Estimating the Optimal Cycle Length for Feeder Transit Services[*]

Shailesh Chandra and Luca Quadrifoglio[†]

Zachry Department of Civil Engineering, Texas A&M University,
College Station, TX 77843-3136

ABSTRACT

The general lack of first/last mile connectivity is one of the main challenges faced by today's public transit. One of the possible actions towards a solution to this problem is the planning, design and implementation of efficient feeder transit services. This paper develops an analytical model which allows for an easy computation of near optimal terminal-to-terminal cycle length of a demand responsive feeder service to maximize service quality provided to customers, defined as the inverse of a weighted sum of waiting and riding times. The model estimates the recommended cycle length by only plugging in geometrical parameters and demand data, without relying on extensive simulation analyses or rule of thumbs. Simulation experiments validate our model, which would allow planners, decision makers and practitioners to quickly identify the best feeder transit operating design of any given residential area.

Keywords: Optimal cycle, demand responsive, feeder transit, transit performance, continuous approximations, first/last mile

1. INTRODUCTION AND BACKGROUND

The US Department of Transportation recently identified the general lack of connectivity as one of the main challenges faced by public transit. Policies which encourage the desired reduction of Vehicle Miles Traveled (VMT), reduction of greenhouse gases and even an increase of "livability" depend on solutions to the issue of first/last mile access to transit and multi-modal connectivity. One of the possible actions towards providing a solution to this problem is the planning, design and implementation of efficient Demand Responsive Feeder Transit services, connecting residential areas to major fixed-route transit networks, also known as Demand Responsive Connectors (DRC).

Demand Responsive Transit (DRT) systems, also known as dial-a-ride transit (DART) or call-n-ride (CnR) systems, have been proven effective in responding to the need of low demand density areas and are welcomed by passengers as they provide an increased flexibility and higher service level compared to 'regular' fixed-route services. These services are, however, much more costly to deploy for the operators. Typical examples, other than feeder services, include ADA Paratransit services (Diana and Dessouky, 2004), rural services (Quadrifoglio et al., 2009) or other flexible hybrid systems, like "route deviation" services.

Mathematical formulations are often carried out for optimizing and managing fleet size for demand responsive or dial-a-ride systems (Horn , 2002). Single vehicle

---

[*] Shortened version of a paper recently accepted for publication in *Transportation Research – Part B*

[*] Corresponding author. Phone: (979) 458-4171, Fax: (979) 862-8498, Email: quadrifo@tamu.edu

study for dial-a-ride can be found in the work of Psaraftis (1980) as an exact dynamic programming solution. The single vehicle case deals with mostly minimizing route duration, ride time, and waiting time of the passengers. Chandra et al. (2011) studied a simulation based approach using a single vehicle case to estimate optimal cycle length for the feeder services within a residential community. These kinds of studies, however, require considerable data input a priori that often are not available in a new area.

Literatures can be found that deal with a single cycle (or headway) optimization for performance of operating a 'fixed' feeder transit services (Chowdhury and Chien, 2011). The same for flexible transit systems, such as DRT, is still not properly addressed due to complex service request times and demand uncertainty. Analytical modeling, simulation and continuous approximations have been used to analyze flexible transit services. Cortés and Jayakrishnan (2002) provide good examples and reviews on these type of research approach. Research specifically on decision tool for the design of feeder transit services is somewhat limited.

Serving passengers with a shuttle in demand responsive operations can be treated as a problem similar to those found in queuing theory, where the shuttles can be considered as dynamic service windows, passengers as service objects, and dispatch time from service windows as waiting times (Daganzo, 1990). However, the influence of service area characteristics cannot be easily integrated with performance in these studies. This paper fits in this category by proposing a model to estimate the most appropriate cycle length to operate the service given the demand and the shape of the service area.

Better planning, design and operation of these services may provide a potential solution to the first/last mile problem, which ultimately contribute towards greater goals like reducing traffic, mitigating related pollution and congestion problems and ultimately increase the overall "livability". What stops this from happening is the lack of existing research and information for transit managers in improving the state-of-the-art practice in operating these flexible transit services, especially related to the 'problems with scheduling' (Potts et al., 2010). In this paper, we are focusing our research efforts to partially fill the above gap. More specifically, we propose a model to describe the relationship between the level of service provided to passengers and the terminal-to-terminal cycle length (one of the decision variables of the service), allowing for a manageable computation of its optimal value that could otherwise be estimated only using extensive simulation analyses or by trial-and-error or rule-of-thumb procedures, as often done in practice. Several sets of simulation experiments and some comparison with existing services appear to validate our model, which would allow planners, decision makers and practitioners to quickly identify a good estimate of the best feeder transit operating design within a given service area under given circumstances.

2. METHODOLOGY

3.1 Service area, demand and optimal cycle

We are considering a generic residential area whose shape can be approximated with a rectangle with length L and width W, served by a single shuttle. The terminal (designated as D) with coordinates (0, 0) is located in the middle of an edge (at $W/2$). The shuttle is departing from D at constant time intervals (cycle) C to serve passengers, which can request to go from D to the service area ("drop-off" customers) or vice versa ("pick-up" customers). Demand within the area is assumed to be spatially uniformly distributed as well as temporally uniformly distributed (Poisson process) within a target time interval T of the day. This demand assumption is reasonable if at the terminal there are enough transit lines with high frequency and/or if the terminal is also a trip production/attraction site, such as a shopping mall or similar.

3.2 Existence of optimal cycle and performance definition

In a given time slot of the day T, the shuttle performs a number of cycles to serve the given demand. The longer the cycle length, the more ride sharing will be used, but passengers will experience longer average riding/waiting time, so shorter cycle lengths are intuitively more desirable for customers. However, the shorter the cycle length, the more the cycles, the more *extra* driving will be needed to go to/from the terminal and fewer customers can be served. At the limit, one customer at a time can be served, not taking advantage of any ride sharing. With negligible demand, this taxi-like operating practice would be plausible and best for customers; but with high enough demand, the service would be oversaturated and queues and spillovers will more likely occur, average waiting time will increase, thus lowering the level of service. These two combined effects cause an optimal cycle length to exist with the right amount of ride-sharing (minimizing disutility U, thus maximizing the average level of service), at which the system will be able to optimally serve all customers and that we aim to find in this paper.

It is generally difficult to identify a unique definition of performance for a transit system as priorities differ among stakeholders. Several authors have used measures such as passenger cost, passengers per vehicle hour, vehicle miles per operator, cost per vehicle mile, cost per vehicle hour, the ratio of cost to fare box revenue and fleet fuel efficiency for the urban public transit (Gleason and Barnum, 1986; Badami and Haider, 2007). However, all seem to agree that transit performance can generally be identified as a combination of operating costs and service quality. In our model, the service quality is expressed as passengers' disutility (cost): a weighted sum of expected waiting time and the expected in-vehicle travel time of passengers (there is no walking time for door-to-door services). Thus, the disutility (cost) $U = \gamma_1 w_t + \gamma_2 r_t$ (w_t = expected waiting time and r_t = expected riding time; γ_1 and γ_2 are the weight factors for the waiting time and riding time respectively, to be properly calibrated). This is used to identify the optimal C to maximize serving quality. Within each operating cycle C, the operations are then conducted so that total distance is minimized (like a Traveling Salesman Problem) to minimize operating costs.

The proposed methodology consists of building an analytical queuing model for estimating the optimal cycle length in two steps. The first step consists of identifying equations linking a given cycle length C and the average number of passengers n that can be served at most by the shuttle within C; the second part consists of building the model of our objective function (disutility) U as a function of C to find the optimum.

3.3 $C(n)$ models

We are summarizing four models, found in the literature, which can be potentially used to describe the relationship between C and n. Note that these models provide an approximation of the optimal C needed to serve n passengers by minimizing the total distance, just as in a Travelling Salesman Problem. These routes are devoid of any consideration for the waiting time or riding time of the individual passengers. We call these relationships between average *C* and *n* as designs and address them as Design (I) to (III).

Design (I). *Nearest-neighbor approach*: The average closest distance between two uniformly distributed passenger demands, as developed by Quadrifoglio et al. (2006) for a very high demand density ρ, can be used for this approach. This model provides a good approximation for the optimal *C* and *n* relationship with very high demand density. After adapting the model to our case by including a needed newly developed mathematical derivation, the relationship can be written as (see Appendix for details),

$$C = \left(\frac{1.15}{V\sqrt{\rho}}\right) + \frac{0.63n}{V\sqrt{\rho}} + (n+1)t_s \tag{1}$$

where, t_s is the time taken at each stop or terminal to pick-up/drop-off of a passenger, and *V* is the average speed of the feeder bus. For a very high demand density ρ, the expression in (1) can be approximated as,

$$C \approx \frac{0.63\sqrt{n}\sqrt{LW}}{V} + (n+1)t_s \tag{2}$$

Design (II). *Approximate TSP solution approach*: Results from work of Beardwood et al. (1959) and Jaillet (1988) can also be used to express another approximate relationship between *C* and very large *n*, with

$$C = \frac{\sqrt{nLW}}{V} + (n+1)t_s \tag{3}$$

Design (III). *No-backtracking approach*: Exploiting the scheduling guideline as proposed by Quadrifoglio and Li (2009), a strategy can be set for serving passengers for a lower passenger demand (low *n*) and a high length-to-width ratio of service area (see Fig. 1). In this case, the vehicle would move through the upper half of the region in a no-backtracking policy left-to-right, and move through the bottom half in a no-backtracking policy right-to-left. The relationship between *C* and *n* is expressed in this case by the following,

$$C = \frac{2L\left(\frac{n}{n+1}\right) + \frac{2W}{3} + \frac{Wn}{6}}{V} + (n+1)t_s \tag{4}$$

With the further assumption $\left(\frac{n}{n+1}\right) \approx 1$, the relationship between C and n in (4) becomes linear and can be expressed as

$$C \approx \frac{2L + \frac{2W}{3} + \frac{Wn}{6}}{V} + (n+1)t_s \qquad (5)$$

Using (5) for computing C does not change its values significantly with respect to (4), especially for higher values of n (Quadrifoglio and Li, 2009). This model also closely estimates the optimal cycle length and performs better for high L/W ratios.

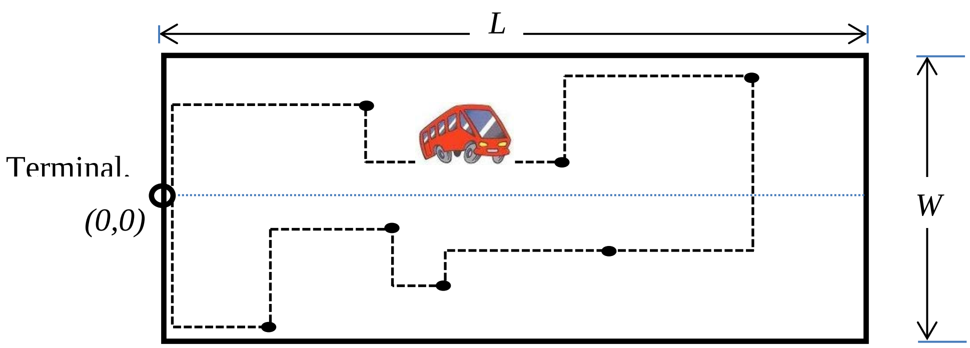

Figure 1: Shuttle pick-up/drop-off strategy.

3.4 Model Comparison

It is important to compare the relative applications of each of the design methodologies discussed above to select the best C and n for a single cycle relationship according to the actual service area and demand. The four relationships are compared to outputs obtained from an insertion heuristic policy (Jaw et al., 1986; Quadrifoglio et al., 2007) and optimality (for lower demand). This comparison for outputs is made with respect to the total travel time, which is C, of the DRT shuttle for a single service cycle. Insertion heuristic can be used for good solutions as compared to optimality obtained using optimization software. In fact, as the demand becomes higher, it is quite impractical to compute optimal solutions, simply because of unreasonable computational times involved in the process (Li and Quadrifoglio, 2010). Besides, insertion heuristic is widely used in practice for most scheduling problems.

Cycle length values were computed for three different shapes of a one square mile service area: Case 1 is a square with $L = W = 1$ mile; Case 2 is a rectangle with $L = 2$ miles and $W = 0.5$ miles; Case 3 is a slimmer rectangle with $L = 3$ miles and $W = 0.33$ miles. An average of 1000 replications were performed in *MATLAB R2010b* for each value of C and the values of n vary between 1 to 40 to cover any variability in demand within each cycle. The average feeder speed was assumed to be *20* mile per hour, posted speed limit found in most residential areas where shuttles operate, and the value for average time t_s spent at a passenger stop was assumed to be *30* seconds. Optimal cycle lengths were obtained using a TSP code in MATLAB and optimal computed using CPLEX 12.1. Following the outputs, a close match between the insertion heuristic and C_s were found. We thus used insertion heuristic as benchmark for all the three design methodologies.

Design method (III), *No-Backtrack*, matched well with insertion heuristic (V) cycle lengths for higher L/W and lower n, which is the more commonly found shape for real residential areas. The cycle length values from design method (II) were closer to insertion heuristic outputs for the square shaped area for almost all values of n. The values for C using design methods (I) and (II) matched more closely with insertion heuristic values as n increased as compared to other design methods. The match was further improved as the demand density became higher (as expected). Also (I) and (II) showed identical cycle length values for all three cases for the same n, as they depend on the total area (LW, which is the same for all) and not on the shape (identified by the L/W ratio).

These results set the stage for building our analytical model and choosing the most appropriate $C(n)$ expression depending on the conditions, as they are useful in estimating cycle lengths for any given service area geometry and demand.

3.5 Feeder Operations

The feeder shuttle is assumed to operate daily for a fixed duration of time. We focus on a portion of the day (T hours), during which $\left[\frac{T}{C}\right] = [\tau]$ cycles ([·] stands for the nearest integer value of ·) of constant duration C will be performed to serve a total of N passengers. Passenger requests for a given cycle are collected during the previous cycle and considered in a FIFO fashion. The above situation is explained through sets of slots (each of duration C) corresponding to each of the dispatch times of the feeder bus in the sketch of Fig. 2.

The variables t_1 and T_1 are the start time of the passenger request times and the feeder bus service times, respectively. Similarly, t_{k+1} and T_{k+1} are the end times of the passenger request times and the feeder bus service times, respectively, for a total of k slots in a day formed equal to $[\tau]$. Note that $(T_1 - t_1) = C$ (and $(T_{k+1} - t_{k+1}) = C$) with $(t_{k+1} - t_1) = T$ (and $(T_{k+1} - T_1) = T$). However, in case of passenger spillovers carried from the previous cycles, the shuttle operates an extra cycle to serve all at end time = T_{k+2}.

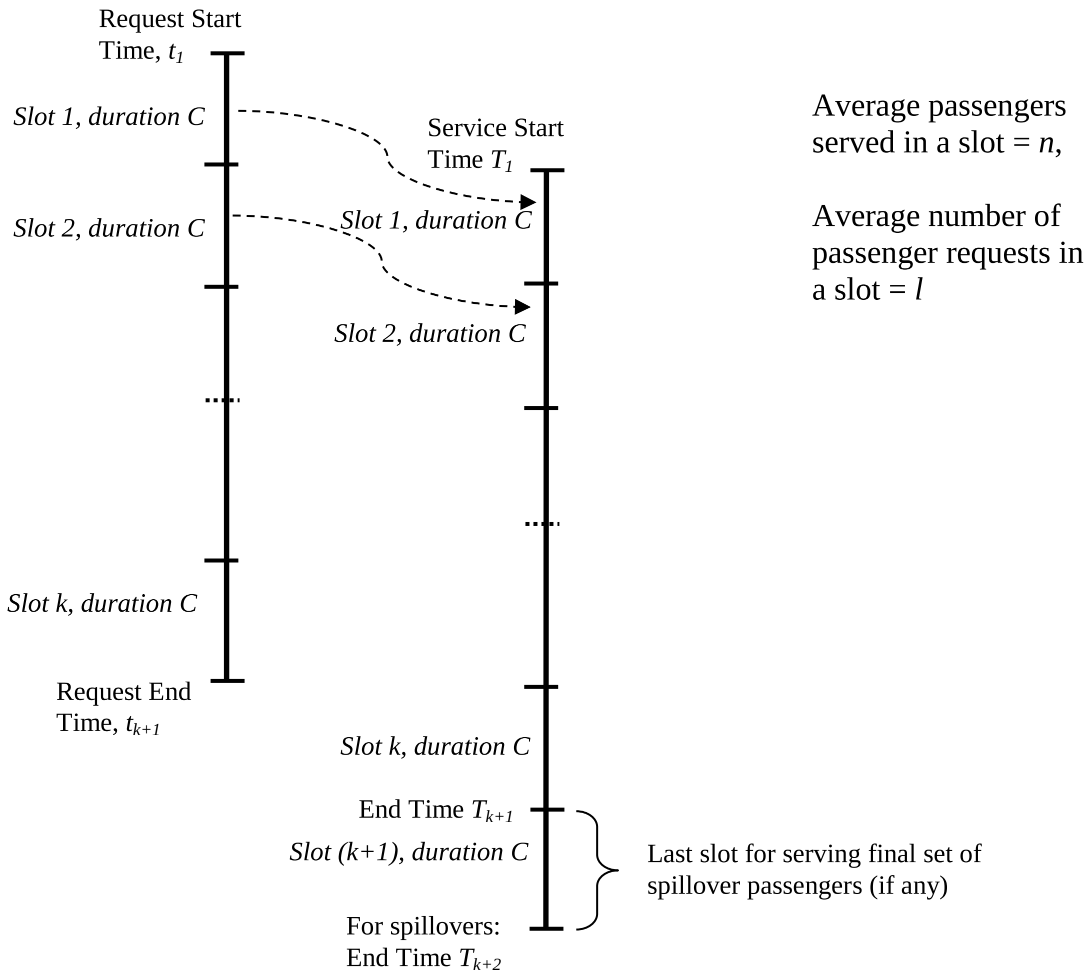

Figure 2: Illustration of passenger requests and service times using slots of duration *C*.

3.6 Service Disutility

In this section we develop a continuous approximation model assuming a deterministic process with constant (average) time intervals between service requests and compare them to simulated realistic Poisson process. There are $\frac{NC}{T} = l$ requests/cycle on average. We know that n is the numbers of passengers which can be served at most on average in a cycle C (as estimated by the models illustrated in the previous section 3.3). If $l \leq n$, the system is under saturation and can comfortably serve all demand in every cycle. We assume that this collected demand (l) is scheduled for service in a given cycle following a near optimal sequence, generated by an insertion heuristic algorithm (this scheduling procedure is the same adopted by SuperShuttle (2012). If instead $l > n$, the system is oversaturated and the (average) residual (l-n) passengers would need to be served in the next cycle, causing a stable queue to build up in the system and additional waiting time to exist.

The disutility (cost) experienced by passengers $U = \gamma_1 w_t + \gamma_2 r_t$ can be modeled as

$$U = \begin{cases} Q_1 & if \ \ l > n \\ Q_2 & if \ \ l \leq n \end{cases} \tag{6}$$

We aim to model the above cases as a function of any demand level and L/W geometric dimension of the service area and the decision variable C, to derive its optimal. Derivation of w_t and r_t in Q_1 and Q_2 involves careful accounting for passenger requests served in a cycle. We define α as the fraction of pick-up passengers (going to service area to terminal).

For Q_1, the average riding time will be easily computed for all passengers as $r_t = C/2$. The waiting time is more complicated. 1-α drop-off passengers (going from terminal to service area) need to wait an average of $C/2$ time (from their show-up time to the beginning of the shuttle ride from the terminal); α pick-up passengers need to wait an average of time C ($C/2$ from their show-up time to the beginning of the shuttle ride from the terminal and an additional $C/2$ till their pick-up time). Thus, average waiting time for all customers served would be $(1+\alpha)C/2$ (see also Quadrifoglio and Li (2009)). However, only n passengers per cycle can be served at their requested cycle; the remaining l-n passengers (both pick-up and drop-off ones) in each cycle will have to be served in the following cycle and there will also be an additive spillover effect. Thus, l-n passengers in the first demand cycle will be served in the second service cycle and would need to wait for an additional average time of C. In the second demand cycle, $2(l\text{-}n)$ customers will be served in the third service cycle (since additional l-n customers will be pushed back by the first service cycle's spillover). In general, at demand cycle k, $k(l\text{-}n)$ customers will be served in service cycle k+1. Thus, the additional waiting time to be considered will be $C\{[(l\text{-}n) + 2(l\text{-}n) + \ldots + \tau(l\text{-}n)]/N\} = C[\tau(1+\tau)/2](l\text{-}n)/N$, considering only $\tau = [\tau]$. Eventually, the overall average waiting time will be $w_t = (1+\alpha)C/2 + C[\tau(1+\tau)/2](l\text{-}n)/N$. Note that the queuing effect we are trying to capture and model is significant, but small enough that we can safely assume that there will not be situations in which customers are pushed back more than one extra cycle within T, since this would not be realistic, as the system would be excessively oversaturated and would need more shuttles, not just an operational fix. In particular, this assumption would be acceptable when $\tau\left(l-n\right) < n$, which ensures that for the number of spilled-over passengers from the last slot τ are less than n and thus all served in the very next cycle. By substituting $\tau = T/C$ and $l = NC/T$, we need to have

$$n > \frac{N}{\left(1+\frac{T}{C}\right)} \tag{7}$$

For Q_2, when demand $l < n$, the shuttle would take an average time $t < C$ to complete every cycle (and wait a slack time C-t at the terminal before the next service cycle). An estimate of t that can be computed using the design methodologies (I), (II), and (III) discussed earlier, plugging in l instead of n and taking the resulting C value as t. Average riding time for all passengers is $r_t = t/2$. The waiting time for drop-off passengers is the same as for Q_1; for pick-up passengers it is $C/2 + t/2$, where the second part takes in account the actual travel time $t < C$ at each cycle. There won't be any stable queue formation in this Q_2 case, so no extra waiting time.

In summary,

$$Q_1 = \gamma_1 C\left(\frac{1+\alpha}{2} + \frac{\tau(1+\tau)}{2}\frac{(l-n)}{N}\right) + \gamma_2\left(\frac{C}{2}\right) \tag{8}$$

$$Q_2 = \gamma_1\left(\frac{C}{2} + \frac{\alpha t}{2}\right) + \gamma_2\left(\frac{t}{2}\right) = \gamma_1\left(\frac{C}{2}\right) + (\alpha\gamma_1 + \gamma_2)\left(\frac{t}{2}\right) \tag{9}$$

As mentioned earlier (section 3.2), the existence of an optimal C to minimize U is intuitive; we are attempting to find it analytically and verify our results.

Keep in mind that n is a function of C and can be obtained by using the reverse expressions from methods (I) – (III) by using approximate expressions wherever necessary. Skipping the easy but cumbersome mathematical passages, we can write $n \approx \left(hC + f\sqrt{C} + g\right)$ with the values of h, f and g listed in Tab. 1.

Table 1: Expressions for h and g

Design Method	h	f	g
(I)	$\frac{1}{t_s}$	$-\sqrt{\left(\frac{0.63LW}{t_s^3V^2}\right)}$	$-1+\frac{0.31LW}{t_s^2V^2}$
(II)	$\frac{1}{t_s}$	$-\sqrt{\left(\frac{LW}{t_s^3V^2}\right)}$	$-1+\frac{0.5LW}{t_s^2V^2}$
(III)	$\frac{1}{\left(\frac{W}{6V}+t_s\right)}$	0	$-\left(\frac{12L+4W+6Vt_s}{W+6Vt_s}\right)$

It is difficult to work with non-linear expressions (due to square root of C involved) if design methods used are (I) and (II) as $f \neq 0$. This also makes it hard to derive closed form expressions as roots of cubic polynomials are not trivial to estimate. Thus, we proceed to give some closed form results only for design method (III) when $f = 0$ and only report optimal cycle lengths obtained graphically for the other two methods along with simulation results later. From now on, unless specified, assume we are only working with design method (III). Thus, writing Q_1 in (8) for $f = 0$, $l = NC/T$ and $\tau = T/C$ we get

$$Q_1 = \frac{\gamma_1}{2}\left((2+\alpha)C - \frac{(hC+g)T}{N} + T - \frac{(hC+g)T^2}{NC}\right) + \gamma_2\left(\frac{C}{2}\right) \tag{10}$$

Q_1 is monotonically decreasing with increasing C and representing the system till the first intersection with Q_2 (9), which is a monotonically increasing linear function of C. This intersection represents the point at which the spillover condition stops and is also the estimate for the optimal C, which we aim to find. The equality $Q_1 = Q_2$ would be a quadratic Equation of the following form:

$$\left(\frac{u}{C}+vC+r\right)=\left(mC+b\right) \tag{11}$$

where, $u=\frac{\gamma_1}{2}\left(-\frac{gT^2}{N}\right), v=\frac{\gamma_1}{2}\left(\left(2+\alpha\right)-\frac{hT}{N}\right)+\left(\frac{\gamma_2}{2}\right), r=\frac{\gamma_1}{2}\left(T-\frac{hT^2}{N}-\frac{gT}{N}\right)$, with h and g as defined in Tab. 1 for method (III).

Depending on the relationship between C and n from design methods (III), the expressions for m and b are as shown in Tab. 2.

Table 2: Expression for right hand side coefficients of Equation (11)

Design Method	m	b
(III)	$\frac{\gamma_1}{2}+\frac{\left(\alpha\gamma_1+\gamma_2\right)}{2}\left(\frac{NW}{6VT}+\frac{Nt_s}{T}\right)$	$\frac{\left(\alpha\gamma_1+\gamma_2\right)}{2}\left(\frac{6L+2W}{3V}+t_s\right)$

Solving (11) gives the smallest value of cycle length which is C optimal (C_{opt}),

$$C_{opt}=\frac{\left(b-r\right)-\sqrt{\left(b-r\right)^2-4u\left(v-m\right)}}{2\left(v-m\right)} \tag{12}$$

with all variables defined earlier. The other root is to be discarded as it would have a higher disutility, could also very well occur beyond realistic capacity of the shuttle, but primarily Q_1 would simply not hold anymore (no more spillover).

For (12) to hold true, it is essential that $\left(b-r\right)^2-4u\left(v-m\right)\geq 0$ (say, Constraint 1) and $\left(v-m\right)>0$ (say, Constraint 2). If Constraint 1 is violated it would mean that there is no real intersection point for the spillover and the no-spillover curves. Therefore, C_{opt} cannot be obtained by the intersection of spillover and no-spillover case. Moreover, this situation would arise when the demands are much higher than sustainable causing a large number of spillovers at every cycle, violating our model assumptions. If Constraint 2 is violated then it would result in a negative value of C_{opt} which is again impractical. However, this situation would never arise for any DRT system since this would occur at very high variable demand (for a very high N in m to make (v-m) non-positive).

3. SIMULATION EXPERIMENTS

The purpose of this section is to validate the derived analytical expressions for optimal cycle length estimation. Disutility values are obtained for different cycle lengths with three different assumptions of dimensions for rectangular service area, denoted by Case 1 (L = 1 mile, W = 1 mile), Case 2 (L = 2 mile, W = 0.5 mile) and Case 3 (L = 3 mile, W = 0.33 mile)

Many residential areas can be approximated by square and rectangular shapes with dimensions shown in the three cases. The simulation is performed by coding the Feeder operations in MATLAB R2010b. The shortest street based path between any two demand points is computed using the Dijkstra's algorithm using the rectilinear distance between the points. The rectilinear distance is also a good approximation for street based distances between two demand points (Quadrifoglio and Li, 2009). This is particularly true for residential areas, most of which have a grid-like rectilinear street pattern, especially the ones using feeders. There were 20 replications used for each simulation.

The passenger demands used are N = 50, 80, 100 and 240 for T = 4hr of operations (assumed to be within peak-hours). The first three demands fall within those that are found in practice from several call-n-ride systems (Potts et al., 2010).

We perform two sets of simulations: one (*Simu1*) with realistic demand distribution taken from the report by Santos et al. (2011). We focus on the morning peak hour distribution periods of 5 am to 9 am for the *Total* distribution. The other one (*Simu2*) is performed with the above N considered constant over T to closely match the assumption of our analytical model.

For simplification, the *Simu1* trip data for the commuters are converted to cumulative distribution functions (CDF) as shown in Fig. 3. The actual CDF (a polynomial of higher degree and hence, difficult to invert) is slightly modified into an easily invertible piecewise linear function for random generation of travel times for passengers. In simulation, this is achieved by generating random real numbers between 0.1 and 1, and simply computing the corresponding travel request times using the linear function. A linear CDF is obtained that expresses the uniform random generation of passengers between 5 am to 9 am (*Simu2*).

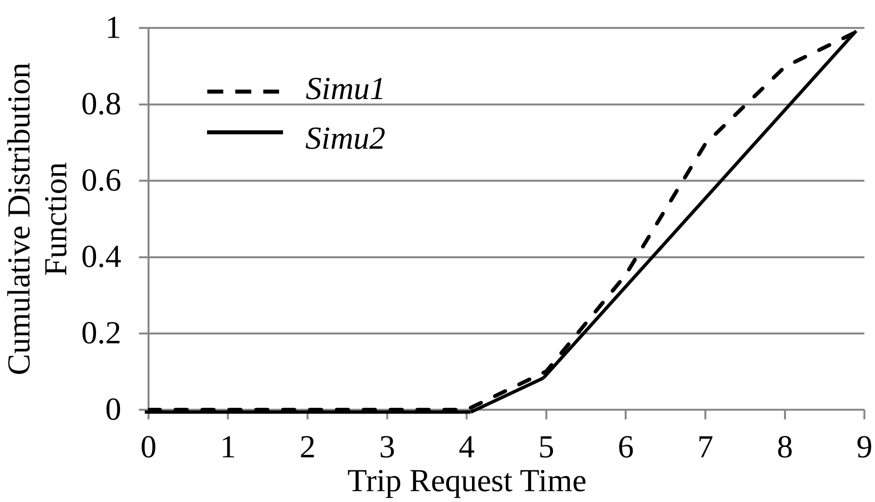

Figure 3: CDF representing trip travel time for commuting (Source: Santos et al., 2011).

Note that the simulated morning peak-hour period will reasonably consist of mostly pick-up passengers ($\alpha \cong 1$). A similar analysis can be performed for the afternoon peak hours with $\alpha \cong 0$. However, our model allows for any kind of combination, just by choosing appropriately the value of α included in the analytical modeling earlier.

The collected demand at each cycle is scheduled with insertion heuristic using up all the cycle time C. Customers are considered in a FIFO fashion and simulated as pick-up or drop-off depending on a random number generated depending on α. Any customer not been able to be served in a cycle will be considered first in the next cycle.
The shuttle is assumed to travel at constant speed V, leave the terminal at the beginning of each cycle, stop a time t_s at each requested location following the generated schedule, and travel back to the terminal.

Other simulation settings used are outlined in Tab. 3.

Table 3: Input settings for the simulation model

Parameter	**Input Numerical Values**		
(L, W)	Case 1: *(1,1)*	Case 2: *(2, 0.5)*	Case 3: *(3, 0.33)*
V	*20* miles per hour (i.e. the average speed in a residential area)		
(γ_1 / γ_2)	*1.8* (Wardman , 2004)		
Fleet size	*1* (A single feeder shuttle)		
Shuttle Capacity	*Infinite* (or large enough) to accommodate all passengers within a given cycle.		
Dwell Time (t_s)	*30* seconds (at the depot as well as at the passenger pick-up/drop-off location)		
Cycle Length C	Minimum: $C_m = (2L+W)/V + 2t_s$ (time needed to reach the farthest point of the service area from the terminal and back in a rectilinear fashion) Maximum : Long enough to capture lowest disutility for each simulation case.		

3.1 Simulation Results and Discussion

Results are shown in Fig. 4-6. We are pleased to observe a very close match of our queuing model Equations (8), (9) and (12) to simulation results, validating our approach and allowing us to adopt the analytical model for design purposes. We are also pleased to observe a close overlap between *Simu1* and *Simu2*. This is particularly important to show that assuming constant demand over T as opposed to more realistic distribution is not a very strong assumption.

Three of the curves are used in the charts of Fig. 4–6 for disutility from three different design methodologies adopted for relationship between C and n: *Nearest Neighbor* uses method (I), *Approximate TSP* uses method (II) and *No-backtracking* uses method (III). Fig. 4–6 only show results for passenger demand = 100 with other optimal cycles and minimum disutility values for other demand levels being tabulated in Tab 4a-c. The term C_m stands for the minimum cycle length needed (see Tab. 3).

All three models provide a good visual match for the no-spillover part (right) of the simulation curves. The *No-backtracking* method visually gives a very close match to the spillover part (left) too, overall providing a good approximation of the behavior of the system in all situations. As expected, curves obtained by the model underestimates the simulation one around the minimum, as delays and temporary queues are formed due to randomness to worsen the level of service, but not captured by our model based on expected values and continuous approximations. More sophisticated queuing models might be more appropriate for close matches of the results, but they would certainly lose in practicality and simplicity, needed to more easily make planning decisions.

The column $|\Delta U|$% in Tab. 4a-c stands for the absolute percentage error calculated on the *Simu2* curve caused by using C_{opt} recommended by the analytical model vs. the actual optimal of the curve. The expression for $|\Delta U|$% is calculated using ($|\Delta U| \times 100 / U_{Simu2}$; see Fig. 6 as an example demonstration for *No-backtracking* design). The best C (C_{opt}) obtained from the theoretical curves are selected (**bold**) based on their $|\Delta U|$% values. The lower the $|\Delta U|$%, the more preferred a particular design method is. Too high values of $|\Delta U|$% are reported as '>100' indicating values higher than 100 in Tab. 4 - it is an indication that the corresponding design model is not accurate and others are preferred. The error terms (expressed as *Err* C_{opt} %) in the tables are similarly calculated with respect to the C_{opt} on *Simu2* but along the C axis. Since we are aiming to identify the best estimate of C_{opt} in order to minimize the disutility (U), it is appropriate to select model performance based on $|\Delta U|$%.

Results clearly show that the *No-backtracking* design method (III) gives the best results for lower demands (50, 80 and 100), but deviates for higher demand (240), where *Approximate TSP* curves provide the closest match. Since Feeder services generally operate at lower demand levels, we can recommend using no-backtracking model (III) for most practical situations. Better performance for (III) is also seen for slimmer service areas (Fig. 8 and Tab 4c) as opposed to square ones (Fig. 6 and Tab 4a) to further validate the concept mentioned earlier (Section 3.3) that the *No-backtracking* design method (III) is more suitable for high length-to-width ratio of the service area.

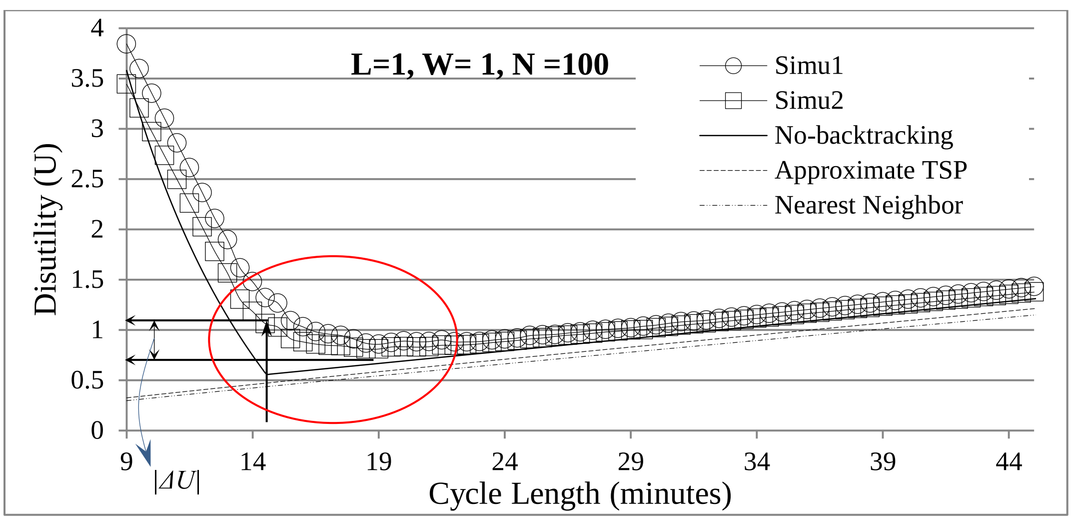

Figure 6: Case 1 disutility versus C for passenger demand N = 100.

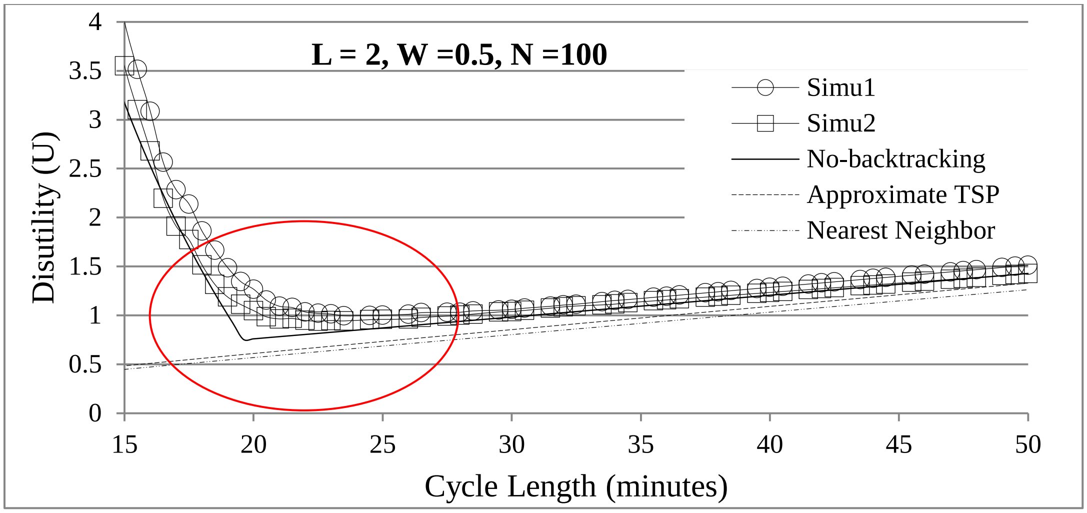

Figure 7: Case 2 disutility versus *C* for passenger demand N = 100.

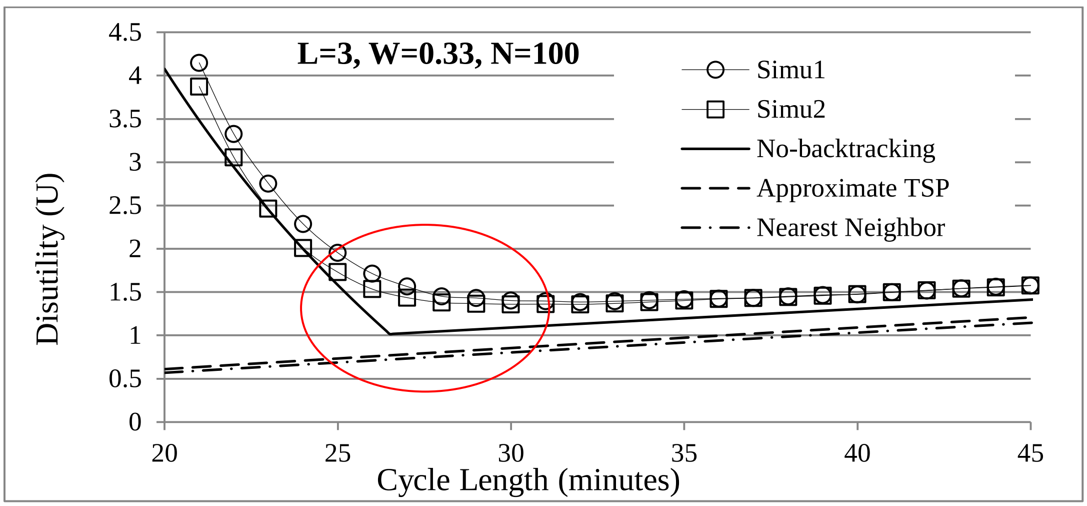

Figure 8: Case 3 disutility versus *C* for passenger demand N = 100.

Table 4a: Optimal cycle lengths (C_{opt} in minutes) for Case 1 with different demands

N	Case 2 : L = 1 mile, W = 1 mile												
	Simu1		*Simu2*		Design Method Used								
					I			**II**			**III**		
	U	C_{opt}	U	C_{opt}	$^*C_{opt}$	$\|\Delta U\|$ %	*Err* C_{opt} (%)	C_{opt}	$\|\Delta U\|$ (%)	*Err* C_{opt} (%)	C_{opt}	$\|\Delta U\|$ %	*Err* C_{opt} (%)
50	0.71	14-15	0.70	13-17	C_m	>100	31	C_m	>100	31	**10.7**	4	17
80	0.81	16-18	0.80	15-19	C_m	>100	40	C_m	>100	40	**12.7**	21	15
100	0.88	19-21	0.82	17-20	C_m	>100	47	C_m	>100	47	**14.5**	67	15
240	1.45	36-37	1.40	35-37	24.4	>100	30	**37.6**	8	16	28.3	>100	19

**All C_m values are 9 min.*

Table 4b: Optimal cycle lengths (C_{opt} in minutes) for Case 2 with different demands

N	Case 1: L = 2 mile, W = 0.5 mile												
	Simu1		*Simu2*		Design Method Used								
					I			II			III		
	U	C_{opt}	U	C_{opt}	$*C_{opt}$	$\|\Delta U\|$ %	*Err* C_{opt} *(%)*	C_{opt}	$\|\Delta U\|$ %	*Err* C_{opt} *(%)*	C_{opt}	$\|\Delta U\|$ %	*Err* C_{opt} *(%)*
50	0.83	17-18	0.82	16-19	C_m	>100	13	C_m	>100	13	**16.0**	0	0
80	0.90	20-22	0.88	19-22	C_m	>100	26	C_m	>100	26	**18.0**	16	5
100	1.00	23-27	0.90	22-26	C_m	>100	36	C_m	>100	36	**19.4**	22	12
240	1.65	39-42	1.55	38-39	24.4	>100	36	**37.6**	11	1	44.2	55	16

**All C_m values are 14 min*

Table 4c: Optimal cycle lengths (C_{opt} in minutes) for Case 3 with different demands

N	Case 3 :L = 3 mile, W = 0.33 mile												
	Simu1		*Simu2*		Design Method Used								
					I			II			III		
	U	C_{opt}	U	C_{opt}	$*C_{opt}$	$\|\Delta U\|$ %	*Err* C_{opt} *(%)*	C_{opt}	$\|\Delta U\|$ *(%)*	*Err* C_{opt} *(%)*	C_{opt}	$\|\Delta U\|$ %	*Err* C_{opt} *(%)*
50	1.10	24-25	1.00	23-25	C_m	>100	13	C_m	>100	13	**22.2**	2	3
80	1.20	27-30	1.20	25-29	C_m	>100	23	C_m	>100	23	**24.6**	6	2
100	1.40	30-33	1.40	27-32	C_m	>100	26	C_m	>100	26	**26.5**	7	2
240	1.80	44-47	1.75	42-46	24.4	>100	42	**37.6**	45	10	58.1	54	26

**All C_m values are 20 min.*

Note that the inequality (7) is found to be true for all our experiments in the ranges of interest around the C_{opt}. In particular, lower demand and/or more compact area cases (like Case 1), which are less critical, are fully satisfied for all values of C. For example, using N = 80, T = 4 hours, C = 9 min (C_m for Case 1), we have that n should be >2.9, which is satisfied by the values in Tab. 1, where n is a value between 3 and 4 for a C = 9 min. For higher C (less critical scenarios), the inequality is satisfied with increasing margin. For higher demand cases and/or more dispersed area cases (like Case 3), we found that some scenarios do not satisfied (7) for lower C (close to C_m), but all begin to be satisfied at some value of C smaller than the C_{opt} (and continue to be valid from that value up), ensuring that our model assumptions are verified around the critical point of interest (C_{opt}).

Besides verifying (7), which is a deterministic approximation, we also similarly verified that all our simulation experiments do not show customers left un-served after one extra cycle.

4. EXTRA DRIVING AND OPTIMAL CYCLE LENGTH

To further explain and validate the intuitive existence of an optimal C discussed in section 3.2 and shown in Fig. 4-6. The expressions $C(n)$ discussed in section 3.3 can be split into two terms. The first term consists of the estimated time taken by the shuttle to go from the terminal to the first passenger in the serving sequence and from the last passenger back to the terminal (this term identifies the estimated *extra* driving to/from terminal). And the second terms consists of the estimated travel time taken to go from the first passenger to the last one, passing by the remaining (n - 2) passengers in the same cycle. For the No-backtracking design method III, Equation (4) can be rearranged to show these two terms, denoted by (i)/(ii).

$$C = \underbrace{\frac{W}{2V} + \frac{4L}{(n+1)V} + 2t_s}_{(i)} + \underbrace{\frac{2L(n-2)}{(n+1)V} + \frac{Wn}{6V} + \frac{W}{6V} + (n-1)t_s}_{(ii)} \qquad (13)$$

We show an example to illustrate the influence of the *extra driving* leading to varying C_{opt}. We assume a fixed demand of N = 100 for the service duration of the feeder with service area dimensions of L = 2 miles and W = 0.5 mile. The chart in Fig. 7 shows the increase in Feeder service duration needed with lower cycle lengths. This increase is attributed due to the extra driving portion represented by Part (i) of Equation (13) which becomes dominating as C continues to decrease below the C_{opt} obtained for T = 4 hours. This is shown using the shaded area in Fig. 7. In fact, with considerably lower cycle length values close to minimum cycle length value, C_m = 14 min, the Feeder shuttle would take more than a day to serve all the demand, which is impractical. An opposite behavior of the extra driving is seen as $C > C_{opt}$, during which, Part (ii) of the Equation (13) dominates. This portion of the cycle length allows shuttle to serve all the demand within T = 4 hours of service duration, however, undermining longer waiting and riding times incurred by the customers. Thus, an optimal cycle length C_{opt} for any given T exists which would ensure an equilibrium between the oversaturation effect of *extra driving* and too much of waiting and riding by the customers.

A companion chart shown in Fig. 8 illustrates the variation of N versus cycle length C for a portion of Feeder service duration, T = 4 hours. A decrease in demand N results in decrease of C_{opt} which allows shuttle to perform more cycles due to greater availability of slack times. This can be visualized clearly with the example for N = 50,80and 100 shown in Fig. 8. For really low values of N, the minimum cycle length (C_m = 14 min) becomes the optimal cycle length, C_{opt}, needed to serve a customer, which is 14 minutes for the assumed service area.

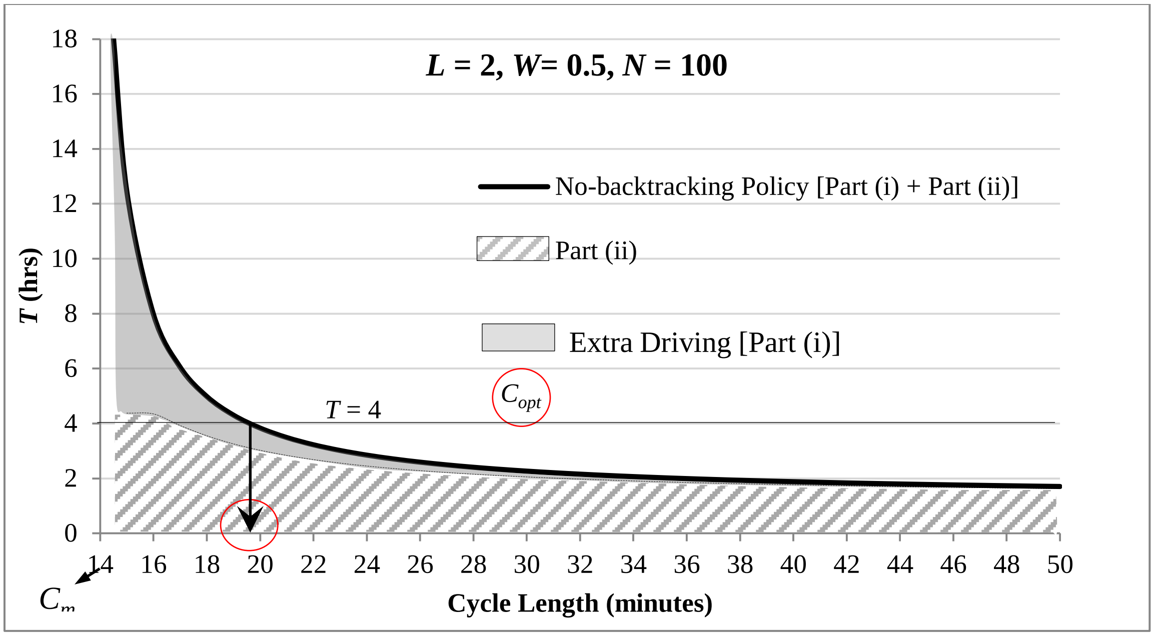

Figure 7: The concept of extra driving with fixed *N*

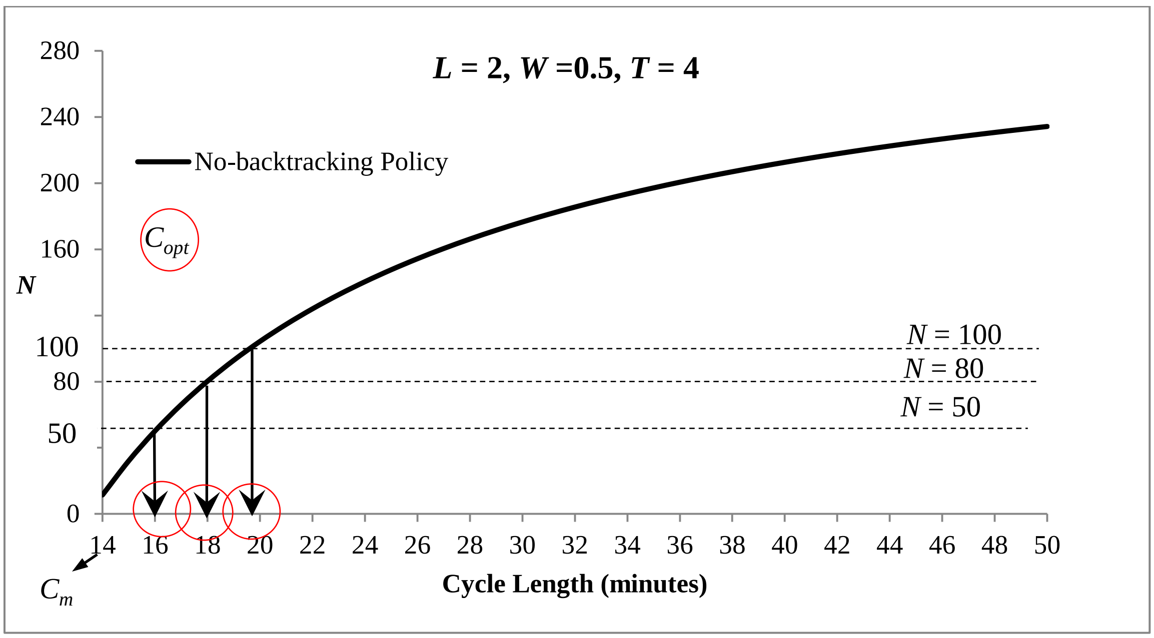

Figure 8: The concept of extra driving with fixed *T*

5. CONCLUSIONS

Feeder services are a potential solution to the first/last mile transit connectivity problem faced by modern society. An improvement of these services would ensure a more pleasant experience for passengers and eventually increase transit ridership, reduce congestion and pollution and increase the livability of residential areas. As noted by

Potts et al. (2010), scheduling problems are a major concern for transit operators for these types of services. This paper proposes the development of an analytical queuing model to estimate the best duration of the cycle length from terminal to terminal using continuous approximation and inputs from demand data and geometrical parameters of the service area. An optimal cycle length C_{opt} must exist to balance two opposite effects: too long cycles would result in excessive riding and waiting time for passengers; too short cycles would cause an oversaturation of the system unable to serve all demand for excessive driving to/from the terminal.

Results give us a handy closed form expression which can be readily used to decide the best dispatch policy to operate a demand responsive feeder transit system. For square service area side length $L = W = 1$ mile, and total demands $N = 50\text{-}100$ over a $T = 4$ hours period, the approximate C_{opt} are found to be 11-15 minutes for the peak hours of commuting. For service areas of length $L = 2$ miles, width $W = 0.5$ mile the estimated optimal cycle lengths C_{opt} vary between 16-20 minutes for the peak hours of commuting. Further, increasing the service area dimensions to a very high length (L= 3 miles) to width (W =0.33 mile) ratio, the estimated optimal cycle lengths are found to be 22-27 minutes. Simulation experiments validate our approach well.

In conclusion, the developed model suggests and encourages transit planners and operators to make use of this methodological approach in selecting the correct operating policy for feeders, whose proper design and operations are becoming increasingly important to enhance the performance of the public transportation system network, within modern sprawled urban and suburban areas. We are aware of the limitations of the model, due to the needed approximation and assumptions used, but we believe our results give a contribution to the research in this area and are useful for practitioners too. Our future work consists of developing models for changes in fleet size and shuttle capacity as an extension of this current work. We are also investigating optimal cycle lengths for non-uniform demand distributions and for appropriate design methods that describe relationships between C and n, other than those mentioned in the paper.

REFERENCES

1. Badami, M.G., Haider, M., 2007. An analysis of public bus transit performance in Indian cities. Transportation Research Part A 41(10), 961-981.
2. Beardwood, J., Halton, J.H., Hammersley, J.M., 1959. The shortest path through many points. Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge Univ Press 55(4), 299-327.
3. Chandra, S., Shen, C.W., Quadrifoglio, L., 2011. A simulation study to derive the optimal cycle length for feeder transit services. International Journal of Information Systems in the Service Sector 3(3), 27-46.
4. Chowdhury, M.S., Chien, S.I.J., 2011. Joint optimization of bus size, headway, and slack time for efficient timed transfer. Transportation Research Record: Journal of the Transportation Research Board 2218, 48-58.
5. Cortés, C. E., Jayakrishnan, R., 2002. Design and operational concepts of a high coverage point-to-point transit system. Transportation Research Record 1783, 178–187.

6. Daganzo, C.F., 1990. Some properties of polling systems. Queueing Systems 6(1), 137-154.
7. Diana, M., Dessouky, M.M., 2004. A new regret insertion heuristic for solving large-scale dial-a-ride problems with time windows. Transportation Research Part B 38(6), 539-557.
8. Gleason, J.M., Barnum, D.T., 1982. Toward valid measures of public sector productivity: Performance measures in urban transit. Management Science 28(4), 379-386.
9. Horn, M.E.T., 2002. Fleet scheduling and dispatching for demand-responsive passenger services. Transportation Research Part C 10(1), 35-63.
10. Jaillet, P., 1988. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. Operations Research 36(6), 929-936.
11. Jaw, J.J., Odoni, A.R., Psaraftis, H.N., 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 20(3), 243-257.
12. Li, X., Quadrifoglio, L., 2010. Feeder transit services: Choosing between fixed and demand responsive policy. Transportation Research Part C 18(5), 770-780.
13. Psaraftis, H.N., 1980. A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem. Transportation Science 14(2), 130-154.
14. Quadrifoglio, L., Chandra, S., Shen, C.W., 2009. Transit services for sprawling areas with relatively low demand density: A pilot study in the Texas border's colonias. Texas Transportation Institute, Texas A&M University System.
15. Quadrifoglio, L., Hall, R.W., Dessouky, M.M., 2006. Performance and design of mobility allowance shuttle transit services: Bounds on the maximum longitudinal velocity. Transportation Science 40(3), 351-363.
16. Quadrifoglio, L., Dessouky, M.M., Palmer, K., 2007. An insertion heuristic for scheduling mobility allowance shuttle transit (MAST) services. Journal of Scheduling 10(1), 25-40.
17. Quadrifoglio, L., Li, X., 2009. A methodology to derive the critical demand density for designing and operating feeder transit services. Transportation Research Part B 43(10), 922-935.
18. Santos, A., McGuckin, N., Nakamoto, H.Y., Gray, D., Liss, S., 2011. Summary of travel trends: 2009 National Household Travel Survey. Federal Highway Administration, Washington, D.C.
19. SuperShuttle, http://www.supershuttle.com/, Accessed on January 7th, 2012.
20. Wardman, M., 2004. Public transport values of time. Transport Policy 11(4), 363-377.

APPENDIX

<u>Derivation of expected distance *E[D]* to the closest passenger demand from the terminal</u>

Within a rectangular service area with length, *L* and width, *W*, spatial-temporal 'uniform' demand distribution of passenger requests follow a Poisson distribution having expected value of ρA_R, with the probability of finding a given number of points (A_n) within an area A_R as

$$P(A_n = q) = \frac{(\rho A_R)^q}{q!} e^{-\rho A_R} \quad \text{where, } \rho = \text{demand density and } q = 1,2,3,\ldots \tag{A-1}$$

The average closest distance between the terminal and a passenger can be obtained for $q = 0$ and $A_R = d^2$ (see Fig. A-1), where d is the variable rectilinear distance between the terminal and the closest passenger.

$$E[D] = \int_0^\infty e^{-\rho d^2} \mathrm{d}d = \frac{1}{2}\sqrt{\frac{\pi}{\rho}} \approx \frac{0.89}{\sqrt{\rho}} \tag{A-2}$$

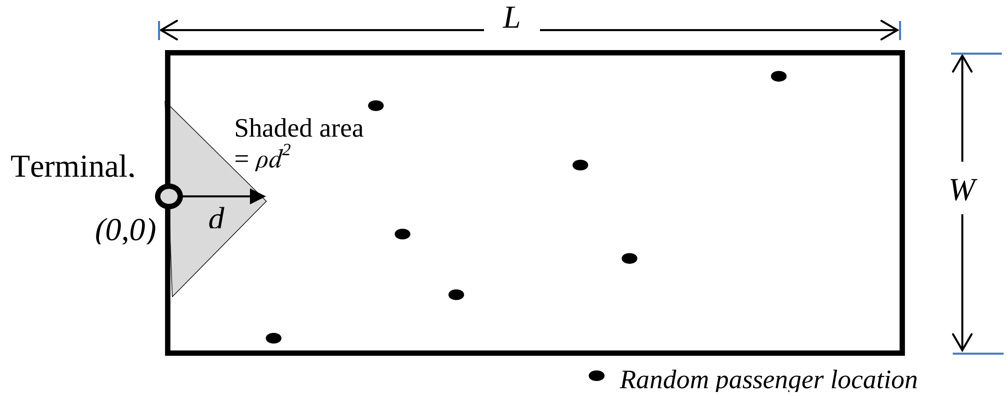

Figure A-1: Illustration of expected closest distance between terminal and a passenger

Using the average closest passenger-to-passenger distance as $\frac{0.63}{V\sqrt{\rho}}$ (by Quadrifoglio et. al (2006)), the cycle length C can be expressed as,

$$C = 2\left(\frac{0.89}{V\sqrt{\rho}}\right) + (n-1)\frac{0.63}{V\sqrt{\rho}} + (n+1)t_s = \left(\frac{1.15}{V\sqrt{\rho}}\right) + \frac{0.63n}{V\sqrt{\rho}} + (n+1)t_s \tag{A-3}$$