

Service Capacity Design Problems for Mobility Allowance Shuttle Transit Systems

Jiamin Zhao

Oracle Corporation
Redwood Shores, CA 94065
jiamin.zhao.zhao@oracle.com

Maged Dessouky*

Daniel J. Epstein Department of Industrial and Systems Engineering
University of Southern California
Los Angeles, CA
90089-0193

maged@usc.edu

phone: 213-740-4891 fax: 213-740-1120

* Corresponding Author

Abstract

Transit agencies are currently seeking ways to improve service flexibility in a cost-efficient manner. Mobility Allowance Shuttle Transit is a transportation concept that merges the flexibility of Demand Responsive Transit systems with the low cost operability of fixed-route bus systems. In this paper, we study the service capacity design problem for such a system. We analyze the relationship between service cycle time, and the length and width of the service area under a desired service level constraint. The analysis can be used as a guide for designing service area parameters. Given that a shuttle travels in a rectilinear pattern and utilizes a non-backtracking nearest-insertion algorithm, we derive closed-form approximate solutions for the service capacity design problem. We show that setting the length of the service area to half the travel speed of the shuttle multiplied by the cycle time is an effective approximation.

Key Words: Transit shuttle, Optimization, Approximation

1 Introduction

Fixed-route bus transit systems are much more cost efficient than most demand responsive transit (DRT) systems (Palmer, Dessouky, and Abdelmaguid, 2004). This is primarily due to the passenger loading capacity of the buses and the consolidation of many passenger trips onto a single vehicle (ridesharing). However, as an alternative to private automobiles, fixed-route bus transit systems have a major deficiency. The general public considers the service to be inconvenient because either the locations of the pick-up and drop-off points do not, or the schedule of the service does not, match the individual rider's desires (lack of flexibility). Also, the total time for a trip is perceived as being greater than that of a private auto.

Commercial DRT systems, such as taxi cabs and shuttle vans, provide much of the desired flexibility. But, these improvements in convenience come at a price. Taxi cabs provide point-to-point pick-up and drop-off, and near real-time scheduling. However, the cost per trip is comparable to that of paratransit DRT (Palmer, Dessouky, and Abdelmaguid, 2004). Shuttle vans provide flexible pick-up points and a cost per trip less than that of a taxi cab. However, drop-off points are limited to popular locations, and often advanced scheduling is required. These restrictions on flexibility allow the shuttle vans to guarantee sufficient ridesharing to operate at their reduced cost per trip. Even so, shuttle vans are still less cost efficient than fixed-route bus transit systems.

Thus, there is a need for a transit system that provides flexible service at a cost efficient price. Mobility Allowance Shuttle Transit (MAST) is a transportation concept that merges the flexibility of Demand Responsive Transit (DRT) systems with the low cost operability of fixed-route bus systems. A MAST service has a fixed base route that covers a specific geographic zone. Shuttles are allowed to deviate from the fixed path to pick up and drop off passengers at their preferred locations. The only restriction on flexibility is that the deviations must lie within a predetermined distance from the fixed base route. By allowing short deviations from a base route, MAST provides point-to-point pick-up and drop-off within a certain geographic zone, and near real-time scheduling comparable to point-to-point transit systems.

The concept of MAST emerges from a real application. The Metropolitan Transit Authority (MTA) of Los Angeles County operates a feeder-line 646, which transports passengers between a large business hub in the San Pedro area of Los Angeles County and a nearby terminal. During daytime, this line serves as a fixed-route bus system. During nighttime, the line changes to a MAST service allowing specific deviations from the fixed route. Those passengers located within .5 miles off the route may call-in for pick-ups at their locations. In addition, customers who board the shuttle at a scheduled stop (e.g., terminal) can request a drop off location that is within .5 miles from the base route.

One critical issue in designing a MAST service is to determine the length of the base route and the allowed deviation from it, which together determine the service area. Since a MAST service typically operates as a feeder line, there is a strict time requirement for the shuttle to arrive at the terminal so that passengers have sufficient time to transfer to the main line service. Clearly, having a large service area provides the flexibility to pick-up and drop-off more passengers from the base route. However, a large service area reduces the probability that the shuttle arrives to the terminal on time. In this paper, we develop a model that can aid transit planners in determining the service area parameters that maximizes the service capacity meanwhile meeting the desired service level (the probability of arriving at the terminal on-time). We are unaware of any such models in the literature.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes a MAST shuttle system; in Section 4 we first derive approximations of the mean and variance of travel time, given that a shuttle travels in a rectilinear pattern and utilizes a non-backtracking nearest-insertion algorithm. Then, we obtain closed-form approximate solutions for the service capacity design problem; in Section 5 we briefly extend our discussion to multi-shuttle systems; Section 6 compares our approximate solutions with simulations; and Conclusions are presented in Section 7.

2 Literature Review

MAST systems have been recently studied by researchers. Quadrifoglio et al. (2006) develop bounds on the maximum longitudinal velocity to evaluate the performance of such systems, and Quadrifoglio et al. (2007b) developed an insertion algorithm to schedule a MAST system; in this paper the efficient use of control parameters significantly improved the performance of the algorithm. Malucelli et al. (1999) also approached the problem including it in a general overview of flexible transportation systems. Crainic et al. (2001) described the MAST concept and incorporated it in a more general network setting providing also a mathematical formulation.

The hybrid type of service that we are studying consists of the same vehicle performing the fixed and variable portions of the trip. There has been some work in studying hybrid systems in which different vehicles perform the fixed and variable portions. In the latter case, local service is provided by on-demand vehicles and line-haul service is provided by a fixed-route line. Passengers switch vehicles at a transfer station. Aldaihani et al. (2004) develop a continuous approximation model for designing such a service. Aldaihani and Dessouky (2003) develop a tabu search heuristic and test it on a data set from Antelope Valley in California. They show that shifting some of the demand to a hybrid service route (18.6% of the requests) reduces the on-demand vehicle distance by 16.6% without significantly increasing the trip times. Cortés and Jayakrishnan (2002) develop algorithms for transporting passengers door-to-door for a high demand coverage system. Their design is based on a system with high demand and restricts the number of transfers to one. Horn (2002a, 2002b) describes an embedded control system for the fleet scheduling and journey planning of a multi-modal transit system.

MAST has a close relation to DRT or Dial-a-Ride systems. Pioneering research on the dial-a-ride problem dates back to the seventies. Wilson et al. (1971) formulated the problem as a dynamic search procedure, inserting a newly arriving passenger's origin and destination into the prospective route of one of the shuttles. Since prediction of passengers' waiting and riding times is one of the key factors in their approach, continuing works were conducted by Daganzo (1978), and Wilson and Hendrickson (1980). Stein (1977, 1978) studied dial-a-ride systems from a probabilistic aspect and

derived an approach toward the design of asymptotic scheduling algorithms. Daganzo (1984) describes a checkpoint demand responsive system that combines the characteristics of both fixed route and door-to-door. In a checkpoint system, a service request is still made but the pick-up and drop-off points are not at the door but at centralized locations called checkpoints. Daganzo compares the riding time of the checkpoint system with the fixed route and door-to-door systems. His analysis found that fixed route systems perform best under high demand levels while the door-to-door performs best under low demand levels.

Although this paper mainly deals with the service area design problem, a solution itself must rely on some statistical features of routing algorithms. There have been numerous works on routing algorithms for dial-a-ride systems. Due to the combinatorial nature of the problem, most of the research efforts focus on heuristic approaches. Savelsbergh and Sol (1995) provide an excellent review of this work. More recent approximation solution procedures include the column generation algorithm by Savelsbergh and Sol (1998), the insertion heuristic by Madsen, Ravn, and Rygaard (1995), the clustering algorithm by Ioachim et al. (1995), the parallel insertion heuristic by Toth and Vigo (1997), the simulated annealing algorithm by Hart (1996), the tabu search heuristic by Cordeau and Laporte (2003), the regret insertion heuristic by Diana and Dessouky (2003), and the construction heuristic by Lu and Dessouky (2006). Although the meta-heuristics generate routes that are on average less costly than the insertion type heuristics, the most common algorithms used in practice are the insertion type heuristics due to their computational efficiency and ease of understanding (Hunsaker and Savelsbergh, 2002; Campbell and Savelsbergh 2004). Furthermore these types of heuristics tend to perform well in a dynamic environment.

3 MAST System Configuration

Figure 1 represents a general geographical configuration of MAST systems. Shuttles run from a terminal toward another one and then return to the original terminal. Shuttles can deviate from the base route within a predetermined service area. Usually, schedules are setup at terminals. If a shuttle arrives at a terminal earlier, it has to wait

until the scheduled departure time; otherwise, it is dispatched immediately after loading and unloading passengers.

[Figure 1]

We assume that a *non-backtracking nearest-insertion* algorithm is used for the dynamic routing and scheduling process. That is, a shuttle always runs from one terminal toward another and serves the request that has the nearest horizontal distance ahead it. The shuttle travels in a *rectilinear* pattern (i.e., it runs horizontally first, then vertically to the request). This can be explained by Figure 2. At time t , there are three requests, a , b and c ; a and b are ahead of the shuttle, while c has just missed the shuttle. The shuttle will not go back to pick up c ; instead, it runs first horizontally then vertically towards a . Suppose request d occurs right after the shuttle serves a , the algorithm reschedules the route as shown in the solid line instead of the one in the dashed line that the algorithm originally scheduled. We also assume that requests are uniformly distributed within the service area, i.e., ρ is a constant; and the shuttle travels at a constant speed, v .

[Figure 2]

Clearly, a nearest-insertion algorithm is a suboptimal routing policy. It is a greedy scheduling strategy that attempts to find a local minimum solution. Using a global optimal search algorithm would make the problem of identifying a closed form analytical equation to the design problem computational intractable; the intent of this paper is not to develop a detailed routing algorithm for MAST systems (see for example Quadrioglio et al. 2007a and 2007b for detailed work in this area), but to develop closed form equations to solve the problem of determining the service area before the routes are generated. Assuming the nearest-insertion algorithm significantly simplifies the analysis and facilitates the development of the closed form analytical models. Furthermore, as outlined in Campbell and Savelsbergh (2004), this type of a simple greedy heuristic is commonly used in practice due their ability to perform well in a dynamic setting whereas the

preplanned globally optimal solutions tend to perform poorly, and the greedy heuristics are computationally efficient which is especially important in a dynamic environment.

4 Service Capacity Design Problems

Before presenting the service design problem, we study the dynamics of the travel and delay distributions for such a system. Theoretical results from the literature and their derivations will help us to obtain the optimal solution for the service design problem. We first consider a single-shuttle system. Capacity of the shuttle is assumed to be infinite. Let L be the length of the base route and $W/2$ the maximum allowed deviation from the base route. Let T and T_r be the scheduled and actual round-trip travel times, respectively. T is a constant (for a single-shuttle system, $T = SH$, the scheduled headway); T_r is a random variable with density function $g(t|L, W, \rho, SH, A)$, where ρ represents demand arrival rate, SH scheduled headway, and A the routing and scheduling algorithm. At a terminal, let $D > 0$ denote the delay time of a dispatch, which is the difference between the actual departure time and the scheduled departure time if the shuttle departs after the scheduled time. Define *service level* as $SL = P(D = 0)$, the probability that a dispatch is on time.

According to Lindley (1952), we have the following lemma:

Lemma 1

If $E(T_r) < T$, then the system is stable, which means the delay distribution converges, and the limiting distribution can be represented by the following relationship

$$D \sim \max(D + T_r - T, 0) \quad (1)$$

where \sim represents that the two random variables have a same distribution function.

Note that round-trip travel times are not really independent because a larger one (especially when it causes a delay) may cause a larger headway and more demands for the next service cycle. This increases the expected round-trip travel time of the consecutive loop. However, if a high service level is maintained, i.e., $P(D = 0)$ is close to 1, it is reasonable to assume that the round-trip travel times are independent because

the variation in travel times is sufficiently compensated by slacks in the schedule. Equation (1) can also be written in a form as

$$P(D \leq t) = P(D + T_R \leq T + t) = \int_0^\infty P(D \leq \tau) \cdot g(T + t - \tau) d\tau \quad (2)$$

Let $F(\cdot)$ be the distribution function of D , then Equation (2) can be expressed as

$$F(t) = \int_0^\infty F(\tau) g(t + T - \tau) d\tau \quad (3)$$

Equation (3) is a Wiener-Hopf integral equation. Explicit solution techniques can be found in related references (Zhao, Dessouky, and Bukkapatnam, 2006; Davies, 1985; Hochstadt, 1973; Gohberg and Feldman, 1974). With a general kernel $g(\cdot)$, it is computationally difficult to solve the equation. Lindly (1952) explained in detail how to solve such an equation with $g(\cdot)$ belonging to a particular family

$\left\{ g_n(t) = \frac{\lambda^{n+1}}{n!} t^n e^{-\lambda t} \right\}_{n=0}^\infty$ (the family of n th order convolution of the exponential distribution).

Assume that $g(\cdot)$ has the following form:

$$g(x) = \lambda e^{-\lambda(x-x_0)} \quad (x \geq x_0) \quad (4)$$

Equation (3) has an analytical solution as

$$F(t) = 1 + \eta \exp(\mu t) \quad (5)$$

where μ is the non-trivial root of equation

$$\frac{\lambda}{\lambda + \mu} = \exp[\mu(x_0 - T)] \quad (6)$$

and η is a constant satisfying

$$1 + \frac{\lambda \eta}{\lambda + \mu} = 0 \quad (7)$$

The probability of on-time returns can be expressed as

$$F(0) = 1 + \eta = -\frac{\mu}{\lambda} \quad (8)$$

Given required service level SL , $F(0) = SL$; combining with Equation (6), we have

$$\lambda(T - x_0) = \alpha \quad (9)$$

where $\alpha = -\frac{\ln(1-SL)}{SL} > 1$ ($0 < SL \leq 1$). The following table lists values of α under different SL .

[Table 1]

Assuming a uniformly distributed service area and a non-backtracking nearest insertion algorithm, given an actual headway H and N be the number of requests during H , then we have

$$E(T_R|H, N) = \begin{cases} \frac{L}{v} + \frac{NW}{3v} + \frac{W}{6v}, & N > 0 \\ \frac{L}{v}, & N = 0 \end{cases} \quad (10)$$

Assume that $P(N=0)$ is very close to 0, then

$$E(T_R) \approx \frac{L}{v} \left(1 + \frac{1}{3} \rho T W^2 \right) + \frac{W}{6v} \quad (11)$$

By further assuming that $P(N \leq 1)$ is very close to 0, we can derive that (see Appendix)

$$\text{var}(T_R) \approx \frac{8\rho L T W^3}{45v^2} \quad (12)$$

Note that, by making the above approximation, we have actually ignored the impact of variation of headways, or in other words, the correlation between two consecutive roundtrip travel times. As stated before, when the desired service level is high, it is safe to ignore the impact of the correlation between two consecutive roundtrip travel times. We will show this in our simulation experiments.

With the approximations of the mean and variance of travel time on hand, we will next solve the service capacity design problems. To obtain closed-form approximations, we assume that $g(\cdot)$, the probability density function of travel time T_R , has the form of Equation (4), i.e., a shifted exponential distribution function. λ and x_0 can be expressed by

$$\lambda = \text{var}(T_R)^{-\frac{1}{2}} \text{ and } x_0 = E(T_R) - \frac{1}{\lambda} \quad (13)$$

By replacing λ and x_0 in Equation (9), we have

$$(\alpha - 1)\sqrt{\text{var}(T_R)} = T - E(T_R) \quad (14)$$

Depending on what parameters among (L, W, T) are given, there are different forms of the service capacity design problem. We will solve two example problems in this section.

Problem 1: Given L and T , determine W

By combining Equations (11), (12), and (14), we have the following equation:

$$a^2W^2 + \sqrt{\frac{8}{15}}(\alpha - 1)aW^{\frac{3}{2}} + \frac{1}{6}W - (vT - L) = 0 \quad (15)$$

where $a = \sqrt{\rho LT/6}$. The equation has a single real positive root if and only if $vT - L > 0$ (let $F(W) = a^2W^2 + \sqrt{\frac{8}{15}}(\alpha - 1)aW^{3/2} + \frac{1}{6}W$, notice that $F'(W) > 0$ when $W > 0$). This result is intuitive since when $L > vT$, a shuttle can never finish a trip within T even if $W = 0$.

When $SL \rightarrow 1$, $\alpha \rightarrow \infty$ and $W \rightarrow 0$. As expected, the results show that the width of the service area is reduced as the desired service level increases, thus limiting the amount of off-route deviations and the chance of arriving to the terminal late. When $SL \rightarrow 0$, $\alpha \rightarrow 1$ and W converges to the positive root of equation $\frac{1}{3}\rho TLW^2 + \frac{1}{6}W - (vT - L) = 0$, which means that W is bounded by $E(T_R) < T$, the stable condition of the system, according to Lemma 1.

Problem 2: Given T , determine L and W simultaneously

The optimization problem can be formulated as:

$$\begin{aligned} \max \Delta &= \rho TLW \\ \text{st } (\alpha - 1)\sqrt{\text{var}(T_R)} &= T - E(T_R) \end{aligned}$$

where $E(T_R)$ and $\text{var}(T_R)$ are expressed by Equations (11) and (12), respectively.

By replacing $E(T_R)$ and $\text{var}(T_R)$, the constraint can be rewritten as

$$\frac{\rho T}{3}(LW)^2 + \sqrt{\frac{8}{45}}(\alpha - 1)\sqrt{\rho T}(LW)^{\frac{3}{2}} + \frac{LW}{6} = L(vT - L) \quad (16)$$

Note that the left-hand side of the constraint is a monotonously increasing function of LW . To maximize LW , we must maximize the right-hand side of the constraint. This yields,

$$L^* = \frac{vT}{2} \tag{17}$$

Then W^* can be obtained by solving the previous problem since T and L^* are already known.

The simple expression of L^* is interesting indeed. It shows that given a service cycle T , the optimal length of the service area is half of the length that the shuttle can travel along a straight line during T . It does not depend on the required service level at all.

Note that the solution $L = vT/2$ is obtained under the assumption that demands are uniformly distributed within the service area. Now suppose that demands are centralized around the base route, the optimal length of the service area would be longer than $vT/2$ and the shape of the service area would be much slimmer because the capacity gain by lengthening the service area would be more significant than by widening the service area. If demand density decreases with distance from the terminal, then the result is just the opposite. The optimal length of service area would be shorter than $vT/2$ and the shape of the service area would look stouter.

5 Discussion on Multi Route and Shuttle Systems

When there are multiple routes around a terminal, a request near the terminal may have multiple choices of being served by different routes (see Figure 3). The purpose of this section is to present some initial findings for multi-route systems and motivate future work in this area.

[insert Figure 3]

There are four routes around the terminal. Areas *I*, *II*, *III*, and *IV* are service areas. Each area falls within the service two separate routes, as hown by the dashed lines in the figure. Thus, area *I* is within the service area of route 1 and of route 2. Requests in area *I*

that can be served by Route 1 fall into two categories: a) a request with pickup point in area I but drop-off point outside area I , or the opposite; b) a request with pickup and drop-off points both in area I (usually one at the terminal). Requests in the first category do not have choices. They can only be served by Route 1; requests in the second category may have multiple choices. They can be served by either Route 1 or Route 2. Let ρ_1 and ρ_2 be densities of request of the first and second types, respectively. Suppose a request in the second category has p chance to be assigned to Route 1 and $(1-p)$ chance to be assigned to Route 2. Then, the overall request density for Route 1 in area I equals $\rho_1 + p\rho_2$. In this way, we have decoupled the multiple routes around a terminal. The only issue is that we have to cope with non-uniformly distributed requests.

We next discuss a case of the single route with multiple shuttles. When there are k ($k > 1$) shuttles in a system, the average headway is reduced to

$$E(H) = \frac{T}{k} \quad (18)$$

By doing slight modification on (11) and (12), we have

$$E(T_R, k) \approx \frac{L}{v} \left(1 + \frac{\rho T W^2}{3k} \right) + \frac{(m+1)W}{6v} \quad (19)$$

$$\text{var}(T_R, k) \approx \frac{8\rho L T W^3}{45k v^2} \quad (20)$$

Notice that the modification does not change the conclusion represented by Equation (17), i.e., given T , $L^* = \frac{vT}{2}$.

With multiple shuttles allowed, we are able to solve the following problem:

Given L and W , determine the minimum number of shuttles

According to Lemma 1, the system is stable if and only if

$$E(T_R, k) < T, \text{ or equivalently,}$$

$$\left(v - \frac{\rho L W^2}{3k} \right) T > L + \frac{W}{6} \quad (21)$$

To make the inequality feasible, we must have

$$v - \frac{\rho LW^2}{3k} > 0 \text{ or } k > \frac{\rho LW^2}{3v} \quad (22)$$

6 Simulations

In this section we use simulations to evaluate our approximate solutions. In the simulation system, demands arise in real-time. A non-backtracking nearest-insertion algorithm is applied for dynamic routing. The cycle time $T = 60$ min and the vehicle speed $v = 0.5$ mile/min. Each simulation consists of 30 runs and each run is for 4000 minutes.

Scenario 1: Evaluate approximations of mean and variance of travel time

First we evaluate our approximations of the mean and variance of travel time, i.e., Equations (11) and (12). The following table summarizes the simulation outputs compared to the approximation results.

[Table 2]

The comparison shows that our approximations are quite close to the simulation results for all the scenarios listed in the table (varying L , W , and ρ). The table shows that the variance of the travel time is greatly affected by the width of the service area supporting the need for smaller service widths when the desired service level increases. Note that the approximation underestimates the variance of travel time because it ignores the variance in the headway effect (the underestimation is more significant when the variance of headway becomes larger). The column under SL in the table shows the fraction of time the bus returns to the terminal on time. Thus, when the desired service level is high ($> 90\%$), the variance in headways can be ignored.

You may also note that $E(T_R)$ in simulation is always larger than that in our approximation. Our approximation is theoretically accurate enough; however, due to our simulation implementation (we always round up the number of request when we simulate

Poisson distribution), $E(T_R)$ in the simulation tends to be larger. Indeed, the upper bound of the error can be derived as $\frac{W}{3v}$.

Scenario 2: Given L , find the maximum allowed W

Our simulation system takes the value of W from the approximate solution as an initial point and performs a line search until we find a W such that the average achieved service level over a small domain around W is within an allowed error from the target service level. The target service level is set to 90%.

The simulation results show that our approximate solution is accurate; the relative error, calculated by $(\text{Approx.} - \text{Sim.})/\text{Sim.} * 100\%$, is less than or around 1%. Note that the error terms are slightly higher for the smaller demand density case because the approximation ignores the variance in headway and its effect is greater with smaller demand density (when demand density is smaller, the width of service area is bigger and the variance of the headway is larger).

[Table 3]

Scenario 3: Given T , find the optimal L and W to maximize the service capacity

Intuitively, the maximum allowed service capacity is a concave function of L . Though not a strict mathematical proof, Equation (16) does provide an insight on the concaveness. This means that we can use simulations to find the optimal solution by a line search procedure. Setting $\rho = 0.04/\text{mil}^2 \cdot \text{min}$, we run simulations with $SL = 0.9$ and $SL = 0.95$, respectively. Service capacities under different L , obtained through simulations, are plotted in Figure 4. In the same figure we also draw capacity curves from our approximation solutions to compare with the simulation results.

[Figure 4]

Simulation results have verified that the maximum service capacity is achieved when L is around $T/2$. Notice that, intuitively, the service capacity from the simulations should be less than that from the approximations since we underestimate the mean and

variance of the travel time. However, the simulations with $SL = 0.95$ display the contrary behavior. This is because our approximate solution is based on the assumption that the travel time is exponentially distributed. This assumption may offset the effect of underestimates on the mean and variance of travel time. This may also help explain why the difference between the approximate and simulations solutions is greater when SL increases.

7 Conclusions

This paper studies the service capacity design problem for MAST systems. We first utilize a result from Theory of Queues to introduce a stable condition for the system dynamics and represent the delay distribution by a Wiener-Hopf integral equation. Then we derive approximations of the mean and variance of travel time under the assumption that a shuttle travels in a rectilinear pattern and uses a non-backtracking nearest-insertion algorithm for real-time routing and scheduling. Closed-form solutions are obtained by using an exponential distribution to simplify the general travel time distribution. Finally, we verify our solutions through simulations.

Some results are interesting and insightful. For example, the optimal length of the service area, given a service cycle time, is half of the distance that a shuttle can travel along a straight line during the cycle period. Another helpful insight given by the approximations is the concaveness of the curve of service capacity in terms of the length of the service area. The simple relationships developed in this study can be used by policy planners to help determine the service areas for given desired service levels.

Future study may include dropping or relaxing some assumptions in this paper, for instance, requests are uniformly distributed within the service area. A more clustered demand pattern may suggest developing special rules that do not allow deviations from the base route in some areas. Another possible extension is to develop similar bounds assuming headway based control versus the schedule-based control used in this paper. The headway based control systems tend to perform better when the schedule headways are short. Although we presented some preliminary results, a more accurate and detailed analysis on the design problem for multi-shuttle systems is required. This analysis could

also try to identify the best mix of number of shuttles for the system and service area. Finally, an interesting avenue of further research is to study the mixture of a MAST system with a checkpoint system as proposed by Daganzo (1984).

Acknowledgements

The research reported in this paper was partially supported by the National Science Foundation under grant NSF/USDOT- 0231665.

Appendix

Derivation of Variance of Travel Time, $\text{var}(T_R)$

Given H , a actual headway, and N , the number of requests during H ($N > 0$), the total travel distance, TL , can be expressed as

$$TL(H, N) = L + \sum_{n=1}^{N-1} |X_{n+1} - X_n| + |X_1| + |X_N| \quad (\text{A1})$$

where X_i ($i = 1, \dots, N$) are i.i.d. variables that are uniformly distributed in $(-W/2, W/2)$. Then

$$E(TL|H, N) = L + \frac{1}{3}(N-1)W + \frac{W}{2} = L + \frac{1}{3}NW + \frac{W}{6}, \quad N > 0 \quad (\text{A2})$$

$$\begin{aligned} \text{var}(TL|H, N) &= \sum_{n=1}^{N-1} \text{var}|X_{n+1} - X_n| + 2 \text{var}|X_1| \\ &+ 2 \sum_{n=1}^{N-2} \text{cov}(|X_{n+1} - X_n|, |X_{n+2} - X_{n+1}|) + 4 \text{cov}(|X_1|, |X_2 - X_1|) \end{aligned} \quad (\text{A3})$$

Note that to make the above formulas valid, we need one more assumption. $P(N \leq 1)$ is very close to zero. Thus it can be ignored.

It can be derived that

$$\text{var}(|X_{n+1} - X_n|) = \frac{1}{18}W^2, \quad \text{var}(|X_1|) = \frac{1}{48}W^2 \quad (\text{A4})$$

$$\text{cov}(|X_{n+1} - X_n|, |X_{n+2} - X_{n+1}|) = \frac{1}{180}W^2, \quad \text{and} \quad \text{cov}(|X_1|, |X_2 - X_1|) = \frac{1}{96}W^2 \quad (\text{A5})$$

Therefore,

$$\text{var}(TL|H, N) = \left(\frac{N}{15} + \frac{1}{180} \right) W^2 \quad (\text{A6})$$

Using the formula for conditional variance, we have

$$\begin{aligned} \text{var}(TL|H) &= E(\text{var}(TL|H, N)) + \text{var}(E(TL|H, N)) \\ &= \left(\frac{\rho HLW}{15} + \frac{1}{180} \right) W^2 + \frac{1}{9} \rho HLW^3 = \frac{8}{45} \rho HLW^3 + \frac{1}{180} W^2 \end{aligned} \quad (\text{A7})$$

Similarly,

$$\begin{aligned} \text{var}(TL) &= E(\text{var}(TL|H)) + \text{var}(E(TL|H)) \\ &= \frac{8}{45} \rho TLW^3 + \frac{1}{180} W^2 + \frac{64}{2025} \rho^2 L^2 W^6 \text{var}(H) \end{aligned} \quad (\text{A8})$$

To compute the $\text{var}(TL)$, we need to compute the variance of headway, $\text{var}(H)$. Note that $\text{var}(H)$ itself depends on $\text{var}(TL)$. The two can be computed recursively. Here we simply assume that with a high service level, $\text{var}(H)$ can be neglected. The item $\frac{1}{180} W^2$ is also small enough to be ignored. Therefore,

$$\text{var}(TL) \approx \frac{8}{45} \rho TLW^3 \quad \text{and} \quad \text{var}(T_R) \approx \frac{8\rho TLW^3}{45v^2} \quad (\text{A9})$$

References

- Aldaihani, M.M., Quadrioglio, L., Dessouky, M.M. and Hall, R.W (2004) “Network Design for a Grid Hybrid Transit Service”, *Transportation Research – Part A*, 38A, 511-530.
- Aldaihani, M.M. and Dessouky, M.M. (2003) “Hybrid Scheduling Methods for Paratransit Operations”, *Computers and Industrial Engineering*, 45, 75-96.
- A. M. Campbell and M. Savelsbergh (2004), “Efficient Insertion Heuristics for Vehicle Routing and Scheduling Problems,” *Transportation Science*, 38, 369-378.
- Cordeau, J. F. and Laporte, G. (2003) “A Tabu Search Heuristic for the Static Multi-vehicle Dial-a-ride Problem”, *Transportation Research - Part B*, 37B, 579-594.

- Cortés, C.E. and Jayakrishnan, R. (2002) “Design and Operational Concepts of a High Coverage Point-to-point Transit System”, *Transportation Research Record*, 1783, 178-187.
- Crainic, T.G., Malucelli, F. and Nonato, M. (2001) “Flexible Many-to-few + Few-to-many = an almost Personalized Transit System”, TRISTAN IV, São Miguel Azores Islands, 435-440.
- Daganzo, C. F. (1978) “An Approximate Analytic Model of Many-to-Many Demand Responsive Transportation Systems”, *Transportation Research*, 12, 325-333.
- Daganzo, C. F. (1984) “Checkpoint Dial-a-Ride Systems”, *Transportation Research*, 18B, 315-327.
- Daganzo, C. F. (1984) “The Length of Tours in Zones of Different Shapes”, *Transportation Research – Part B*, 18B, 135-145.
- Diana, M. and Dessouky, M. (2004) “A New Regret Insertion Heuristic for Solving Large-scale Dial-a-ride problems with Time Windows,” *Transportation Research - Part B*, 38B, 539-557.
- Hart, S. M. (1996) “The Modeling and Solution of a Class of Dial-a-Ride Problems Using Simulated Annealing”, *Control and Cybernetics*, 25, 131-157.
- Horn, M.E.T. (2002a) “Multi Modal and Demand Responsive Passenger Transport Systems: a Modeling Framework with Embedded Control Systems”, *Transportation Research – Part A*, 36A, 167-188.
- Horn, M.E.T. (2002b), “Fleet Scheduling and Dispatching for Demand Responsive Passenger Services”, *Transportation Research – Part C*, 10C, 35-63.
- B. Hunsaker and M. Savelsbergh (2002), “Efficient Feasibility Testing for Dial-a-Ride Problems,” *Operations Research Letters*, 30, 169-173.
- Ioachim, I., Desrosiers, J., Dumas, Y., Solomon, M., and Villeneuve, D. (1995), “A Request Clustering Algorithm for Door-to-door Handicapped ‘Transportation’”, *Transportation Science*, 29, 63-78.
- Jaw, J.-J., Odoni, A. R., Psaraftis, H. N., and Wilson, N. H. M. (1986) “A Heuristic Algorithm for the Multi-Vehicle Advance Request Dial-a-Ride Problem with Time Windows”, *Transportation Research – Part B*, 20B, 243-257.

- Lindley, D. V. (1952) "The Theory of Queues with a Single Server", *Proceedings of Cambridge Philosophy Society*, 48, 277-289.
- Lu, Q. and Dessouky, M. (2004) "An Exact Algorithm for the Multiple Vehicle Pickup and Delivery Problem", *Transportation Science*, 38, 503-514.
- Lu, Q. and Dessouky M. (2006) "New Insertion-based Construction Heuristic for Solving the Pickup and Delivery Problem with Hard Time Windows," *European Journal of Operational Research*, 175, 672-687.
- Madsen, O. B. G., Ravn, H. F., and Rygaard, J. M. (1995) "A Heuristic Algorithm for a Dial-a-Ride Problem with Time Windows, Multiple Capacities, and Multiple Objectives", *Annals of Operations Research*, 60, 193-208.
- Malucelli, F., Nonato, M. and Pallottino, S. (1999) "Demand Adaptive Systems: Some Proposals on Flexible Transit", in *Operations Research in Industry*, T. Cirania, E. Johnson, and R. Tadei (eds), 157-182.
- Palmer, K., Dessouky, M. M., and Abdelmaguid, T. (2004) "Impacts of Management Practices and Advanced Technologies on Demand Responsive Transit Systems," *Transportation Research, Part A: Policy and Practice*, 38, 495-509.
- Quadrifoglio, L., Dessouky, M.M. and Hall, R.W (2006) "Performance and Design of Mobility Allowance Shuttle Transit (MAST) Services: Bounds on the Maximum Longitudinal Velocity," *Transportation Science*, 40, 351-363.
- Quadrifoglio, L., Dessouky, M. M., and Ordóñez, F. (2007a) "Mobility Allowance Shuttle Transit (MAST) Services: MIP Formulation and Strengthening with Logic Constraints," to appear *European Journal of Operational Research*.
- Quadrifoglio, L., Dessouky, M.M. and Palmer, K. (2007b) "An Insertion Heuristic for Scheduling Mobility Allowance Shuttle Transit (MAST) Services", *Journal of Scheduling* 10, 25-40, 2007.
- Savelsbergh, M. W. P, and Sol, M. (1995) "The General Pickup and Delivery Problem," *Transportation Science*, 29, 17-29.
- Savelsbergh, M. W. P, and Sol, M (1998), "Drive: Dynamic Routing of Independent Vehicles," *Operations Research*, 46, 474-490.

- Stein, D. M. (1977) "Scheduling Dial-a-Ride Transportation Systems: An Asymptotic Approach", Harvard University, Division of Applied Science, Technical Report No. 670.
- Stein, D. M. (1978) "Scheduling Dial-a-Ride Transportation Systems", *Transportation Science*, 12, 232-249.
- Toth, P., and Vigo, D. (1997) "Heuristic Algorithm for the Handicapped Persons Transportation Problem," *Transportation Science*, 31, 60-71.
- Wilson, N. H. M., Sussman J. M., Wong Ho-Kwan, and Higonnet, T. (1971) "Scheduling Algorithms for a Dial-a-Ride System", M.I.T, Urban Systems Laboratory, Technical Report USL TR-70-13.
- Wilson, N. H. M., and Hendrickson, C. (1980) "Performance Models of Flexibly Routed Transportation Services", *Transportation Research – Part B*, 14B, 67-7.
- Zhao, J., Dessouky, M. M., and Bukkapatnam, S. (2006) "Optimal Slack Time for Schedule-Based Transit Operations," *Transportation Science*, 40, 529-539.

Tables

Table 1 α under different SL

SL	0.7	0.75	0.8	0.85	0.9	0.95	0.99
α	1.72	1.85	2.01	2.23	2.56	3.15	4.65

Table 2 Comparing simulation and approximation on mean and variance of travel time

L (mil)	$\rho = 0.01/\text{mil}^2 \cdot \text{min}, W = 2 \text{ mil}$						$\rho = 0.04/\text{mil}^2 \cdot \text{min}, W = 1 \text{ mil}$					
	Simulation				Approximation		Simulation				Approximation	
	$E(T_R)$	$\text{Var}(T_R)$	$\text{var}(H)$	SL	$E(T_R)$	$\text{var}(T_R)$	$E(T_R)$	$\text{var}(T_R)$	$\text{var}(H)$	SL	$E(T_R)$	$\text{var}(T_R)$
10	37.86	33.66	0.00	0.9995	36.67	34.13	36.89	17.17	0.0	1.0	36.33	17.07
11	41.55	41.42	0.188	0.997	40.27	37.55	40.49	18.30	0.0	1.0	39.93	18.77
12	44.94	41.33	0.629	0.981	43.87	40.96	44.19	19.61	0.004	0.9995	43.53	20.48
13	48.91	45.61	3.04	0.933	47.47	44.37	47.65	21.99	0.134	0.991	47.13	22.19
14	52.36	50.60	8.42	0.811	51.07	47.79	51.37	23.63	1.00	0.944	50.73	23.89
15	55.87	54.84	23.08	0.573	54.67	51.20	55.09	27.47	5.78	0.767	54.33	25.60
16	59.87	66.61	56.04	0.162	58.27	54.61	58.61	30.06	19.51	0.348	57.93	27.31

Table 3 Given L , determine maximum W

L (mil)	W (mil)					
	$\rho = 0.01/\text{mil}^2 \cdot \text{min}$			$\rho = 0.04/\text{mil}^2 \cdot \text{min}$		
	Sim.	Approx.	Err.	Sim.	Approx.	Err.
10	2.52	2.55	1.2%	1.355	1.355	0.00%
11	2.35	2.38	1.3%	1.257	1.262	0.40%
12	2.20	2.22	0.91%	1.177	1.177	0.00%
13	2.06	2.07	0.49%	1.099	1.100	0.09%
14	1.93	1.94	0.52%	1.027	1.029	0.19%
15	1.79	1.81	1.1%	0.963	0.962	0.10%
16	1.68	1.70	1.2%	0.898	0.900	0.22%

Figures

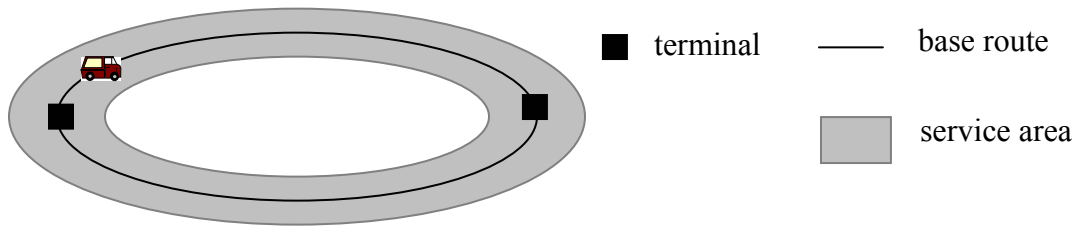


Figure 1 General Representation of MAST Systems

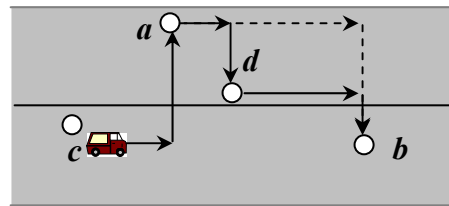


Figure 2 Non-Backtracking Nearest Insertion Algorithm

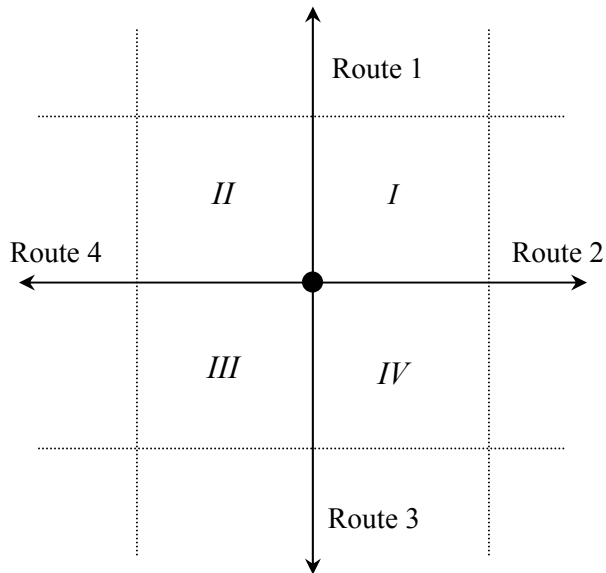


Figure 3 Multiple Route Example

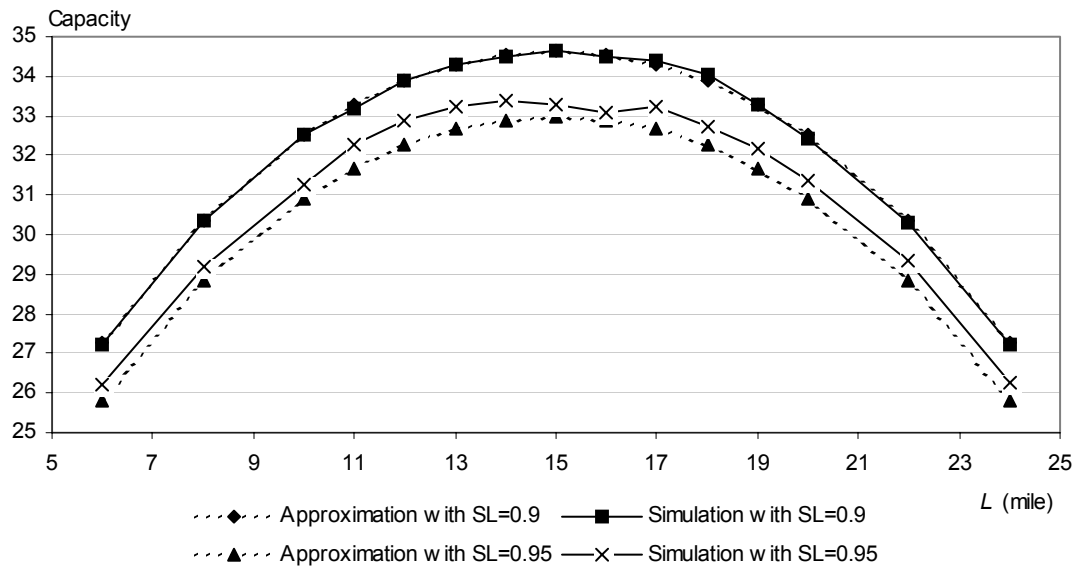


Figure 4 Comparison of approximate solution and simulation result on service capacity