Full Title: Dynamic Interlining in Bus Operations

Abstract:
Dynamic interlining of buses is an operational strategy for routes that have a terminal station at a common hub. The strategy keeps a portion of the fleet unassigned to any specific route and allows them to be shared among the routes belonging to the hub (shared fleet). The shared fleet is then dispatched on an on-demand fashion to avoid delays and regulate service. In this paper, a simulation model is developed to examine the impacts of dynamic interlining on service reliability. Using a subnetwork of Boston area bus lines as a case study, the performance of the strategy as well as factors that affect its performance are investigated. Results show that dynamic interlining can improve service reliability and decrease the variations of departure headways at the hub. The size of the shared fleet has the most dominant impact on the performance. Dispatching rules that govern the use of the shared fleet can also impact the performance.

Manuscript Classifications:
- Data and Information Technology; Travel Time, Speed and Reliability ABJ30SB
- General; Operations and Traffic Management; Bus Transit Systems AP050; Headway; Schedule; Transit Management and Performance AP010; Operations and Control; Performance Measures; Transit Capacity and Quality of Service AP015; Efficiency; Quality; General; Public Transportation; Bus Transit Systems AP050; Bus Rapid Transit; Schedule; General; Bus Bunching; Bus Bunching; Vehicle and Crew Scheduling

Manuscript Number: 20-06038

Article Type: Presentation and Publication

Order of Authors:
- Seyedmostafa Zahedi, M.S
- Haris N. Koutsopoulos, PhD
- Zhenliang (Mike) Ma, PhD
Dynamic Interlining in Bus Operations

Seyedmostafa Zahedi
Department of Civil Engineering
Northeastern University, Boston, Massachusetts, 02115
Email: zahedi.s@husky.neu.edu

Haris N. Koutsopoulos
Department of Civil Engineering
Northeastern University, Boston, Massachusetts, 02115
Email: h.koutsopoulos@northeastern.edu

Zhenliang (Mike) Ma
Department of Civil Engineering
Monash University, Victoria, Australia
Email: mike.ma@monash.edu

Word Count: 4400 words + 4 tables (250 words per table) = 5400 words

Submitted August 1st/2019
ABSTRACT
Dynamic interlining of buses is an operational strategy for routes that have a terminal station at a
common hub. The strategy keeps a portion of the fleet unassigned to any specific route and allows
them to be shared among the routes belonging to the hub (shared fleet). The shared fleet is then
dispatched on an on-demand fashion to avoid delays and regulate service. In this paper, a
simulation model is developed to examine the impacts of dynamic interlining on service reliability.
Using a subnetwork of Boston area bus lines as a case study, the performance of the strategy as
well as factors that affect its performance are investigated. Results show that dynamic interlining
can improve service reliability and decrease the variations of departure headways at the hub. The
Size of the shared fleet has the most dominant impact on the performance. Dispatching rules that
govern the use of the shared fleet can also impact the performance.

Keywords: Service Reliability, Dynamic Interlining, Bus Operations, Simulation
INTRODUCTION

Improving transportation service quality to attract more riders is the goal of many transit agencies worldwide as it has direct positive effect on passenger’s behavior (1). One key aspect of service quality is service reliability, which is the ability of the agency to deliver what it has planned for in a timely manner. Service reliability is challenging as transit operations are stochastic in nature. Things usually do not go according to plan as variations in travel demand, traffic conditions, running times and etc, are hard to predict. To overcome the uncertainties associated with operations, transit planners often schedule their services for near to worst case scenarios. For example, the calculation of the fleet size for a bus route with simple cyclical operation relies on the time it takes a bus to complete a round trip on that route. However, running times of buses on a route is not a random variable (2). If the planner designs the fleet size and schedules a service based on the median value of the running times, then half of the time the buses will be delayed as it would take them more time to complete their trip than the time it was planned for. To increase reliability, planners add buffer times to deal with uncertainties. They use values of higher percentiles of running time distributions to decrease the number of occasions when a departure is delayed due to the late arrival of a bus (3). However, the gains in reliability take place at the expense of increased fleet size required to operate the resulting schedule and consequently, reduced operating efficiency. The lower efficiency often manifests itself in buses remaining idle at end stations or drivers intentionally slowing down or holding at stops to avoid being early (4).

Another way to deal with uncertain operations is to insert flexibility into the planned service. A flexible plan is easier to deliver than a fixed one. For instance, Sanchez et al. proposed a schedule free operation in which buses do not have specific arrival/departure times (5). Further and Nash proposed a vehicle pooling strategy, which does not dedicate buses to a specific route but uses the buses on a first come first served basis among all the routes (6). Other strategies allow the operator to adapt to emerging conditions using ad-hoc modifications in the operation such as holding (7) or short turnings (8).

With many bus networks structured as a collection of hubs with routes emanating from them, there is potential to reduce fleet size without sacrificing (or even exceeding) reliability by dynamically interlining buses. Instead of dedicating a predetermined number of buses to each route that are calculated using high percentile of running times, buses (or at least a portion of them) are shared among the routes in a dynamic manner, depending on emerging conditions.

This paper aims to explore the impacts of dynamic interlining on service reliability and identify key factors that affect its performance. An extreme case that has all the fleet shared in a pool has been discussed in (6), but there were no discussion on a mixed fleet of dedicated and shared buses, or the impacts of different dispatching rules. The following questions will be investigated:

- How does the size of the shared fleet affect the performance?
- How should the shared fleet be utilized?
- How does dynamic interlining affect the idle times and efficiency of buses?
- Can dynamic interlining help reduce fleet size?

The rest of the paper is organized as follows. Section 2 defines the idea of dynamic interlining and provides analytical reasoning to show the potential of it. Section 3 discusses the approach and
the details of the developed simulation model. Section 4 investigates the merits of the idea through a number of experiments and lastly, section 4 reports on the conclusions and learned lessons.

DYNAMIC INTERLINING

Traditional transit scheduling assigns a fixed fleet of buses to a route that either serves only that route or is statically interlined with other routes. These buses are dedicated and will be dispatched at a scheduled departure time or at a certain frequency headway. The number of vehicles $N$ required to operate a service with a target level of reliability depends on the design cycle length $C$ and frequency of service $f$, $N = \lceil C \times f \rceil$ (8). However, the vehicle-hours that buses are busy and are actually on the road depend on the average running times of the buses $\bar{R}$, and the number of times they are dispatched, $(\bar{R} \times f)$. The difference between $\lceil C \times f \rceil$ and $(\bar{R} \times f)$ is the average time buses are staying idle. This is a measure of the price agencies are paying to achieve a target level of reliability.

Dynamic interlining allows some of idle buses to be dispatched on other routes instead of waiting for a specific departure time on a dedicated one. It takes some of the buses that typically are dedicated to a route out, and puts them into a “pool” as a shared fleet. As shown in Figure 1, four buses are taken out from three different routes and put at a hub location where they can be shared by all routes. If a bus is late or is stuck in a traffic jam and cannot execute the next scheduled departure, then a shared bus in the “pool” would be dispatched instead. In this manner, the shared buses can be dispatched more frequently and decrease the aforementioned inefficiency while increasing the chances of meeting the schedule for the different routes.

![Figure 1 Dynamic interlining and shared fleet](image)

We can model the operation of a route with a queue in which buses are servers and scheduled dispatches are the demand or the customers (Figure 2a). The amount of time that a customer (a dispatch) keeps a bus busy is $R$ which follows the bus running time distribution of that route. The probability of the event that the system cannot serve the demand at the specific scheduled time is equal to the probability that the number of customers exceed the number of servers i.e. when we have more than “n” dispatching requests in the system Equation 1.
If the buses of another route (shown by a prime) can be borrowed as shown in Figure 2b, new opportunities will open to meet the demand because we can accommodate excessive customers. Therefore, the probability to meet the demand increases by Equation 2:

\[
P_{n+1} (P_{n'-1} + P_{n'-2} + \ldots) + P_{n+2} (P_{n'-2} + P_{n'-3} + \ldots) + \ldots + P_{n+n'} \cdot P_{n'-n'}
\]  

The increased reliability in meeting the schedule, is because of the fact that idle buses on the lending route are being utilized. The comparison can be more tangible in the extreme case of N, M/M/1 queues and 1, M/M/N queue. Combining the servers and having them serve all the demand in one queue results in smaller waiting times and queue lengths.

**APPROACH**

A simulation model was developed to examine the effects of dynamic interlining. The model is an abstraction of the actual bus operations by assuming that buses operate as a shuttle traveling between two terminals. In the simulation, buses have only two checkpoints which are the stations at either end of the route; and they only need to meet the schedule at those points (intermediate stops are not modeled). Each route has at least one of its end stations at the hub where it can use the shared fleet if needed. End-to-end running times are modeled as a random variable using real distributions obtained from AVL data. Routes operate according to a schedule. Based on the schedule, buses are then assigned to each route (labeled dedicated) using a deficit function (9) and a percentile of running time distributions (see fleet size calculation). The time-based simulation advances the clock and goes through all the stations to check if it is time for a dispatch. If a station needs to dispatch a bus on a route, and that station is located on a hub, then the list of dedicated buses to that route is searched. If a dedicated bus is available on that station it will be dispatched. If not, based on the dispatch strategy (see dispatch strategy) and availability of shared buses on that hub, a decision is made to use a shared bus. If the station is not located in a hub, any bus that reached the station earlier will be dispatched (first come first serve). Figure 3 shows a flowchart of the simulation.
Fleet Size Calculation

As mentioned earlier, the deficit function is used to calculate the fleet size for each scenario. The deficit function is a step function measured at a particular terminal. It calculates the minimum number of buses required to implement a schedule \( S_i \) between terminals \( K_i \) during time period \( T_i \) which equals the maximum value of the deficit function at all \( T_i \)s. The deficit function is shown below by Equation 3:

\[
\text{Min } N(S) = \sum_{k \in K} D(k, s) = \sum_{k \in K} \max_{t \in T} d(k, t, s) \quad (3)
\]

Where \( N(S) \), is the required fleet size for the system \( S \), \( D(k, s) \) is the deficit value for terminal \( k \) based on schedule \( s \) and \( d(k, t, s) \) is the deficit as a function of time.

The time it is expected to take a bus to reach a destination where it can be dispatched for another scheduled trip, affects the deficit function and consequently the number of required buses to meet a schedule. The “fleet-size-calculator” engine of this simulation, uses the percentile of the travel time distribution to determine the travel time to be used. The higher the percentile, the larger the fleet size. The resulting fleet size is the number of dedicated buses for the base case. The number of shared fleet that will be used in the simulation is another input. The engine takes buses from the dedicated fleet to fill the shared pool. The taking away of the dedicated buses for the shared pool is done in a manner that the remaining dedicated fleet corresponds to similar travel time percentile for each line. Table1 shows several instances of fleet assignments.
Table 1

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Total fleet</th>
<th>Dedicated buses</th>
<th>Shared buses</th>
<th>Corresponding percentile for dedicated buses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>68</td>
<td>0</td>
<td>85th</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
<td>66</td>
<td>2</td>
<td>75th</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>61</td>
<td>6</td>
<td>65th</td>
</tr>
</tbody>
</table>

**Dispatch strategy**

The simulation uses different dispatching strategies for different stations and different buses. Dedicated buses can only be dispatched on the route to which they belong. Once there is the time for a departure from a station on a route, all dedicated buses for that route are searched. Among the dedicated buses that are available at that station (there could be none), the one that was available earlier will be dispatched. The dispatch strategy for shared buses differs based on the location of departure. If the station requiring a dispatch is not at a hub, then the dispatch strategy for shared buses is the same as that of a dedicated bus. However, if the station is at a hub, the shared buses will only be dispatched if there is no dedicated bus available.

To make good use of the shared fleet at the hub, stations experiencing larger delays will have priority over those with smaller ones. This strategy will be called the larger delay strategy (LD) throughout the rest of this paper.

Also, as small values of delay are usually negligible, a dispatching threshold is used to ensure that shared buses at the hubs are not dispatched unless the next available dedicated bus is late more than the threshold.

**Performance Metrics**

The coefficient of variation of headways is used as a measure of effectiveness (MoE) to evaluate the performance of the system under various scenarios. Large variations in headways on high frequency routes result in large waiting times and can cause bunching (10).

Another MoE which does not specifically affect the passengers but can be important for the operators is the times when buses are idle. Idle times, affect the efficiency of the system. They also impacts the size of the hubs, as larger number of vehicles require a larger physical hub. The simulation also calculates and reports the times when each bus were on the road and the times that each bus arrived early and was waiting at a terminal for the next dispatch.

Number of delayed departures and the magnitude of the delays are also used to evaluate the performance as some agencies incentivize their operators to keep delays under a limit and penalize large delays.
EXPERIMENTAL DESIGN

To understand the impacts of dynamic interlining on service reliability and required fleet size, different scenarios were conducted. The scenarios were defined based on fleet size, dispatching strategy and size of shared fleet.

Two major terminals in the Boston area (Harvard and Dudley) are used as hubs in the experiments. Using the GTFS files available at Massachusetts Bay Area Transit Authority’s (MBTA) website, information on bus routes that use these two hubs were collected. Routes with low frequency, complex interlining and variations were dropped from the analysis in order to have a more homogenous set of operating conditions. Four routes (1, 66, 71 and 73) were chosen. The first two (1 and 66) run between the two hubs while 71 and 73 start from Harvard and go to a none-hub station. Configuration of the network under study and characteristics of the routes are shown in Figure 4.

![Network configuration and route headways](image)

Two years AVL data was used to create an empirical end-to-end running time distribution for the peak period (7:00 AM to 9:00AM) for each route. After removing outliers, the distributions were used in the simulation to generate random running times for buses that are being dispatched.

Scenarios

A total of 60 scenarios were tested. Each scenario is a combination of a total fleet size, shared fleet size, and dispatch threshold. Since ridership was not considered in the simulation and dispatch strategies, all routes are treated the same. In scenarios with small number of shared fleet, having a minimum threshold for dispatching of shared buses is necessary. Otherwise, the shared fleet will be depleted very soon for small delays. Table 2 shows the list of the scenarios that were examined.
Table 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Total Fleet</th>
<th>Shared Fleet</th>
<th>Dispatch Threshold (s)</th>
<th>Dispatch Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>0</td>
<td>0</td>
<td>LD</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
<td>3</td>
<td>180</td>
<td>LD</td>
</tr>
<tr>
<td>3</td>
<td>63</td>
<td>7</td>
<td>120</td>
<td>LD</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>10</td>
<td>60</td>
<td>LD</td>
</tr>
<tr>
<td>5</td>
<td>63</td>
<td>13</td>
<td>0</td>
<td>LD</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
<td>16</td>
<td>0</td>
<td>LD</td>
</tr>
<tr>
<td>7</td>
<td>63</td>
<td>20</td>
<td>0</td>
<td>LD</td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>20</td>
<td>180</td>
<td>LD</td>
</tr>
<tr>
<td>9</td>
<td>63</td>
<td>63</td>
<td>0</td>
<td>LD</td>
</tr>
<tr>
<td>10</td>
<td>61</td>
<td>61</td>
<td>0</td>
<td>LD</td>
</tr>
</tbody>
</table>

Results

The coefficient of variation of headways at both stations of each route is plotted against the number of shared fleet for each scenario in Figure 5. Different colors correspond to different routes. The solid lines are the stations that belong to Harvard, the dashed lines are the ones that are located at Dudley and the dotted lines belong to no hub. The columns at the bottom of the graph show the number of buses that each route donates to the shared fleet. Each cell in the columns is a bus, and the number in the cell is the route it was taken from.

The overall trend of the plot indicates that as more buses are shared, the variation of headways decreases, which means higher reliability. In the initial stages of sharing, where the shared fleet is small, for example 3, the impact of sharing varies on different routes, as some routes contribute more to the shared fleet than others. The routes that contribute their dedicated buses to the shared fleet experience more varied headways (e.g. routes 71, and 73 at Harvard). The performance of the routes who benefit from the shared fleet while not contributing to it will improve (routes 1 and 66). However, as the shared fleet size increases and all routes start contributing, then all the routes start receiving the benefits of sharing. The last scenario in Table 2 assumes all the buses are shared. This design results in better performance for all routes (all-shared scenario).

Looking at the all-shared scenario, the four stations that have the best performance are the stations that belong to the Harvard hub. The next two belong to Dudley, and the rest belong to no hubs. This indicates that in the hubs where there are more lines, more departures and consequently more shared buses coming and going, the performance of routes improves more significantly than the hubs with limited number of routes. Also, the performance of the stations that do not belong to any hubs seems to be independent of sharing. Since these stations have no access to the pool of shared vehicles, they have to dispatch any bus that arrives at their location. As a result, their
performance only depends on the departures from the other end at the hub and the time it takes for the buses to arrive. Sharing buses can only affect and regulate the departures at the hub but has no control over the randomness of running times. Stations 71_No_Hub and 73_No_Hub in Figure 5 indicate that running time variability is dominant and sharing buses does not significantly affect the variations of headways at these stations.

**Figure 5** Impacts of dynamic interlining on the variation of headways

**Figure 6** shows the cumulative distribution of delays for scenarios 1 to 9 (excluding 8). This distribution is obtained by the delays observed in 300 repetitions of each scenario. The base case, (scenario 1) is the third graph from the top, which means scenario 2, and 3 are increasing the frequency of delayed departures. However, the slope of the graphs significantly changes at their corresponding thresholds, meaning that delays larger than the threshold have decreased. Given that small delays are usually negligible, having a dispatching threshold coupled with LD dispatching strategy can help decrease the frequency of large delays. Scenarios 4 to 9 show that having enough shared fleet can decrease the frequency of late departures by almost 50%.
Figure 6 Impact of dynamic interlining on delayed departures

Also, having better performance when all 63 buses are shared, indicates that it is possible to use a smaller total fleet size and still have the performance of the base case with the larger fleet but no shared buses. Table 3 compares the performance of this base case with scenario 10 which has 61 buses that are all shared. This case is clearly better than the dedicated one with no fleet.

Table 3 Impact of dynamic interlining on fleet size

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Harvard</th>
<th>Dudley</th>
<th>No Hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>66</td>
<td>71</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.32</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>0.11</td>
<td>0.16</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Dispatching threshold also has an impact on the performance of dynamic interlining. This threshold should be relative to the number of shared fleet. If the size of the shared fleet is small, it makes sense to keep the shared buses for delays that are large. On the other hand, if for example half of the fleet is shared, a larger threshold can force the shared buses to remain idle while the routes are repeatedly having delays that are only slightly smaller than the threshold. Table 4 compares scenario 7 with another scenario (8) that has the same number of shared buses (20) but implements a threshold of 3 minutes. The figure clearly shows that implementing the threshold is hurting the performance.
Table 4 Impact of dispatching threshold on the performance

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Harvard</th>
<th>Dudley</th>
<th>No Hubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66</td>
<td>71</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>0.11</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td>8</td>
<td>0.11</td>
<td>0.22</td>
<td>0.28</td>
</tr>
</tbody>
</table>

DISCUSSION

As mentioned earlier, one important parameter for transit agencies is bus idle time. The results of the simulations showed no statistically significant change in the average idle time of buses in the case of shared fleet. A value of around 9 minutes per hour was observed for each scenario that used 63 buses. Since the fleet size is 63 and the average idle time is 0.15 hours, then using Little’s law, on average, about 9.9 buses are waiting idle in the system. It can also be mathematically shown that the average number of idle buses serving routes that start from a hub does not change, and is only a function of fleet size, average running time of buses (E(R)) and average headway (E(h)). Consider a “hub and spoke” network configuration such as the one in Figure 1 with m routes where buses are serving round trips from the hub. An observer is monitoring the system during time period T and whenever a bus is dispatched on any line, the observer measures and records the running time of that bus. Note that there are no assumptions about the dispatching strategies or sharing policies. The total running time is the vehicle hours they used to serve those trips for the whole network. The ratio of this value by the time period T, is the equivalent number of buses that were always on the road. (Equation 4).

\[ \text{Equivalent number of busy buses} = \frac{R_1 + R_2 + R_3 + \ldots + R_n}{T} \]  

(4)

Re-arrange the terms in Equation 4 we obtain:

\[ \left( \frac{R_{11} + R_{12} + \ldots + R_{1n_1} + \ldots + R_{m1} + R_{m2} + \ldots + R_{mn}}{T} \right) = \left( \frac{n_1 * E(R_1) + n_2 * E(R_2) + \ldots + n_m * E(R_m)}{T} \right) = \left( \frac{n_1}{n_1} \frac{E(R_1)}{T} + \frac{n_2}{n_2} \frac{E(R_2)}{T} + \ldots + \frac{n_m}{n_m} \frac{E(R_m)}{T} \right) \]

(5)

Equation 5 states that the average number of buses that are on the move in a system is equal to the sum of the average fleet sizes needed for each line for the observed average headway. This means that sharing of buses does not require any additional space for operation, and do not change the idle time.

One important limitation of having a shared fleet is when the schedule has simultaneous dispatches for the routes (aka pulse operation). If the routes at a hub have dispatches at the same time then dynamic interlining of the routes is not as helpful because early buses and late buses
cannot be dispatched on behalf of each other. The increase in reliability shown by Equation 2, is smaller as it is less probable for different queues to be in different states. Dynamic interlining works best when there is a continual rate of departures from the hub. Having enough departures in small time intervals means that a shared bus can be assigned to an on-time dispatch easier and faster given that there are enough shared buses entering the hub in that interval. This is why the performance of the hub at Harvard improves the most. The larger number of routes emanating from Harvard, increases the number of dispatches as well as the number of shared buses which allows the operator at the hub to better assign buses to dispatches.

CONCLUSION

In this paper, a simulation model is developed to examine the impacts of dynamic interlining on service reliability as well as exploring the key factors that affect its performance. The results of the simulation showed that dynamic interlining can improve service reliability and as the size of the shared fleet increases, gains in reliability also increase. The performance can further be improved by implementing dispatching strategies to make better use of the shared fleet. Moreover, the results indicate that dynamic interlining does not change bus idle times and consequently does not change the average number of idle buses. Future research can focus on optimal design of dispatching strategy and also on relaxing some of the underlying assumptions in the simulation, (e.g. potential travel time correlation between consecutive buses).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: H.koutsopoulos, S.Zahedi, Z.Ma; Simulation Developement: S.Zahedi, H.koutsopoulos, Z.Ma; analysis and interpretation of results: S.Zahedi, H.koutsopoulos, Z.Ma; draft manuscript preparation: S.Zahedi, H.koutsopoulos, Z.Ma. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES


