal distance between two consecutive checkpoints with no back-tracking policy, the second term indicates the travel time for vertical deviation with n_0 stops scheduled, and the third term stands for the service time at n_0 stops. Extending to different units of the service area, the expected ride time of a PD customer is shown by Eq. 3.6. The derivation process is detailed in Appendix A.

$$E(T_{rd}^{PD}) = E_0^{PD} + (C-2)t/3 \tag{3.6}$$

Since all the requests are uniformly distributed, the NP (ND) stop of an NPD (PND) request is expected to be located at the middle of two consecutive checkpoints, which means the numbers of requests prior to and after it within a basic unit of service area should be the same. Thus, the expected ride time of a PND or NPD customer whose pickup or drop-off checkpoint is located within a basic unit of service area has the following equation:

$$E_0^{PND/NPD} = E_0^{PD}/2 \tag{3.7}$$

The expected ride time of a PND/NPD customer is half the value of a PD customer within one basic unit of service area. Similarly, considering the possibility of traversing different units of service area, the expected ride time of a PND/NDP customer is:

$$E(T_{rd}^{PND/NPD}) = \frac{1}{2}E_0^{PD} + \frac{C-2}{3}t \tag{3.8}$$

Note that if the two non-checkpoint stops of an NPND request are scheduled within two consecutive checkpoints, the ride time of this NPND request is expected to be one third of the total average ride time between the two consecutive checkpoints (analogous to $E(|x-y|) = (U-L)/3, \text{if } x,y \in [L,U]$). Thus the expected ride time of an NPND customer with two stops scheduled within one basic unit of service area is given by Eq. 3.9. The expected ride time of an NPND customer is formulated in Eq. 3.10. The detailed derivation can be found in the Appendix.

$$E_0^{NPND} = E_0^{PD}/3 \tag{3.9}$$

$$E(T_{rd}^{NPND}) = \frac{E_0^{PD}}{3(C-1)} + \frac{C(C-2)}{3(C-1)}t \tag{3.10}$$

Thus, the expected ride time of all the customers with different types of requests can be calculated by Eq. 3.11:

$$
\begin{aligned}
E(T_{rd}) &= E(T_{rd}^{PD}) \cdot |PD| + E(T_{rd}^{PND}) \cdot |PND| \\
&+ E(T_{rd}^{NPD}) \cdot |NPD| + E(T_{rd}^{NPND}) \cdot |NPND|
\end{aligned}
\tag{3.11}
$$

3.2.2 Waiting Time

Since all the requests discussed here do not exceed the saturation demand and they are uniformly distributed without any obvious variation in this static situation for the analytical modeling, it can be concluded that the customer will be picked up within two trips (one cycle) of a vehicle for any type of request. So the expected waiting time of a customer with any type of request is equal to the total time of one trip. The following equation holds:

$$
\begin{aligned}
E(T_{wt}^{PD}) &= E(T_{wt}^{PND}) = E(T_{wt}^{NPD}) \\
&= E(T_{wt}^{NPND}) = (C-1)t
\end{aligned}
\tag{3.12}
$$

Thus, we can obtain the expected value of waiting time of all the customers with different types of requests:

$$
E(T_{wt}) = N(C-1)t \tag{3.13}
$$

3.2.3 Miles Traveled

For the expected miles traveled by the vehicle during the whole service time, there are two terms formulated here. The first term $E(M_0)$ is the total horizontal miles that a vehicle has to travel. The second term $ext\_E(M)$ is the extra miles that a vehicle is supposed to travel due to the insertion of non-checkpoint stops. Thus the expected miles traveled by a vehicle during the whole service period is formulated as following:

$$
\begin{aligned}
E(M) &= E(M_0) + ext\_E(M) \\
&= R \cdot L + w[1/4 \times 2 + (n_0 - 1)/3]R(C-1)
\end{aligned}
\tag{3.14}
$$

Combining the three performance measures, the utility function for one-vehicle case is:

$$
\begin{aligned}
U_1 = & \frac{w_1}{v}\{R \cdot L + \frac{w[R(C-1)}{6} + \frac{(\beta+\gamma+2\delta)N}{3}]\} \\
& + w_2(C-1)tN + w_3N\{\frac{L}{(C-1)v} \\
& + \frac{w}{v}[\frac{1}{4} \times 2 + \frac{1}{3}[\frac{(\beta+\gamma+2\delta)N}{R(C-1)} - 1]] \\
& + s_0 \times \frac{(\beta+\gamma+2\delta)N}{R(C-1)}\}[\alpha + \frac{\beta+\gamma}{2} + \frac{\delta}{3(C-1)}] \\
& + \frac{w_3N(C-2)t[\alpha+\beta+\gamma+C\delta/(C-1)]}{3}
\end{aligned}
\tag{3.15}
$$

3.3 Analytical Modeling for the Two-Vehicle Case

For the two-vehicle MAST system, note that the average waiting time is determined by two extreme cases: the shortest waiting time (equal to 0) and the longest one. Also note that the system is symmetrical such that the case with time headway t_v is equivalent to the case with time headway $2(C-1)t - t_v$. Thus we have the following relationship for the expected waiting time:

$$\begin{aligned} &E(T_{wt}^{PD}) = E(T_{wt}^{PND}) = E(T_{wt}^{NPD}) = E(T_{wt}^{NPND}) \\ &= \begin{cases} (C-1)t - t_v/2, & for \quad t_v < (C-1)t \\ t_v/2, & for \quad (C-1)t \le t_v \le 2(C-1)t \end{cases} \end{aligned} \tag{3.16}$$

Apparently, in the range of $[0, 2(C-1)t]$, the optimal t_v for Eq. 3.16 is $t_v = (C-1)t$ because of the symmetry of the system. In the following derivation it is assumed that $t_v = (C-1)t$, which means one vehicle starts from checkpoint 1, and the other one starts from checkpoint C simultaneously. Thus we have:

$$E(T_{wt}) = [N(C-1)t]/2 \tag{3.17}$$

Similar to the one-vehicle case, the expected miles traveled and the expected ride time for the two-vehicle case are formulated as follows:

$$\begin{aligned} E(M) &= 2 \times [E(M_0) + ext\_E(M)] \\ &= 2\{R \cdot L + w[\frac{1}{4} \times 2 + \frac{1}{3}(\frac{n_0}{2} - 1)]R(C-1)\} \end{aligned} \tag{3.18}$$

$$E_0^{PD} = \frac{L}{(C-1)v} + \frac{w}{v}[\frac{1}{4} \times 2 + \frac{1}{3}(\frac{n_0}{2} - 1)] + s_0 \times \frac{n_0}{2} \tag{3.19}$$

$$E(T_{rd}^{PD}) = E_0^{PD} + (C-2)t/3 \tag{3.20}$$

$$E(T_{rd}^{PND/NPD}) = E_0^{PD}/2 + (C-2)t/3 \tag{3.21}$$

$$E(T_{rd}^{NPND}) = \frac{E_0^{PD}}{3(C-1)} + \frac{C(C-2)}{3(C-1)}t \tag{3.22}$$

Combining the three performance measures, the utility function for the two-vehicle case is:

$$\begin{aligned} U_2 = &\frac{2w_1}{v}\{R \cdot L + w[\frac{R(C-1)}{6} + \frac{(\beta+\gamma+2\delta)N}{3}]\} \\ &+ w_3 N\{\frac{L}{(C-1)v} + \frac{w}{v}[\frac{1}{2} + \frac{1}{3}[\frac{(\beta+\gamma+2\delta)N}{2R(C-1)} - 1]] \\ &+ s_0 \times \frac{(\beta+\gamma+2\delta)N}{2R(C-1)}\}[\alpha + (\beta+\gamma)/2 + \frac{\delta}{3(C-1)}] \\ &+ \frac{w_3 N(C-2)t[\alpha+\beta+\gamma+C\delta/(C-1)]}{3} \\ &+ \frac{w_2(C-1)tN}{2} \end{aligned} \tag{3.23}$$

3.4 Critical Demand

The utility functions for the one-vehicle and two-vehicle cases are derived and shown in Eq. 3.15 and Eq. 3.23, respectively. By equating these two utility functions and solving for N, the critical demand N_c can be obtained. At this critical demand, the one-vehicle and two-vehicle systems will have the same system performance (including both operation cost based on vehicle miles traveled and service quality provided to customers). In other words, transit demand beyond this critical demand point would necessitate an increase in the fleet size.

By equating the two utility functions, the following quadratic equation can be obtained:

$$A_1N^2 + A_2N + A_3 = 0 \tag{3.24}$$

where,

$$\begin{aligned} A_1 = & \frac{w_3(\beta+\gamma+2\delta)}{R(C-1)}(\frac{w}{6v}+\frac{s_0}{2}) \\ & \cdot [1-\delta-\frac{\beta+\gamma}{2}+\frac{\delta}{3(C-1)}] \end{aligned} \tag{3.25}$$

$$A_2 = \frac{w_2t(C-1)}{2} \tag{3.26}$$

$$A_3 = -\frac{w_1}{v}[RL+\frac{wR(C-1)}{6}] \tag{3.27}$$

The critical demand can be obtained by solving the quadratic equation.

3.5 Experiments

In this section we conduct two types of experiments. First, we analytically derive the critical demand for switching between the one-vehicle and two-vehicle MAST systems and conduct numerical analysis. We also find the optimization results for the formulated MIP model using CPLEX©. The optimization results confirm the derived critical demand. Second, we perform a sensitivity analysis for the weight of vehicle miles traveled.

All the runs are conducted using CPLEX 12.0 x64 with default settings using a desktop computer with Core 2 CPU @3.00 GHz and 8GB RAM. Table 3.1 summarizes the basic model input parameters.

As mentioned before, here L denotes the distance between the two terminals, W denotes the maximum allowable deviation distance on the y-axis, C denotes the number of checkpoints, R denotes the number of trips, $\delta_{s,s+1}$ denotes the rectilinear travel time between two consecutive checkpoints, b_s denotes the service time for boarding and disembarking at each stop, t denotes the time interval between departure times of two consecutive checkpoints and w_1, w_2, w_3 are the objective function weights.

Table 3.1: System Parameters of Analytical Modeling

L	10 miles
W	1 mile
C	3
R	6
$\delta_{s,s+1}(s=1,...,TC-1)$	12 min
$b_s(s=1,...,TS)$	18 sec
$w_1/w_2/w_3$	0.4/0.4/0.2
t	25 min

Table 3.2: Utility Values from Analytical Results and CPLEX Results

N	Analytical Model		CPLEX	
	One-Vehicle	Two-Vehicle	One-Vehicle	Two-Vehicle
8	192.3	211.2	194.9	216.1
10	225.2	233.8	228.8	246.6
12	258.3	256.5	252.9	255.6
14	291.6	279.2	304.2	268.0
16	327.5	304.6	322.8	305.3
18	361.1	327.5	369.2	333.2
20	394.8	350.5	409.3	354.1

3.5.1 Validation of the Analytical Model

For various situations with different numbers of customers N, based on the previously derived utility functions and the given model input parameters (from this experiment on, R=6 is used), the analytical utility results for the one-vehicle case (ANA-1) and two-vehicle case (ANA-2) are calculated and shown in Table 3.2. The optimization results from the MIP model are obtained using CPLEX and also listed in Table 3.2. These CPLEX results are approximated by two quadratic trend lines for both the one-vehicle case (Poly.(SIM-1)) and two-vehicle case (Poly.(SIM-2)) and are plotted in Fig. 3.2, which also includes two lines representing the analytical results.

From Fig. 3.2 the following observations can be made with regards to the utility function curves for the one-vehicle and two-vehicle MAST cases.

- The analytical results match the CPLEX results for both cases even though there still exist some small deviations (e.g., when N is above 18 in the one-vehicle case). The analytical results are a little smaller than the corresponding CPLEX results, which might be caused by some idealized considerations of the analytical modeling that overestimate the system performance.

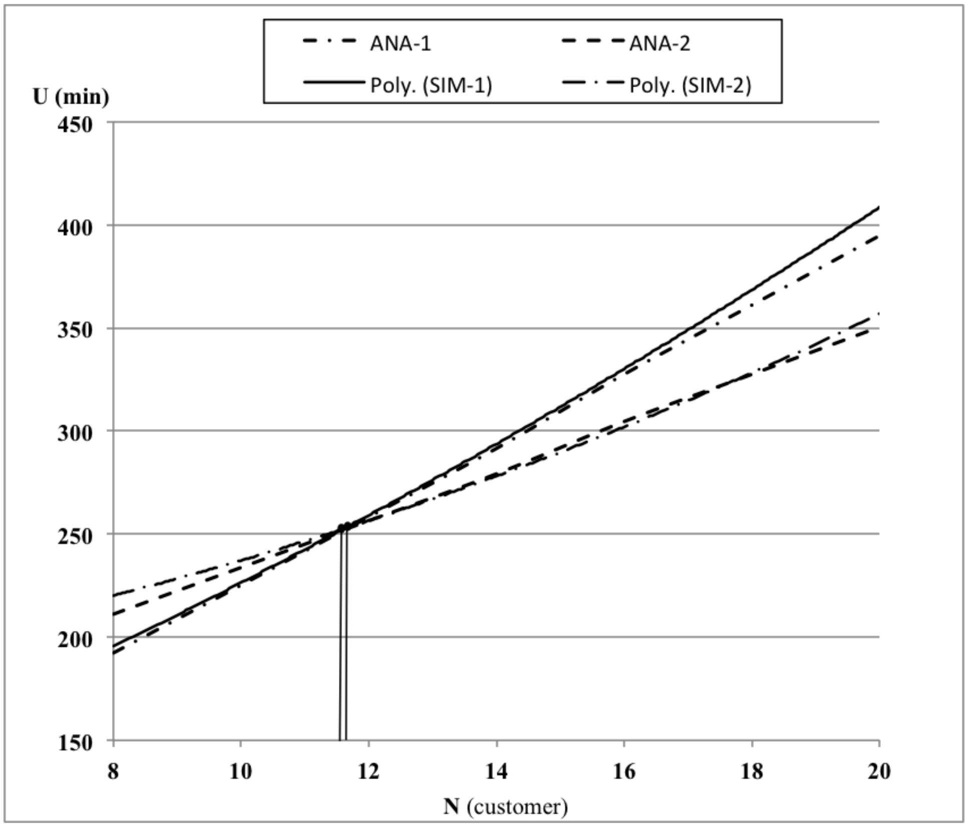

Figure 3.2: Utility function curves for one-vehicle case and two-vehicle case

- The critical demand (the intersection point) at which the one-vehicle case and the two-vehicle case have the same utility function value is around 12, corresponding to the critical demand density $n_0 = 1$ (see Eq. 3.2 for definition of n_0). Below this critical demand value, applying the one-vehicle MAST system can result in lower utility function value (better performance). Beyond this critical demand point, the two-vehicle MAST system is preferable.
- In general, for each case the CPLEX result curve fits the analytical result curve very well, suggesting that both the analytical and optimization methods can be used to estimate the actual utility function values and identify the critical demand.

3.5.2 Sensitivity Analysis

Special attention is paid to w_1, since the first term in the utility function 3.1 (i.e., the total miles traveled) reflects the cost increase when another vehicle is introduced into the fleet. To see how the critical demand Nc varies as a result of changing w_1, we set $w_2 : w_3 = 1 : 2$, $w_1 + w_2 + w_3 = 1$ and increase w_1 from 0.25 to 0.5. Table 3.3 and Fig. 3.3 show the results.

Both curves in Fig. 3.3 clearly indicate that N_c becomes larger with the increase of w_1. In other words, if we put more weight on the total miles traveled, the critical demand to switch from the one-vehicle MAST system to its two-vehicle counterpart will also increase. This increase is expected because when switching from the one-vehicle system to two-vehicle system, the last two

Table 3.3: N_c for Various w_1

	$w_1 = 0.25$	$w_1 = 0.4$	$w_1 = 0.5$
MIP	6.52	11.56	15.58
Analytical	5.88	11.64	17.28

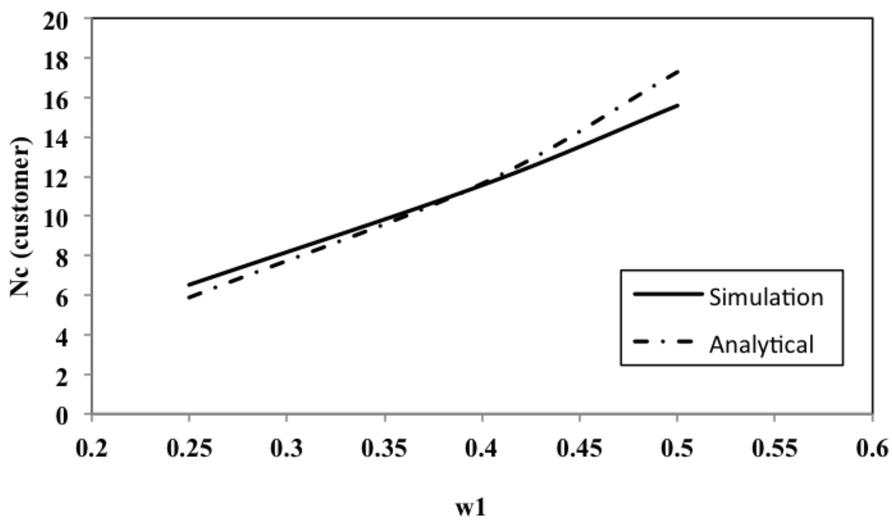

Figure 3.3: N_c with various w_1 (CPLEX vs. analytical results)

terms in the utility function (Eq. 3.1) reflecting the service quality are significantly decreased, whereas the first term is nearly doubled. Thus, the changes in w_1 will affect this trade-off among the three terms and lead to the increasing trend of critical demand as depicted in Fig. 3.3.

3.6 Chapter Summary

In this chapter we provide an analytical modeling framework to help MAST operators with their system planning and to identify the critical transit demand (N_c), which is used to decide when to switch from the one-vehicle MAST system to two-vehicle MAST system. Utilizing this analytical model and the MIP formulation, we also compare the utility function values generated by the two methods for the one-vehicle MAST system and the two-vehicle MAST system. Finally, a sensitivity analysis is conducted to find out the impact of a key modeling parameter w_1 on the critical demand.

All the analyses conducted in the research are based on a rectangular study area with model input parameters specified in Table 3.1. Experiments are conducted to find out the critical demand for switching between the one-vehicle and two-vehicle MAST systems. The results show that for the same model input both the MIP formulation and the developed analytical model generate approximately the same utility function values (i.e., system performance) and critical demands. The reasonable match between the two sets of outputs demonstrates the validation and the effectiveness of the proposed MIP formulation and the analytical framework for critical demand modeling.

Since the MAST problem is NP-hard, the proposed multi-vehicle MIP formulation can only optimally solve small to moderate-size problems (e.g, demand, number of checkpoints, etc.). Future work will include the development of proper valid inequalities/equalities and/or logic constraints to strengthen the proposed formulations and heuristic algorithms, allowing the formulated problems to be solved in real time or at large scales. It would also be interesting to extend the analytical modeling framework to consider different MAST configurations (e.g., three or more vehicles) and to identify the optimal fleet size as a function of demand.

CHAPTER 4 INSERTION HEURISTIC

4.1 Problem Description

Specifically, the multi-vehicle MAST system (m-MAST) analyzed in this paper consists of a set of vehicles V with predefined schedules along a fixed-route of C checkpoints (i=1,2,...,C). These checkpoints include two terminals (i=1 and i=C) and the remaining C-2 intermediate checkpoints. A rectangular service area is considered in this study as shown in Fig. 4.1 [14], where L is the distance between the two terminals, and W/2 is the maximum allowable deviation distance on each side of the fixed-route. A trip r is defined as a portion of the schedule beginning at one of the terminals and ending at the other after traversing all the intermediate checkpoints. Each vehicle performs R trips back and forth between the two terminals (see Fig. 4.1). Since the end terminal of a trip r corresponds to the start terminal of the following trip $r+1$, the total number of stops at the checkpoints of one vehicle is $TC_0 = (C-1)R+1$, and the total number of stops at the checkpoints of all vehicles is $TC = TC_0 \times |V| = [(C-1)R+1] \times |V|$. Hence, the initial schedule's array is represented by an ordered sequence of stops $s = 1,...,TC$.

We identify the checkpoints with $s = 1,...,TC$, and the non-checkpoint customers' requests (NP or ND) with $s = TC+1,...,TS$, where TS represents the current total number of stops. Each stop s of vehicle v has an associated departure and arrival times, respectively identified by $t_{s,v}$ and $t'_{s,v}$. As mentioned, the scheduled departure times $t_{s,v}$ of the checkpoints ($\forall s \leq TC$) are fixed and assumed to be constraints of the system, which can not be violated, while the $t_{s,v}$ of the non-checkpoint stops ($\forall s > TC$) and all the $t'_{s,v}$ are variables of the system.

Different from single-vehicle MAST, the optimization of multi-vehicle MAST is a more restricted problem in which three types of decisions have to be made:

1. Assignment: the solution has to assign requests to vehicles in a way the objective function is minimized.

2. Routing: the solution has to specify routes for vehicles so that the total miles are small.

3. Scheduling: the solution has to give schedules (sequence) to pick up and drop off customers so that the waiting and ride times are short.

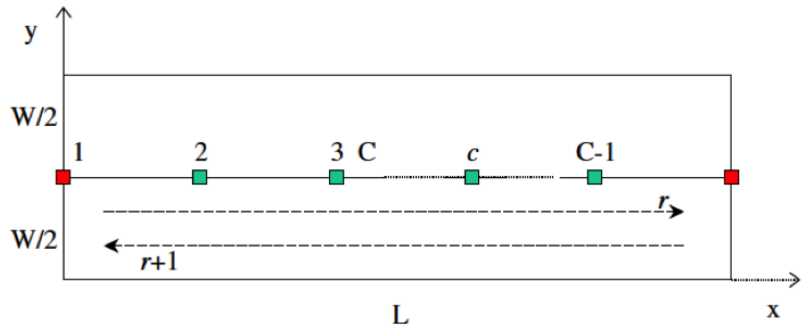

Figure 4.1: Multi-vehicle MAST system (Quadrifoglio et al., 2008)

The objective of this research is to develop an insertion heuristic algorithm that can reasonably approximate the optimality of the problem in a polynomial time so that the real-time dynamic operation of the service is possible. In another perspective, let $\alpha(s)$ represent the current position of stop s in the schedule, $\forall s$. The problem defined by an m-MAST system is to determine the indices $\alpha(s)$ and the departure/arrival times $t_{s,v}$, $\forall s > TC, \forall v$, and $t'_{s,v}$, $\forall s$, in order to minimize objective function while not violating the checkpoints fixed departure times $t_{s,v}$, $\forall s \leq TC, \forall v$, and the customers precedence constraints. The problem is solved by means of a cheapest insertion heuristic algorithm described in the following section.

It is assumed that vehicles are homogeneous, following checkpoints back and forth with a predefined time headway. The vehicles have unlimited capacity, which simplifies the mathematical problem, without compromising adherence to reality, as even small vehicles will almost never be filled up to capacity, due to much more restrictive time constraints.

4.2 Algorithm Description

4.2.1 Control Parameters

Usable Slack Time

Slack time is a crucial resource in the MAST system for vehicles to deviate from the main route to serve *NP* and *ND* requests. The initial slack time of vehicle v between two consecutive checkpoints s and $s+1$, $st^{(0)}_{s,s+1,v}$, is defined by

$$st^{(0)}_{s,s+1,v} = t_{s+1,v} - t_{s,v} - d_{s,s+1}/v_{speed} - h_{s+1}, \quad \forall s = 1,...,TC-1 \tag{4.1}$$

where v_{speed} is the vehicle speed, h_{s+1} the time allowed at stop $s+1$ for passengers boardings and dis-embarkments, and $d_{s,s+1}$ the distance between s and $s+1$. As more pickups and drop-offs occur off the base route, the current slack time $st_{s,s+1,v}$ available is progressively reduced. Initially,

$$st_{s,s+1,v} = st^{(0)}_{s,s+1,v}, \quad \forall s = 1,...,TC-1 \tag{4.2}$$

At time t_{now}, usable slack time $st^u_{s,s+1,v}$ of vehicle v between stop s and $s+1$ is defined as follows:

$$st^u_{s,s+1,v} = \begin{cases} \pi^{(0)}_{s,s+1} st^{(0)}_{s,s+1,v}, \text{ if } t_{now} < t_{s,v} \\ [1+(\pi^{(0)}_{s,s+1}-1)(1-\frac{t_{now}-t_{s,v}}{t_{s+1,v}-t_{s,v}})]st^{(0)}_{s,s+1,v}, \text{ if } t_{s,v} \leq t_{now} \leq t_{s+1,v} \\ st^{(0)}_{s,s+1,v}, \text{ if } t_{now} > t_{s+1,v} \end{cases} \tag{4.3}$$

where $\pi^{(0)}_{s,s+1}$ is the parameter controlling the usage of slack time, the lower it is set, the more slack time is reserved for future insertions.

Setting $\pi^{(0)}_{s,s+1}$ too low would prevent the algorithm from working properly. From [2], we know that the minimum value of $\pi^{(0)}_{s,s+1}$ should be:

$$\pi^{(0)min}_{s,s+1} = (W/v_{speed} + h_q)/st^{(0)}_{s,s+1,v} \tag{4.4}$$

Backtracking Distance

The vehicles could drive back and forth with respect to the direction of a trip r while serving customers in the service area, not only consuming the extra slack time but also having a negative impact on the customers already onboard, who may perceive this behavior as an additional delay. Therefore, we limit the amount of backtracking in the schedule. The backtracking distance indicates how much the vehicle drives backward on the x-axis while moving between two consecutive stops to either pick up or drop off a customer at a non-checkpoint stop with respect to the direction of the current trip. Formally, given two consecutive stops identified by a and b, such that $\alpha(a)+1 = \alpha(b)$, and the vector $\hat{d_{a,b}}$ representing the vector from a to b, the backtracking distance $bd_{a,b}$ is defiend as the negative component of the projection of $\hat{d_{a,b}}$ along the horizontal unit vector $\hat{d_r}$, representing the direction of the current trip r ($1 \rightarrow C$ or vice versa) as follows:

$$bd_{a,b} = -min(0, \hat{d_r} \cdot \hat{d_{a,b}}) \tag{4.5}$$

We define the control parameter $BACK > 0$ that is the maximum allowable backtracking distance that the vehicle can ride between any two consecutive stops. Clearly, with $BACK \geq L$, any backtracking is allowed.

4.2.2 Feasibility

While evaluating a customer request, the algorithm needs to determine the feasibility of the insertion of a new stop $s = q$ between any two consecutive stops a and b already scheduled. The extra time needed for the insertion is computed as follows:

$$\Delta t_{a,q,b} = (d_{a,q} + d_{q,b} - da, b)/v_{speed} - h_q \tag{4.6}$$

Let m and $m+1$ be the checkpoints before and after stops a and b in the schedule. It is feasible to insert q between a and b if the following conditions hold:

$$
\begin{cases}
\Delta t_{a,q,b} \leq min(st_{m,m+1,v}, st^{u}_{m,m+1,v}) \\
bd_{a,q} \leq \text{BACK} \\
bd_{q,b} \leq \text{BACK}
\end{cases}
\tag{4.7}
$$

4.2.3 Cost Function

For each feasible insertion of a stop q, the algorithm computes the following quantities:

- ΔRT: the sum over all passengers of the extra ride time, including the ride time of the customer requesting the insertion.
- ΔWT: the sum over all passengers of the extra waiting time.

Finally, the cost function is defined as:

$$
COST = w_1 \times \Delta t_{a,q,b} + w_2 \times \Delta RT + w_3 \times \Delta WT \tag{4.8}
$$

4.2.4 Buckets

Considering the schedule's array, Each checkpoint c is scheduled to be visited by each vehicle a number of times, with different stop indices $s(r,c,v)$ (stop index of the r-th occurrence of checkpoint c in the schedule of vehicle v), depending on the fleet size and how many trips R are planned.

For each intermediate checkpoint $c = 2, ..., C-1$ and each $v \in V$ the indices $s(k,c,v)$, which identify them in the schedule, are computed by the following sequence:

$$
\begin{aligned}
s(r,c,v) = & 1 + (C-1)(r-1) + \frac{(C-1) + (-1)^r[(C-1) - 2(c-1)]}{2} \\
& + (v-1)TC_0 \quad \forall r = 1, ..., R, \forall v = 1, ..., Ve
\end{aligned}
\tag{4.9}
$$

For the terminal checkpoints 1 and C, since their frequency of occurrence is halved, the sequences are as follows:

$$
s(r,1,v) = 1 + 2(C-1)(r-1) + (v-1)TC_0 \quad \forall r = 1, ..., 1 + \lfloor R/2 \rfloor, \forall v = 1, ...Ve \tag{4.10}
$$

$$
s(r,C,v) = C + 2(C-1)(r-1) + (v-1)TC_0 \quad \forall r = 1, ..., 1 + \lceil R/2 \rceil, \forall v = 1, ...Ve \tag{4.11}
$$

Definition 4. *For every checkpoint c and every $v \in V$, a bucket of c and v is a portion of the schedule delimited by two successive occurrences of c by the same vehicle, namely all the stops s in the current schedule's array such that* $\alpha[s(r,c,v)] \leq \alpha(s) < \alpha[s(r+1,c,v)]$ *for any allowable r, as described in Eq. 4.9 - Eq. 4.11.*

The buckets' definition for *NPND*-type customers needs to be slightly revised. We characterize the sequence representing the occurrences of any terminal checkpoint ($c = 1$ or C):

$$
s(r, 1 \text{ or } C, v) = 1 + (C-1)(r-1) + (v-1)TC_0 \forall r = 1, ..., R+1, \forall v = 1, ..., Ve \tag{4.12}
$$

For $NPND-type$ customers, a bucket represents all the stops s such that $\alpha[s(r,1 \text{ or } \mathrm{C},v)] \leq \alpha(s) < \alpha[s(r+1,1 \text{ or } \mathrm{C},v)]$ for any allowable r as described in Eq. 4.12.

4.2.5 Assignment and Insertion Procedure

PD Type

PD-type requests do not need any insertion procedure, since both pick-up and drop-off points are checkpoints and they are already part of the schedule. However assignment procedure is needed. The assignment is made by choosing the vehicle traveling as the desired direction of the customer that can provides the earliest pickup time .

PND Type

PND-type customers need to have their ND stop q inserted in the schedule. The algorithm checks for insertions feasibility in the buckets of the P checkpoint. Since the ND stop cannot be scheduled before P, the first bucket to be examined is the one starting with the first occurrence of P following the current position of the vehicle, that is the bucket delimited by $s(k,P,v)$ and $s(k'+1,P,v)$, with $(k',v) = min_{k,v}s(k,P,v)$, s.t. $t_{s(k,P,v)} \geq t_{now}$. Among the feasible insertions between all pairs of consecutive stops a, b in the first bucket of all the vehicles, the algorithm selects the one with the minimum COST and then stops. The customer is therefore scheduled to be picked up at $s(k,P,v)$ and dropped off at the ND inserted stop q. If no feasible insertions are found in the first bucket, the algorithm repeats the procedure in the second bucket, assuming that the customer will be picked up at the beginning of it corresponding to the following occurrence of P, that is $s(k'+1,P,v)$. This process is repeated bucket by bucket until at least one feasible insertion is found.

NPD Type

NPD-type customers need to have their NP stop q inserted in the schedule. The algorithm runs in a very similar way except for changing the insertion from ND to NP.

NPND Type

A NPND-type customer requires the insertion of two new stops q and q'; therefore, the insertion procedure will be performed by a $O(|V| \cdot |TS|^2)$ procedure, meaning that for each feasible insertion of the NP stop q, the algorithm checks feasibility for the ND stop q'. A NPND feasibility is granted when both NP and ND insertions are simultaneously feasible. The search for NPND feasibility is performed with the additional constraint of having q scheduled before q'.

Recall that buckets correspond to the trips for an NPND type customer. The search for NPND feasibility is performed in at most two consecutive buckets meaning that when checking for NP insertion feasibility in bucket i and $i+1$, the algorithm looks for ND insertion feasibility only in bucket i and $i+1$. For each of the vehicles, the algorithm starts checking the NPND feasibility in the first bucket delimited by the current position of the bus (x_b, y_b) and the end of the current trip r. This is the first occurrence in the schedule of one of the terminal checkpoints $s = 1$ or $s = C$, namely $s(k', 1 \text{ or } C, v) = min_{k,v} s(k, 1 \text{ or C}, v)$, s.t. $t_{s(k,1 \text{ or } C)} \geq t_{now}$. Among all the feasible NPND insertions in the first bucket, the algorithm selects the one with the minimum COST. If no NPND feasibility is found, the algorithm will then check pairs of two consecutive buckets at a time, increasing the checking-range by one bucket at each step (buckets 1/2, then buckets 2/3, . . . , i/(i + 1), etc.). While checking buckets i/i + 1, we already know that NPND insertion is infeasible in bucket i (because it has been already established before in the procedure while checking buckets (i-1)/i). Therefore, while NP insertion feasibility needs to be considered in both buckets (since NPND insertion infeasibility in bucket i does not prevent NP insertion to be feasible in i), ND insertion needs to be checked only in bucket i + 1. The procedure will continue until at least one NPND feasible insertion is found.

Rejection Policy

The general assumption while performing the insertion procedure is a no-rejection policy from both the MAST service and the customers. Thus, the algorithm attempts to insert the customers' requests checking, if necessary, the whole existing schedule of all the vehicles bucket by bucket. So generally pending requests will not be rejected, rather they may be postponed. However, in a static environment, where the trips of service are very short, requests are more likely to be rejected.

4.2.6 Update Time Windows

The algorithm provides customers at the time of the request with time windows for their pickup and drop-off locations. Assuming the customer is assigned to vehicle v, the earliest departure time $et_{q,v}$ from q is computed as follows:

$$et_{q,v} = t_{a,v} + d_{a,q}/v_{speed} + h_q \tag{4.13}$$

where $t_{a,v}$ represents the current departure time from stop a of vehicle v. Also the departure time of q is initialized likewise:

$$t_{q,v} = t_{a,v} + d_{a,q}/v + h_q = et_{q,v} \tag{4.14}$$

It can be easily shown that $et_{q,v}$ is a lower bound for any further updates of $t_{q,v}$. The algorithm then computes the latest departure time from q, $lt_{q,v}$, as follows:

$$lt_{q,v} = et_{q,v} + st_{m,m+1,v} \tag{4.15}$$

We prove that $lt_{q,v}$ is an upper bound for $t_{q,v}$ by the following contradiction argument. Let us use the superscript β (with $\beta = 0, ..., f$) to indicate the β-th update of a variable and suppose that $t_{q,v}^{(f)} > lt_{q,v}$, we have $t_{q,v}^{(f)} - t_{q,v}^{(0)} > lt_{q,v} - t_{q,v}^{(0)}$. We also know that:

$$\begin{aligned} t_{q,v}^{f} - t_{q,v}^{(0)} =& (t_{q,v}^{f} - t_{q,v}^{f-1}) + \cdots + (t_{q,v}^{\beta} - t_{q,v}^{\beta-1}) + \cdots + (t_{q,v}^{1} - t_{q,v}^{0}) \\ =& \Delta t_f + \cdots + \Delta t_\beta + \cdots + \Delta t_1 = \sum_{k=1}^{f} \Delta t_k \end{aligned} \tag{4.16}$$

and from Eq. 4.14 and Eq. 4.15, $lt_{q,v} - t^{(0)}_{q,v} = lt_{q,v} - et_{q,v} = st_{m,m+1,v}$, but this would imply $\sum_{k=1}^{f} t_{k,v} > st_{m,m+1,v}$, meaning that the sum of the extra time needed for insertions after the insertion of q had exceeded the total slack time available after the insertion of q and this is a contradiction, since the feasibility check would have prevented this from happening. Therefore, Eq. 4.15 says that future possible insertions between m and q will delay $t_{q,v}$ to a maximum total amount of time bounded by the currently available slack time.

In a similar fashion, the earliest and latest arrival times, $et'_{q,v}$ and $lt'_{q,v}$, are computed. As a result, the customer, once accepted, is provided with $et_{q,v}, lt_{q,v}, et'_{q,v}$, and $lt'_{q,v}$, being aware that their actual times $t_{q,v}$ and $t'_{q,v}$ will be bounded by these values:

$$et_{q,v} \leq t_{q,v} \leq lt_{q,v} \tag{4.17}$$
$$et'_{q,v} \leq t'_{q,v} \leq lt'_{q,v} \tag{4.18}$$

While a P request has $et_{P,v} = t_{P,v} = lt_{P,v}$ because the departure time from a checkpoint is a constant in a MAST system, a D request will have $et'_{D,v} \leq t'_{D,v} \leq lt'_{D,v}$. Clearly, NP and ND requests will also have $et_{NP,v} \leq t_{NP,v} \leq lt_{NP,v}$ and $et'_{ND,v} \leq t'_{ND,v} \leq lt'_{ND,v}$.

Algorithm 1 Overall Scheme

```
for t = service_start to service_end do
  if If customer request received then
    for each vehicle v (v = 1,2,...,ve) do
      while No feasible insertion found do
        1. Check the current bucket for the feasible insertion spots for customer i's NP or
           ND request.
        2. Go to check the next bucket
      end while
      Record sub_mincost(v), the min. insertion cost of v
    end for
    if at least one vehicle have feasible insertion spot then
      1. assign customer i to the v with minimum sub_mincost(v) for ∀v = 1,2,...,ve
      2. customer i accepted
      3. update
    else
      customer i rejected
    end if
  end if
end for
```

4.2.7 Overall Approach

The overall approach is described by Algorithm 1. The overall time complexity of the algorithm is $O(T \cdot |N| \cdot |V|)$, where T is the overall service horizon, $|N|$ is the total number of customers and $|V|$ is the total number of vehicles.

4.3 Experimental Results

The target of this section is to show that the proposed insertion heuristic can be used as an efficient scheduling tool for m-MAST systems. Three series of experiments are conducted. First we experiment on the control parameters to find the best configuration for the heuristic. Second, 2-MAST system and 1-MAST system are compared to confirm the potential of m-MAST to handle heavy demand. Last, the algorithm is compared to optimality obtained through solving the MIP using *CPLEX*©.

4.3.1 Performance Measures and System Parameters

The performance measures of interests include:

- PST_v: Percentage of the total initial slack time of vehicle v ($\sum st^0_{s,s+1,v}$) consumed.
- WT: Average waiting time (the difference between actual pick-up time and request time) per customer, as mentioned before.
- RT: Average ride time (the difference between drop-off time and pickup time) per customer, as mentioned before.
- M_v: Total mileage traveled by vehicle v.
- BF: Balance factor, an intuitive factor shows the ratio of the number of passengers served by the 2 vehicles. Ideally, this value should be close to 1.
- $Rej.Rate$: Rejection rate shows the percentage of customers that are not accepted.
- Z: The objective function Z in this section conforms to that in the MIP in Chapter 2, namely the weighted sum of mileage, customer ride times and waiting times.

Table 4.1 shows a summary of the parameters that are used in the experiments.

Table 4.1: System Parameters of the Insertion Heuristic

L	10 miles
W	1 mile
C	3
R	4, 6
v	25 miles/h
$b_s(s = 1, ..., TS)$	18 sec
$w_1/w_2/w_3$	0.4/0.4/0.2
t_v	50 min

The customer types are assumed to be distributed as in Table 4.2.

As in previous chapters, it is assumed that the checkpoint requests (P and D) are uniformly distributed among the C checkpoints, and that non-checkpoint requests (NP and ND) are

Table 4.2: Customer Type Distribution

PD	PND	NPD	NPND
10%	40%	40%	10%

uniformly distributed in the service area. The simulation is run for 50 hours (equivalently, $R = 60$).

4.3.2 Algorithm Performance

First we seek the saturation level of the 2-MAST system by examining the WT, PST_v and $Rej.Rate$ values for different values of θ. Note that we set the control parameters $BACK = L$ and $\pi^{(0)}_{s,s+1} = 1, \forall s = 1, ..., TC - 1$, allowing the most freedom on checking insertion feasibility.

As analyzed in Chapter 3, given that the demand is uniform, for systems well below saturation level, the WT values are expected to be characterized by Eq. 3.17. Applying the parameters in Table 4.1, this value becomes 25 min. A slightly larger value of WT shows that the system is near the saturation level, but still below it. The system on average is stable. If instead the WT value increases over the simulation time drastically, then it means the demand is above the saturation level, resulting in system instability. Another indicator is PST_v which shows how much slack time has been used by vehicles. PST_v values around 90% indicate that the demand is around saturation level. $Rej.Rate$ also gives indication of whether the demand is above saturation level. Under a moderate demand, the MAST system only rejects customers that appear towards the very end of the service, thus giving a very low level of $Rej.Rate$ (usually <1%, as will be shown in experiments later). If the demand is above saturation level, it is more likely more requests would be postponed to later rides, and finally be rejected.

As is shown in Table 4.3, the saturation level of 2-MAST under configuration A is around $\theta = 45$ customers per hour (configuration A2). This conclusion is drawn based on three facts: 1) PST values are well below 90% for both vehicles under $\theta = 40$, meaning this is below saturation level. 2) While under $\theta = 45$, the PST values are approaching 90% and the system begins to reject requests, WT is still stable, meaning the demand is around the saturation level. 3) Above $\theta = 45$ the WT increases significantly, and PST values are approaching 100%, meaning the demand is above the saturation level and increasing demand will cause system instability.

Therefore, without setting the control parameters, the 2-MAST system is able to handle a demand rate of 45 customers per hour using the proposed insertion heuristic.

Then we want to observe the effect of modifying the usable slack time $st^u_{s,s+1}$ by varying the values of $\pi^{(0)}_{s,s+1}$. Performance corresponding to each $\pi^{(0)}_{s,s+1}$ is compared mainly by means of Z. Other performance values are also examined. The simulation time is still 50 h ($R = 60$). Note that from Eq. 4.4, $\pi^{(0)min}_{s,s+1} = 0.22$. Table 4.4 summarized the results.

Table 4.4 shows the positive effect of tuning $\pi^{(0)}_{s,s+1}$. As decreasing $\pi^{(0)}_{s,s+1}$ from 1 to $\pi^{(0)min}_{s,s+1}$, the Z values (as well as other performance measures) reach their minimum values with configuration B5 at $\pi^{(0)}_{s,s+1} \approx 0.3$. As a result, the system drops well below the saturation level since

Table 4.3: Saturation Level for 2-MAST Under Configurations A

Configuration	A1	A2	A3
θ (customers per hour)	40	45	50
$BACK$ (miles)	L	L	L
$\pi^{(0)}_{s,s+1}$	1	1	1
Performance			
WT (min)	36.45	37.21	48.32
PST_1 (%)	84.0	89.2	93.9
PST_2 (%)	88.0	88.5	95.8
RT (min)	23.00	22.70	23.88
M (miles)	2042.6	2049.0	2094.8
BF	1.012	1.01	1.028
$Rej.Rate$ (%)	0	0.31	0.4
Saturation level?	Below	Yes	Above

Table 4.4: Effect of $\pi^{(0)}_{s,s+1}$ Under Configurations B

Configuration	B1=A2	B2	B3	B4	B5	B6	B7
θ	45	45	45	45	45	45	45
$BACK$ (miles)	L	L	L	L	L	L	L
$\pi^{(0)}_{s,s+1}$	1	0.75	0.5	0.4	0.3	0.22	0.2
Performance							
WT (min)	37.21	37.35	32.66	32.78	31.29	30.70	29.35
PST_1 (%)	89.2	87.0	84.6	82.4	81.2	77.3	74.2
PST_2 (%)	88.5	88.4	84.0	83.6	81.0	77.9	74.0
Sat. level?	Yes	Yes	Below	Below	Below	Below	Below
RT (min)	22.70	22.59	22.24	22.30	22.12	22.32	22.34
M (miles)	2049.0	2033.8	1990.85	1974.2	1949.9	1905.8	1867
Z	39024.4	39009.8	36551.8	36635.2	35788.8	35653.6	34494
BF	1.01	1.06	1.00	1.00	1.02	1.01	1.02
$Rej.Rate$ (%)	0.31	0.22	0.22	0.22	0.22	0.22	1.82

configuration B3. These results show the benefit of controlling the consumption of slack time and saving it for future insertions, thus resolving the "myopia dilemma".

It is of interest to observe the effect of limiting the backtracking distance. We perform another set of runs (configurations C) by keeping $\theta = 45$ and $\pi^{(0)}_{s,s+1} = 0.3$ and varying the $BACK$ parameter from L to 0. The results are in Table 4.5.

The best configuration according to Z (and all other performance measures) is found by setting $BACK$=0.2 miles, corresponding to C6.

Then we look for the new saturation level under more efficient parameter settings by performing another set of runs under configurations D.

Table 4.5: Effect of *BACK* Under Configurations C

Configuration	C1=B5	C2	C3	C4	C5	C6	C7
θ	45	45	45	45	45	45	45
BACK (miles)	*L*	0.8	0.5	0.3	0.2	0.1	0
$\pi^{(0)}_{s,s+1}$	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Performance							
WT (min)	31.29	31.29	30.94	30.53	30.29	30.85	30.85
PST_1 (%)	81.2	81.2	79.3	78.8	77.7	78.1	77.5
PST_2 (%)	81.0	81.0	79.7	78.5	77.6	78.4	79.5
Sat. level?	Below	Below	Below	Below	Below	Below	Below
RT (min)	22.12	22.12	22.19	22.08	22.01	22.24	22.28
M (miles)	1949.9	1949.9	1929.6	1919.4	1906.0	1913.94	1917
Z	35788.8	35788.8	35674.1	35380.2	35197.3	35662.7	35697
BF	1.02	1.02	1.01	1.01	1.00	1.00	1.01
$Rej.Rate$ (%)	0.22	0.22	0.22	0.22	0.22	0.22	0.22

Table 4.6 shows the results. It can be seen that by properly setting up the parameters, the system is able to handle a demand at least 22% larger than the initial configuration A2.

Table 4.6: New Saturation Level for 2-MAST Under Configurations D

Configuration	D1=C6	D2	D3	D4	D5
θ (customers per hour)	45	50	55	60	65
BACK (miles)	0.2	0.2	0.2	0.2	0.2
$\pi^{(0)}_{s,s+1}$	0.3	0.3	0.3	0.3	0.3
Performance					
WT (min)	30.29	32.68	37.10	56.97	107.97
PST_1 (%)	77.7	85.3	90.2	96.6	98.1
PST_2 (%)	77.6	82.8	90.3	96.2	98.3
Saturation level?	Below	Below	Yes	Above	Above
RT (min)	22.01	22.30	23.42	25.37	27.96
M (miles)	1906.0	1955.72	2000.2	2050.25	2065.3
BF	1.00	1.03	1.04	1.07	1.035
$Rej.Rate$ (%)	0.22	0.12	0.07	0.2	5.32

The new saturation level is $\theta = 55$. Note that from [2] we know the saturation level of 1-MAST is $\theta = 25$. This means by adding one vehicle into the fleet, the MAST system can handle more than double the demand of 1-MAST. This "$1+1>2$" fact proves encouraging potential of m-MAST.

4.3.3 2-MAST vs. 1-MAST

The performances of 2-MAST and 1-MAST are compared under their tuned control parameter settings (see [2] for details of proper parameter setting for 1-MAST) under three different levels of demand. The results are summarized in Table 4.7.

Table 4.7: 2-MAST vs. 1-MAST

θ	2-MAST (1-MAST)				
	Z	WT	RT	M	$Rej.Rate$
15	10946.9 (14965.1)	25.25 (51.92)	19.11 (21.35)	1487.8 (862.0)	0% (0.4%)
20	14814.4 (20610.9)	26.09 (52.19)	20.24 (23.57)	1564.0 (921.8)	0% (0.7%)
25	17882.3 (27533.2)	25.39 (58.64)	19.95 (24.36)	1626.9 (969.9)	0% (0.9%)

We make the following observation:

1. 2-MAST provides a service of WT nearly half of that of 1-MAST, which is certainly more attractable to customers. RT and $Rej.Rate$ also has a better value.
2. Although M value of 2-MAST is larger than 1-MAST due to the increase in fleet size, the overall objective value Z of 2-MAST is still significantly better than that of 1-MAST, indicating the superbness of 2-MAST.

4.3.4 Heuristic vs. Optimality

Although the worst-case analysis of approximation scheme is of theoretical interest, it becomes intractable in this research because of the existence of complicated time constraints and weighted combination of objective function. As a result we conduct several numerical experiments based on random generated demand to evaluate the performance of the algorithm. In this section, the results produced by the proposed heuristic is compared with the optimal results by solving the integer program using *CPLEX*©, a commercial solver.

The choice of $\pi^{(0)}_{s,s+1}$ and *BACK* is based on configuration C5. The results of two different settings of $R = 6$ and $R = 4$ are shown in Table 4.8 and Table 4.9, respectively. Note that the demand is represented by $(|PD|, |PND|, |NPD|, |NPND|)$.

Table 4.8: Heuristic vs. Optimality, R=6

		Heuristic				Optimal			
Demand	apx/opt	obj.	M	RT	WT	obj.	M	RT	WT
$(2,5,5,2)$	1.06	284.8	129.1	196.4	411.5	267.5	125.5	163.2	409.0
$(2,6,6,2)$	1.09	292.1	131.8	214.9	398.2	268.5	126.1	166.7	403.8
$(2,7,7,2)$	1.10	322.3	129.4	254.8	480.6	291.7	125.0	213.4	431.7
$(2,8,8,2)$	1.13	364.8	140.8	324.9	498.3	321.5	128.6	281.6	426.8
$(2,9,9,2)$	1.11	395.2	134.5	312.3	705.9	354.2	125.7	236.6	694.7

Based on the results of Table 4.8 and Table 4.9, the following observations can be made:

1. From the ratio of apx/opt we can see, the performance of the heuristic is reasonably good.
2. The ratio is reasonably stable which isn't growing intractably big as the demand increases.

Table 4.9: Heuristic vs. Optimality, R=4

		Heuristic				Optimal			
Demand	apx/opt	obj.	M	RT	WT	obj.	M	RT	WT
$(2,3,3,2)$	1.04	224.6	88.9	223.8	248.5	215.1	86.1	193.5	275.4
$(2,4,4,2)$	1.13	215.6	86.0	196.2	272.7	190.3	81.6	151.2	257.5
$(2,5,5,2)$	1.14	275.7	90.6	203.8	536.0	242.5	89.5	176.6	429.5
$(2,6,6,2)$	1.04	242.7	92.5	213.7	342.1	232.7	83.8	193.5	374.4
$(2,7,7,2)$	1.12	284.3	88.7	255.2	485.2	252.8	83.0	213.2	439.0

4.4 Conclusion and Future Research

In this chapter, we develop an insertion heuristic for scheduling m-MAST service. The algorithm allows customers to place a request, and once accepted, provides them with time windows for both pickup and drop-off points. Due to the dynamic nature of the environment, the algorithm makes effective use of a set of control parameters to reduce the consumption of slack time and enhance the algorithm performance. Due to the existence of complicated time constraints and weighted objective function, we resort to experiments to evaluate the algorithm. The results of simulations show the efficacy of the algorithm and its optimal control parameter setting and demonstrate that the algorithm can be used as an effective method to schedule m-MAST service. By comparing the performance of 2-MAST and 1-MAST, the potential of m-MAST to provide a more attractable service and a better overall operation cost is shown. In addition, a comparison versus optimality values computed by *CPLEX*© in a static scenario shows that the results obtained by the heuristic are not far from optimum.

CHAPTER 5 SUMMARY AND CONCLUSIONS

The mobility allowance shuttle transit system is a promising innovative concept that combines the low cost of fixed-route transit and the flexibility of demand-responsive transit. Previous literature addressed the design and scheduling issues of single-vehicle MAST, but so far, no research on multiple-vehicle MAST has been done.

In this research we first give the formal definition of the optimization problem of scheduling m-MAST service and provide the $\mathcal{NP}$-hardness through reduction from m-PDP. A mixed-integer program (MIP) of m-MAST is developed. Then we provide an analytical modeling framework to help MAST operators with their system planning and to identify the critical transit demand, which is used to decide when to switch from the one-vehicle MAST system to a two-vehicle MAST system. Utilizing this analytical model and the MIP formulation, we also compare the utility function values generated by the two methods for the one-vehicle MAST system and the two-vehicle MAST system. Experimental results show that for the same model input both the MIP formulation and the developed analytical model generate approximately the same utility function values (i.e., system performance) and critical demands. The reasonable match between the two sets of outputs demonstrates the validation and the effectiveness of the proposed MIP formulation and the analytical framework for critical demand modeling.

Since the MAST problem is $\mathcal{NP}$-hard, the proposed multi-vehicle MIP formulation can only optimally solve small to moderate-size problems (e.g, demand, number of checkpoints, etc.). Future work will include the development of proper valid inequalities/equalities and/or logic constraints to strengthen the proposed formulations and heuristic algorithms, allowing the formulated problems to be solved in real time or at large scales. It would also be interesting to extend the analytical modeling framework to consider different MAST configurations (e.g., three or more vehicles) and to identify the optimal fleet size as a function of demand.

The second major contribution of this research is that we develop an insertion heuristic for scheduling m-MAST service. The algorithm allows customers to place a request, and once accepted, provides them with time windows for both pickup and drop-off points. Due to the dynamic nature of the environment, the algorithm makes effective use of a set of control parameters to reduce the consumption of slack time and enhance the algorithm performance. Due to the existence of complicated time constraints and weighted objective function, we resort to experiments to evaluate the algorithm. The results of simulations show the efficacy of the algorithm and its optimal control parameter setting and demonstrate that the algorithm can be used as an effective method to schedule m-MAST service. By comparing the performance of 2-MAST and 1-MAST, the potential of m-MAST to provide a more attractable service and a better overall operation cost is shown. In addition, a comparison versus optimality values computed by *CPLEX*© in a static scenario shows that the results obtained by the heuristic are not far from optimum.

Although the experiments show the efficacy of the algorithm, it is highly likely that the performance is subject to the distribution of the demand. Thus theoretical analysis of worst-case performance of the proposed algorithm is of interest from the perspective of computer science and the perspective of engineering.

References

[1] D. Koffman, “Operational experiences with flexible transit services: a synthesis of transit practice,” TCRP Synthesis 53, Transportation Research Board, Washington DC, 2004.

[2] L. Quadrifoglio, M. M. Dessouky, and K. Palmer, “An insertion heuristic for scheduling mobility allowance shuttle transit (MAST) services,” *Journal of Scheduling*, vol. 10, no. 1, pp. 25–40, 2007.

[3] L. Quadrifoglio, R. W. Hall, and M. M. Dessouky, “Performance and design of mobility allowance shuttle transit services: Bounds on the maximum longitudinal velocity,” *Transportation Science*, vol. 40, no. 3, pp. 351–363, 2006.

[4] L. Quadrifoglio and M. M. Dessouky, “Sensitivity analyses over the service area for mobility allowance shuttle transit (MAST) services,” in *Computer-aided Systems in Public Transport*, vol. 600 of *Lecture Notes in Economics and Mathematical Systems*, pp. 419–432. Springer, Berlin, 2008.

[5] J. Zhao and M. Dessouky, “Service capacity design problems for mobility allowance shuttle transit systems,” *Transportation Research Part B: Methodological*, vol. 42, no. 2, pp. 135–146, 2008.

[6] M. W. P. Savelsbergh and M. Sol, “The general pickup and delivery problem,” *Transportation Science*, vol. 29, no. 1, pp. 17–29, 1995.

[7] T. R. Sexton and L. D. Bodin, “Optimizing single vehicle many-to-many operations with desired delivery times: I. Scheduling,” *Transportation Science*, vol. 19, no. 4, pp. 378–410, 1985.

[8] J.-F. Cordeau, “A branch-and-cut algorithm for the dial-a-ride problem,” *Operations Research*, vol. 54, no. 3, pp. 573–586, 2006.

[9] J.-F. Cordeau and G. Laporte, “The dial-a-ride problem (DARP): Variants, modeling issues and algorithms,” *4OR: A Quarterly Journal of Operations Research*, vol. 1, pp. 89–101, 2003.

[10] Q. Lu and M. M. Dessouky, “An exact algorithm for the multiple vehicle pickup and delivery problem,” *Transportation Science*, vol. 38, pp. 503–514, November 2004.

[11] C. E. Cortes, M. Matamala, and C. Contardo, “The pickup and delivery problem with transfers: Formulation and a branch-and-cut solution method,” *European Journal of Operational Research*, vol. 200, no. 3, pp. 711–724, 2010.

[12] S. Ropke and J.-F. Cordeau, “Branch and cut and price for the pickup and delivery problem with time windows,” *Transportation Science*, vol. 43, no. 3, pp. 267–286, 2009.

[13] G. Berbeglia, J.-F. Cordeau, and G. Laporte, “Dynamic pickup and delivery problems,” *European Journal of Operational Research*, vol. 202, no. 1, pp. 8–15, 2010.

[14] L. Quadrifoglio, M. M. Dessouky, and F. Ordez, “Mobility allowance shuttle transit (MAST) services: MIP formulation and strengthening with logic constraints,” *European Journal of Operational Research*, vol. 185, no. 2, pp. 481–494, 2008.

[15] H. N. Psaraftis, “A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem,” *Transportation Science*, vol. 14, no. 2, pp. 130–154, 1980.

[16] H. N. Psaraftis, “An exact algorithm for the single vehicle many-to-many dial-a-ride problem with time windows,” *Transportation Science*, vol. 17, no. 3, pp. 351–357, 1983.

[17] M. Fortini, A. Letchford, A. Lodi, and K. Wenger, “Computing compatible tours for the symmetric traveling salesman problem,” *Mathematical Programming Computation*, vol. 3, pp. 59–78, 2011.

[18] D. Teodorovic and G. Radivojevic, “A fuzzy logic approach to dynamic dial-a-ride problem,” *Fuzzy Sets and Systems*, vol. 116, no. 1, pp. 23–33, 2000.

[19] J. K. Lenstra and A. H. G. R. Kan, “Complexity of vehicle routing and scheduling problems,” *Networks*, vol. 11, no. 2, pp. 221–227, 1981.

[20] A. M. Campbell and M. Savelsbergh, "Efficient insertion heuristics for vehicle routing and scheduling problems," *Transportation Science*, vol. 38, no. 3, pp. 369–378, 2004.

[21] Q. Lu and M. M. Dessouky, "A new insertion-based construction heuristic for solving the pickup and delivery problem with time windows," *European Journal of Operational Research*, vol. 175, no. 2, pp. 672 – 687, 2006.

[22] N. Christofides, "Worst-case analysis of a new heuristic for the traveling salesman problem," Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA, 1976.

[23] D. J. Rosenkrantz, R. E. Stearns, and P. M. Lewis, II, "An analysis of several heuristics for the traveling salesman problem," *SIAM Journal on Computing*, vol. 6, no. 3, pp. 563–581, 1977.

[24] C. Archetti, L. Bertazzi, and M. G. Speranza, "Reoptimizing the traveling salesman problem," *Networks*, vol. 42, no. 3, pp. 154–159, 2003.

[25] S. Arora, "Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems," *J. ACM*, vol. 45, pp. 753–782, September 1998.

Appendix A DERIVATION OF EXPECTED RIDE TIME

All the notation used in this appendix is consistent with those in Chapter 3. In this appendix, the formulas for $E(T_{rd}^{PD})$, $E(T_{rd}^{PND/NPD})$ and $E(T_{rd}^{NPND})$ are derived. Due to the existence of the discretely located checkpoints, it is necessary to consider the possibility of traversing various checkpoints when formulating the expected values of performance measures. Two consecutive checkpoints (both in time and in space) are taken as a basic unit, then sum up the values of performance measures for different units and average them to obtain the expected value.

$E(T_{rd}^{PD})$ indicates the average ride time of all the possible pairs of pickup and drop-off checkpoints. Without loss of generality, assume the bus is progressing from left to right as depicted in Fig. A.1. Thus, a PD customer picked up at checkpoint j should have $C-j$ possible drop-off checkpoints and $C-j$ different ride times, which are: E_0^{PD},$E_0^{PD}+t$, ... , $E_0^{PD}+(C-1-j)t$ as shown in Fig. A.1. The expected ride time for this customer can be calculated as

$$\begin{aligned} E(T_{rd}^{PD}) &= \frac{\sum_{j=1}^{C-1}(C-j)[E_0^{PD}+(j-1)t]}{\sum_{j=1}^{C-1}(C-j)} \\ &= E_0^{PD} + \frac{t \cdot C(C-2)(C-1)/6}{C(C-1)/2} \\ &= E_0^{PD} + \frac{C-2}{3}t \end{aligned} \tag{A.1}$$

Similarly we can derive $E(T_{rd}^{PND/NPD})$, the only difference between $E(T_{rd}^{PND/NPD})$ and $E(T_{rd}^{PD})$ is the first term. By replacing the first term in $E(T_{rd}^{PD})$ (i.e., E_0^{PD}) with $E_0^{PND/NPD}$, the $E(T_{rd}^{PND/NPD})$ can be derived.

The derivation of $E(T_{rd}^{NPND})$ is different in some degree. From Fig. A.2, it can be seen that an NPND customer with his/her NP stop between checkpoints j and $j+1$ has $C-j$ possible units of service area for ND stop within one trip. Since the requests are uniformly distributed, there are

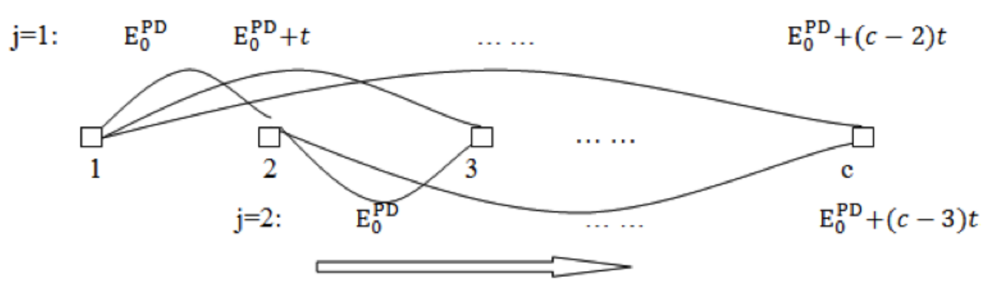

Figure A.1: Illustration for derivation of $E(T_{rd}^{PD})$

$C-j$ different expected ride times, which from left to right are: $E_0^{NPND}, t, 2t, \ldots, (C-1-j)t$. Considering the two directions in a two-trip cycle, we get the formulation of $E(T_{rd}^{NPND})$ as follows:

$$E(T_{rd}^{NPND}) = \frac{(C-1)E_0^{NPND} + 2\sum_{j=1}^{C-2}(C-1-j)\cdot j\cdot t}{(C-1)^2} = \frac{E_0^{NPND}}{C-1} + \frac{C(C-2)}{3(C-1)}t \tag{A.2}$$

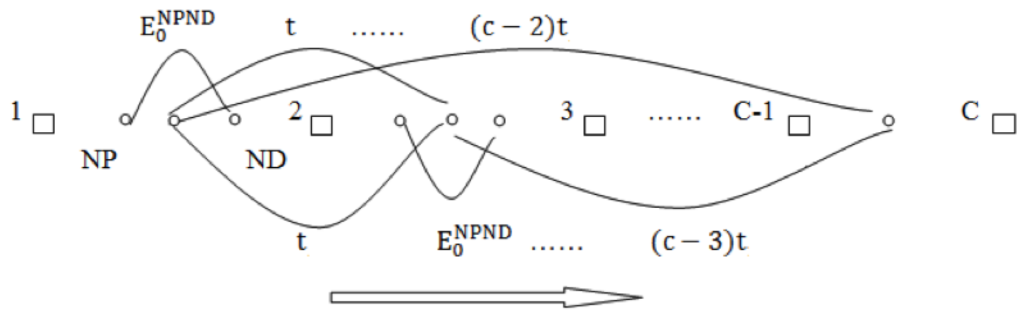

Figure A.2: Illustration for derivation of $E(T_{rd}^{NPND})$

University Transportation Center for Mobility™

Texas Transportation Institute

The Texas A&M University System

College Station, TX 77843-3135

Tel: 979.845.2538   Fax: 979.845.9761

utcm.tamu.edu

Texas Transportation Institute