Research Article

Vehicle Routing and Scheduling of Flex-Route Transit under a Dynamic Operating Environment

Yue Zheng [1], **Liangpeng Gao** [2], **and Wenquan Li** [3]

[1]*School of Modern Posts & Institute of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing, China*
[2]*Institute of Transportation, Fujian University of Technology, Fuzhou, China*
[3]*School of Transportation, Southeast University, Nanjing, China*

Correspondence should be addressed to Liangpeng Gao; liangpenggao.acad@gmail.com

To improve the reliability, responsiveness, and productivity of the flex-route transit service, this paper investigates the vehicle scheduling and routing problem under a dynamic operating environment. First, we discuss the new operating polices after the introduction of intelligent transportation systems (ITSs), including automatic vehicle location (AVL) system, mobile data terminal (MDT), and computer-aided dispatch (CAD) system. Second, a mixed integer programming (MIP) formulation is employed to solve the offline routing problem. Third, an online scheduling scheme is presented to tackle different dynamic events, such as dynamic requests, travel time fluctuations, cancellations of requests, and customer no-shows. Finally, simulation experiments based on a real-life flex-route transit service are conducted to assess the influence of different dynamic events. The results demonstrate that the proposed scheduling scheme is reliable for coping with various dynamic events, and our findings can be used to guide the policy making of flex-route transit services.

1. Introduction

Currently, public transit systems play an indispensable role in people's daily life. High volume transit modes such as subway and conventional bus lines can provide convenient, rapid, and efficient service in high-density urban areas. However, in many low-density suburban and rural areas, these traditional fixed-route transit services are considered to be inefficient and inconvenient due to the rigid itineraries and infrequent intervals. Unsurprisingly, the dependency on private cars is reinforced, which poses severe challenges to environmental sustainability and social equity.

To address these challenges, many efforts have been focused on providing new mobility solutions in these low demand regions (e.g., sprawling suburban areas) to not only meet the diverse demands of passengers but also improve the quality of service. Among the many types of flexible transit services, flex-route transit (also referred to as route deviation [1]), mobility allowance shuttle transit [2], and demand adaptive transit system [3]) is the most popular

operating polic [4]. This innovative service is a hybrid public transportation mode that combines the low-cost operability of fixed-route transit with the flexibility of demand-responsive transit. Similar to the conventional fixed-route service, flex-route transit has a base route with several mandatory checkpoints located in high-density demand zones. Typically, these checkpoints are assigned with fixed departure times to serve regular customers and are synchronized with other public transportation lines. On the other hand, flex-route transit shares some common features with demand-responsive transit because the vehicles are allowed to deviate from the base route for a certain distance (usually between 0.25 and 0.75 miles) to serve curb-to-curb requests. Practical operation experiences have shown that ridership and passenger satisfaction remarkably increased after implementing flex-route transit service [5, 6].

In practice, the passengers who need deviation service are required to make reservations at least two hours ahead of time. Therefore, most studies focus on fully static systems in

which all the requests are known in advance and all the unforeseeable events that may happen during the flex-route transit operation are neglected [7–11]. However, real-life situations could be more complex, and the distributions of some accidents cannot be clearly identified. The pre-determined schedule is likely to be modified in real time due to dynamic events such as travel time fluctuations [12–14], cancellations of requests, customer no-shows [15, 16], and dynamic requests [17, 18]. All these unpredictable events severely influence the on-time performance of the flex-route transit and inevitably degrade the system reliability [19]. That may be the one reason why the application of this innovative transit service is quite limited and far behind prior expectations. Therefore, a flexible and efficient real-time scheduling scheme is needed to tackle these practical problems.

Recent developments in intelligent transportation system (ITS) technologies, such as automatic vehicle location (AVL) systems and digital communications, provide solid technical support for real-time fleet management. In particular, these technologies allow a dispatching center to constantly track the real-time positions of the vehicles, traffic information, and up-to-date information about customer requests. In addition, advances in computing power have accelerated the data-processing abilities so that any schedule modifications can be processed and transmitted to the drivers in real time. With the help of these ITS technologies, the dispatcher is now able to quickly track these dynamic events and adjust the schedule of the flex-route transit in real time. A case study on real-life ITS-assisted flex-route transit [5, 20] showed that the schedule reliability was greatly enhanced after the implementation of several ITS projects.

In this paper, we study the vehicle routing and scheduling problem of flex-route transit under a dynamic operation environment. As an extension of our previous research [21], an online scheduling model is developed to deal with the various dynamic events, including customer travel time fluctuations, no-shows, cancellations, and dynamic requests. To the best of our knowledge, the dynamic scheduling problem of flex-route transit has not been examined in past studies. The structure of this paper is organized as follows. The system structure and operating policy of the flex-route transit system are described in detail in Section 2. Then, an offline routing mode and an online scheduling model are proposed in Section 3. In Section 4, we evaluate the system performance under different dynamic events through simulation experiments. A summary of the findings and future works is presented in Section 5.

2. System Description

2.1. Service Area and Demand.
To simplify the modeling process, the flex-route transit system is assumed to operate within a rectangular service area of width W and length L (see Figure 1). All the reserved pick-up/drop-off points should be located within the scope of the service area. There are C checkpoints along the base route (denoted by $c = 1, 2, \ldots, C - 1, C$), which are located at major connection points

or high-density demand zones (see Figure 1). The departure times of the checkpoints are fixed, which can be regarded as hard constraints for vehicle operations.

The flex-route transit service can respond to four different types of requests: type I—the pick-up and drop-off points are both at checkpoints; type II—the pick-up point is at a checkpoint, but the drop-off point is not at a checkpoint; type III—the pick-up point is not at a checkpoint, but the drop-off point is at a checkpoint; type IV—neither the pick-up point nor the drop-off point is at a checkpoint. Type I requests use flex-route transit as a traditional fixed-route transit service. Therefore, they do not need to book the service in advance. The other three types of requests are required to make a reservation by calling or using the Internet to schedule their non-checkpoint stops.

To accommodate possible deviation requests, the scheduled running time between adjacent checkpoints must be greater than the direct running time. The difference between them is called slack time. In flex-route transit, the slack time allocated for the deviation service in each route segment is predetermined based on the expected demand level. If the slack time between the checkpoints is not used up, the vehicles must wait and experience additional idle time until the scheduled departure time. In contrast, when the predetermined slack time is not sufficient to accommodate the unexpected higher demand, some passengers may be rejected (see Figure 1).

2.2. Operating Policy.
In real-life operation, due to the low demand level of suburban areas, the headways of flex-route transit systems are usually more than half an hour [1]. Hence, unlike other demand-responsive shared-ride systems [22, 23], passengers are more likely to reserve a specific trip based on the timetables rather than specify a pick-up time window. As shown in Figure 2, each passenger (except type I) needs to specify a desired trip, a pick-up location, and a drop-off location with the reservation center before the reservation deadline. The reservation center provides the connection between the customers and the dispatching center. It is responsible for recording the demand information and dealing with some dynamic events such as service cancellations, updating demand information, and dynamic requests. Generally, the reservation should be made between two days beforehand to one hour earlier than the departure time of the trip. The requests are processed on a first-come, first-serve basis. Once the request is received, the dispatching center immediately informs the customer whether or not the request can be accommodated. After the reservation deadline, all accepted passenger requests collected for a given trip are used as inputs for the offline routing model to calculate the optimal routing plan. After finishing the calculation process, the planned pick-up times are sent to the accepted customers before the trip begins.

There are two common categories of data that need to be treated differently. The first category of data, called static data, consists of data that are relatively stable and do not need to be updated frequently. Examples include the road
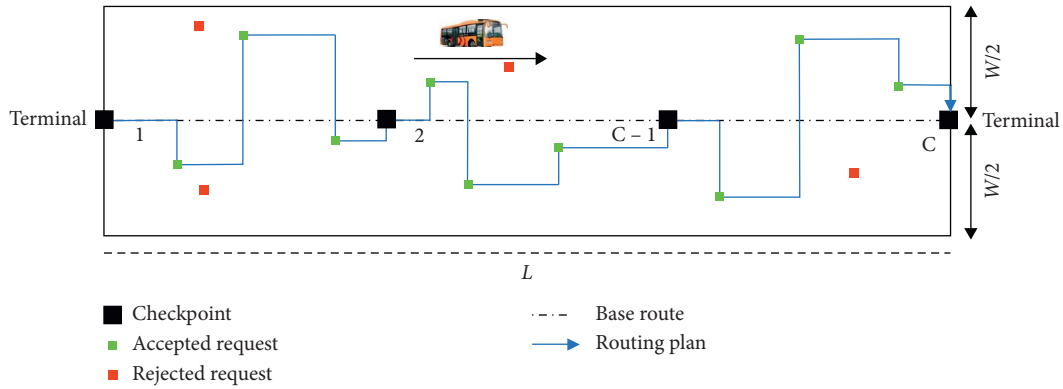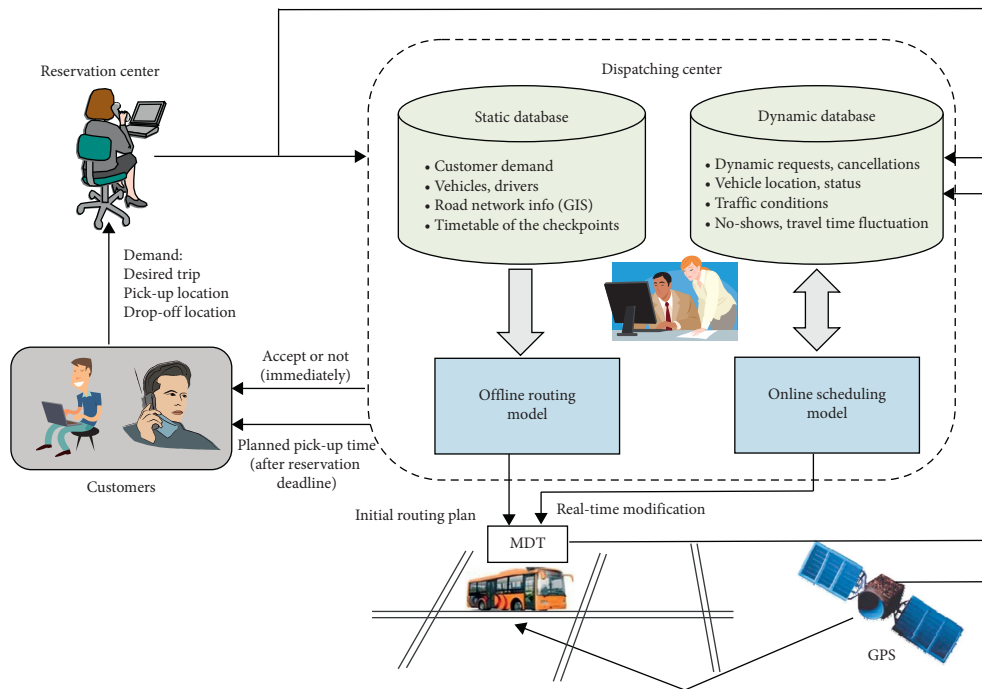
FIGURE 1: Flex-route transit service.

FIGURE 2: System structure of the flex-route transit system.

network topology, customer demand, and information of fleet, drivers, and timetables. Before the actual operation of the service, the initial routing plan of the vehicle is calculated by the offline routing model using the static database. The second category is a dynamic database that includes vehicle location, traffic conditions, and dynamic events. During the operation, the AVL system provides the dispatcher with up-to-date location information for each vehicle via the global positioning system (GPS) equipment. The dispatcher continuously monitors any operational changes in the system and updates them in real time. Once a dynamic event occurs, the dispatcher calls on the online scheduling model to reschedule the planned route, and the modified routing plan is transmitted to the drivers through a mobile data terminal (MDT). Both the offline and online models can be calculated using the computer-aided dispatch (CAD) system installed in the dispatching center to obtain the initial routing plans and responses to dynamic events in real time.

3. Vehicle Scheduling Model

3.1. Offline Routing Model. A mixed integer programming (MIP) formulation is proposed for the offline routing problem. We define a directed graph $G = (N, A, T)$ to describe the vehicle routing network, where N represents all the stops for a trip, $A = N^2$ is the set of arcs, and $T = (t_{ij})$ is the matrix of the travel time associated with A. The objective of the problem is to determine the optimal path without violating the constraints.

The system parameters are defined as follows:

C = number of checkpoints

Q = maximum vehicle capacity

$N_0 = \{1, 2, \ldots, C\}$ = set of stops at the checkpoints

δ_i = scheduled departure time of checkpoint i $\forall i \in N_0$

$K = K_{\text{Type I}} \cup K_{\text{Type II}} \cup K_{\text{Type III}} \cup K_{\text{Type IV}}$ = set of all accepted requests

$TS = C + |K_{\text{Type II}}| + |K_{\text{Type III}}| + 2 \times |K_{\text{Type IV}}| =$ total number of visiting stops in a trip

$N_n = \{C+1, \ldots, TS\} =$ set of non-checkpoint stops reserved by passengers

$N = N_0 \cup N_n =$ set of all stops

$A =$ set of all arcs in the network

$q_i =$ number of passengers boarding ($q_i > 0$) or alighting ($q_i < 0$) at stop i

$t_{ij} =$ rectilinear travel time from node i to node j $\forall i, j \in N$

$dwell\_n =$ dwelling time at every non-checkpoint stop

$dwell\_c = dwell\_n$ dwelling time at every checkpoint

$ps(k) =$ pick-up stop of each request k $\forall k \in K$

$ds(k) =$ drop-off stop of each request k $\forall k \in K$

The decision variables of the model are as follows:

$x_{ij} = \{0, 1\}$ $\forall (i, j) \in A =$ indicating an arc (i, j) is chosen ($x_{ij} = 1$) or not ($x_{ij} = 0$)

$t\_dep_i =$ departure time of stop i, $i = 1, 2, \ldots, TS$

$t\_arr_i =$ arrival time of stop i, $i = 1, 2, \ldots, TS$

$w_i =$ vehicle's passenger load at node i, $i = 1, 2, \ldots, TS$

$p_k =$ pick-up time of request k $\forall k \in K$

$d_k =$ drop-off time of request k $\forall k \in K$

Given the above definitions, the problem can be formulated as a mixed integer linear program in which ω_R and ω_V are the costs of the passengers' in-vehicle time and the operating cost of the service vehicle.

$$\min \omega_R \left(\sum_{k \in K} (d_k - p_k) \right) + \omega_V \left(\sum_{i,j \in A} t_{ij} x_{i,j} \right), \quad (1)$$

subject to

$$\sum_{i=1}^{TS} x_{1i} = 1, \quad (2)$$

$$\sum_{i=1}^{TS} x_{i1} = 0,$$

$$\sum_{i=1}^{TS} x_{Ci} = 0, \quad (3)$$

$$\sum_{i=1}^{TS} x_{iC} = 1,$$

$$\sum_{j=1}^{TS} x_{ji} = \sum_{j=1}^{TS} x_{ij} = 1, \quad i = N/\{1, C\}, \quad (4)$$

$$x_{ii} = 0, \quad \forall i \in N, \quad (5)$$

$$\sum_{i,j \in V} x_{ij} \le |V| - 1, \quad \forall V \subset N, \quad (6)$$

$$t\_dep_i = \delta_i, \quad \forall i \in N_0, \quad (7)$$

$$w_1 = q_1, \quad (8)$$

$$w_j \ge w_i + x_{ij} q_j - MQ(1 - x_{ij}), \quad \forall (i, j) = A, \quad (9)$$

$$t\_arr_j \ge t\_dep_i + x_{ij} t_{ij} - MT(1 - x_{ij}), \quad \forall (i, j) = A, \quad (10)$$

$$t\_dep_i = t\_arr_i + dwell\_n, \quad \forall i \in N_n, \quad (11)$$

$$t\_dep_i \ge t\_arr_i + dwell\_c, \quad \forall i \in N_0/\{1\}, \quad (12)$$

$$0 \le w_i \le Q, \quad \forall i \in N, \quad (13)$$

$$p_k = t\_dep_{ps(k)}, \quad \forall k \in K, \quad (14)$$

$$d_k = t\_arr_{ds(k)}, \quad \forall k \in K, \quad (15)$$

$$d_k > p_k, \quad \forall k \in K. \quad (16)$$

The objective function (1) minimizes the total cost of two factors, namely, the in-vehicle time cost of all passengers and the driving cost of the vehicle; this definition incorporates the costs associated with the transit operators as well as the service quality of the system. Constraints (2) and (3) address the special case of the incoming and outgoing degree of the first stop and last stop, respectively. For the first terminal checkpoint, there is no incoming arc and one outgoing arc; for the last terminal checkpoint, only one incoming arc is allowed, and there is no outgoing arc. For other intermediate nodes, constraint (4) indicates that the incoming/outgoing degree of each node is equal to 1. Constraint (5) forbids the self-connection of each node. Constraint (6) ensures that no illegal subtours are contained in the optimal solution by eliminating all the loops that make the other points disjointed. Constraint (7) guarantees that the scheduled departure times of the checkpoints are fixed. Constraint (8) ensures that for the first terminal checkpoint, the vehicle's passenger load is equal to the number of boarding passengers. Constraint (9) defines the vehicle's passenger load at each node as the passenger occupancy at the preceding node plus or minus the number of passengers boarding or alighting (depending on the sign of q_j). Constraint (10) defines that for any two nodes that have a connection, the arrival time of node j should be no later than

the departure time of node i plus the travel time between the two nodes. We let MQ and MT be a sufficiently large number to ensure that constraints (9) and (10) are irrelevant when there is no connection between the two nodes. Constraint (11) makes sure that the departure time of every non-checkpoint stop is always equal to the arrival time plus the dwelling time. Since there may be some idle time at each checkpoint, constraint (12) guarantees that the departure time of the checkpoint is always later than the arrival time plus the dwelling time. Constraint (13) ensures that the vehicle capacity is not violated during the operation. Constraint (14) establishes the equality for each request between the pick-up time and the departure time of its corresponding node. Similarly, constraint (15) establishes for each request the equality between the drop-off time and the arrival time of its corresponding node. Constraint (16) ensures that the pick-up time of each request should be scheduled before the corresponding drop-off time.

3.2. Dynamic Events.
In reality, some dynamic events may happen during the operation of the flex-route transit service. In our study, we consider four kinds of uncertain events which are summarized as follows. Other rare events, such as vehicle breakdown or traffic jam, are not considered in our research.

(1) Dynamic requests: customers call for service after the reservation deadline of the trip. The pick-up or drop-off points are randomly distributed over the service area.

(2) No-shows: customers do not show up at their pick-up points.

(3) Cancellations: customers cancel requests before or during their reserved trip.

(4) Travel time fluctuation: the travel time between two sites varies due to the changing travel speed.

3.3. Online Scheduling Model.
Recent advances in information and communication have facilitated real-time fleet management. In our assumed flex-route transit system, the real-time location information for each vehicle can be provided by the AVL system. The up-to-date status of customers (such as before service time, on board, serviced, and no-show) can be transmitted through the MDT. Customers' dynamic requests and cancellations can also be accommodated by the reservation center via telephone or through the Internet. The offline and online models can be executed using the CAD system. With these technologies, the vehicle routing and scheduling can now be performed dynamically, introducing greater opportunities to reduce operational costs and improve customer service.

Figure 3 shows the framework of the online scheduling model. In this model, all the events are also sent to the event queue and are processed based on the first-come, first-serve policy. During the operation, the dispatcher continuously fetches one event from the event queue and processes it based on the event type. After finishing the online calculation, the dispatcher sends the updated routing plan to the driver through the MDT and informs the reservation center on how to respond to the customers. The specific scheduling schemes for the corresponding dynamic events are described in the following sections.

3.3.1. Dynamic Request Events.
In a partially dynamic environment, a number of real-time requests, which are not known to the dispatcher at the time of planning, are revealed gradually after the reservation deadline. All the dynamic requests are assumed to be immediate requests that wish to be served instantly. To accommodate these dynamic requests, an insertion heuristic algorithm was adopted to insert these requests into the initial routing plan determined by the offline routing model. The dynamic requests are processed in time order, and the routing plan is updated immediately when a dynamic request is accepted.

Specifically, once a dynamic request is received, the algorithm first checks the feasibility of the insertion. Assume that the requested new stop q is located between checkpoints n and $n + 1$. The algorithm tries to insert the stop into any two consecutive stops a and b that are scheduled within checkpoint n and $n + 1$. If any of these insertions can meet the slack time constraints and vehicle capacity constraints mentioned in equations (7) and (13), the requested new stop is accepted; otherwise, the requested new stop is rejected. It is worth mentioning that type IV passengers need to check the feasibility of both the pick-up point and drop-off point, and only when both stops are inserted successfully can the request be accepted. Since there is likely to be more than one feasible position, the next step is to determine the best insertion position with the minimum incremental costs. The incremental costs of inserting the new stop q between any two consecutive stops a and b can be calculated by the following equation:

$$\Delta F = \omega_V \times \Delta t + \omega_R \times \Delta R + \omega_A \times \Delta A, \qquad (17)$$

where Δt is the extra travel time of the vehicle after the insertion (calculated by equation (18)), ΔR is the sum of the extra in-vehicle time of the passengers caused by the insertion, ΔA is the sum of the extra waiting time of the passengers caused by the insertion, and ω_V, ω_R, and ω_A are the corresponding weights.

$$\Delta t = t_{a,q} + t_{q,b} - t_{a,b} + dwell\_n. \qquad (18)$$

Once the best insertion with the minimum incremental costs is determined, the new stop is inserted into the route, and the updated vehicle routing plan and schedule will be sent to the driver immediately. Meanwhile, the decision regarding the acceptance and the scheduled pick-up time of the dynamic request is also sent to the customer.

3.3.2. No-Show Events.
In practice, some customers may be absent without giving any advanced notice when a vehicle arrives at their reserved pick-up points. A previous study [5] indicated that customer no-shows are frequently observed during the actual operation of flex-route transit services. Therefore, it is necessary to consider no-show events and update the route dynamically.
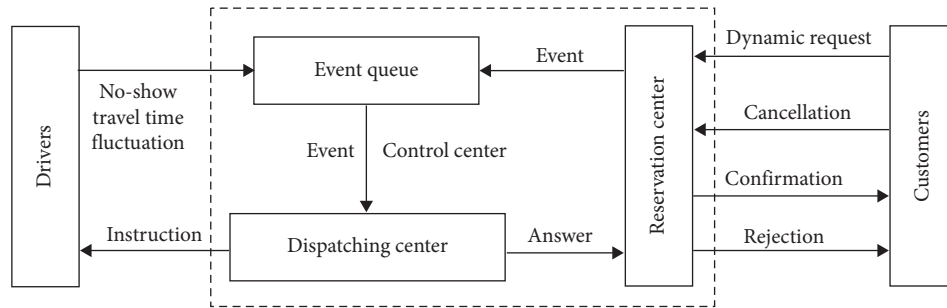
Figure 3: The framework of online scheduling model.

During the operation of the flex-route transit service, once a no-show event is detected, the driver sends a data message (for example: "no-show") to the dispatching center through the MDT. The dispatcher checks the drop-off point of the absent customer in the route. If the drop-off point is at a checkpoint (i.e., type III passengers), the current routing plan remains unchanged. Otherwise, the drop-off point is deleted in the current routing plan without changing the visiting sequence of the other stops. After determining the visiting sequence, the next step is to update the time schedule of the vehicle. More specifically, the schedule of stops before the deleted drop-off point remains unchanged. For the stops after the deleted drop-off point, if it is a drop-off point, the vehicle departs immediately after the customer is dropped off. If it is a pick-up point, the vehicle has to wait at the stop until the planned pick-up time of the customer. This scheme can save the unnecessary travel distance of the vehicle and the in-vehicle time of the passengers onboard. Meanwhile, the on-time performance of the service is also guaranteed.

3.3.3. Cancellation Events. There are two kinds of cancellations that should be treated differently. If the customer cancels the trip before the reservation deadline of the trip, the request is removed from the static database, and the optimal initial routing plan is calculated using the rest of the demand. In contrast, if the cancellation is made after the reservation deadline (i.e., a real-time cancellation), it may affect the scheduling and routing of the vehicle. In particular, once a cancellation event is received by the reservation center, the dispatcher first removes the non-checkpoint pick-up and drop-off stops in the initial routing plan and then reconnects the remaining stops with the shortest distance. Similar to the scheme of no-show events, the vehicle departs immediately after serving the drop-off passengers and waits until the previously notified pick-up times at the pick-up points.

3.3.4. Travel Time Fluctuation. In the offline routing model, the travel time between any two stops is calculated based on the location coordinates, average travel speed, and road networks. However, travel speeds are rarely constant in reality but are instead subject to many stochastic variations, such as random fluctuations in travel demands, frequent interruptions of traffic controls, and unpredictable occurrences of traffic incidents. These travel time fluctuations

inevitably affect the on-time performance of the flex-route transit system, but it can only be known to the dispatching center when the vehicle arrives at the stops. In our study, we assume that the vehicles are operated based on the following schemes.

(i) For the non-checkpoint pick-up points, if the vehicle arrives earlier than expected, the vehicle has to wait for the customers until the planned arrive time. If the vehicle arrives late, passengers would experience a waiting time and the vehicle would depart immediately after picking them up.

(ii) For the non-checkpoint drop-off points, regardless of whether running ahead of or behind schedule, the vehicle departs immediately after dropping off the passengers.

(iii) For the checkpoints, if the vehicle arrives early, it needs to spend additional idle time at the checkpoint until the scheduled departure time. If the vehicle arrives late, passengers who need to get off may experience excessive in-vehicle time and the passengers waiting at the checkpoints also experience a delay. After serving the passengers, the vehicle departs immediately.

Although excessive in-vehicle time and waiting time due to travel time fluctuations have a negative effect on perceived service quality, a workable balance between on-time performance and serving route-deviation requests is difficult to achieve. In our study, the service time fluctuation is not taken into account because it can be considered as part of the travel time between two stops.

4. Result Analysis

4.1. Simulation Settings. In this section, numerical experiments are conducted based on the MTA Line 646, which operates under a flex-route policy in Los Angeles County. This line has been widely regarded as a benchmark and testbed of flex-route transit for comparing and evaluating different system settings, models, and solution methods [2, 8, 10, 11, 24, 25]. The default parameter values are shown in Table 1. Line 646 has three checkpoints that are evenly distributed along the base route. A constant vehicle speed of 25 miles/h and Manhattan distances are assumed in calculating the travel times between the stops. The

TABLE 1: Parameter values.

Parameter	Value
L	10 miles
W	1 miles
C	3
θ	12 passenger/trip
V_b	25 miles/h
ω_V	\$60/vehicle/h
ω_R	\$20/passenger/h
ω_A	\$15/passenger/h
ω_I	\$30/passenger/h
dwell_n	0.3 min
dwell_c	1 min
T_r	40 min
$\eta_1/\eta_2/\eta_3/\eta_4$	0.1/0.4/0.4/0.1

operating cost of the service vehicles and the values of the customer cost indicators are set based on the National Database of the US (2010) and [26]. The cost of the idle time at checkpoints is assumed to be twice the cost of the waiting time. The system is designed with an expected demand level of $\theta = 12$ passengers/trip. For each trip, the requests are generated based on the expected demand level and pre-defined type proportion. The non-checkpoint pick-up or drop-off points are uniformly distributed throughout the service area. In the simulation model, each request is assigned a sequence number in each trip to indicate the order of the reservation time. Based on the theoretical model proposed in [11], we can derive the single-trip time of the flex-route transit T_r as 40 min. Therefore, the scheduled departure times of the three checkpoints are 0, 20, and 40 min.

4.2. Simulation Results.

To average out the effects of randomness in the request generation, 1000 replications of a single-trip operation were run to provide statistical estimates of the system performance for each kind of dynamic event. The offline routing model was solved by GRUOBI (version 9.0) through the YALMIP platform to achieve the optimal results [27].

The system performance of the flex-route transit service is evaluated using the following indicators: (1) the riding time (R), i.e., the average riding time per passenger, including the vehicle running time and service time at each stop (also equal to the in-vehicle time minus the idle time); (2) the waiting time (A), i.e., the average waiting time per passenger at stops; (3) the idle time (I), i.e., the extra waiting time on-board per passenger if the vehicle arrives before the scheduled departure times; (4) the rejection rate (Rej), i.e., the fraction of passengers rejected by the system; (5) operating cost (O_p), i.e., the operating cost per passenger for a trip (it can be calculated by equation (19), where T is the average operation time in one trip and θ_{accept} is the average number of accepted passengers per trip); and (6) system cost (F), i.e., the sum of the operating cost per customer and the average customer cost, which can be calculated by equation (20).

$$O_p = \frac{\omega_V \times T}{\theta_{accept}}, \tag{19}$$

$$F = O_p + \omega_R \times R + \omega_A \times A + \omega_I \times I. \tag{20}$$

4.2.1. Dynamic Requests.

To explore the effect of dynamic requests on the operating efficiency, the degree of dynamism (DOD) is used as an index in our experiments. The degree of dynamism is defined as the ratio between the number of dynamic requests N_d and the total number of requests N_{tot}:

$$DOD = \frac{N_d}{N_{tot}}. \tag{21}$$

For a fixed number of total requests received per trip, we vary the DOD from 0% (pure static scenario) to 100% (pure dynamic scenario) to generate different numbers of static and dynamic requests. The simulation results in Table 2 illustrate that with the increasing DOD, the riding time R decreases from 15.76 min to 15.60 min and then stabilizes after DOD = 75%. This trend is mainly caused by the decreased number of accepted passengers and the fewer detours that are taken by the vehicles. For the same reason, the idle time I is also considerably increased. In contrast, the waiting time A increases dramatically with the increasing DOD. This is because in scenarios with a high DOD, the dynamic requests may be frequently inserted before the prebooking requests or even before some dynamic requests that have already received the scheduled pick-up time. This unexpected delay inevitably deteriorates the reliability of the service and degrades the service level of flex-route transit systems. It is interesting to see that the rejection rate first experiences a gradual increase and then deceases after DOD is higher than 50%. This finding is not in accord with our common intuition that "better routing can be achieved when more information is known." When the system is highly (but not fully) dynamic, there are some advanced requests assigned to the routes. Hence, the route with assigned stops is less flexible for rerouting, and the detour to the accept dynamic requests is constrained by these determined visiting sequences. The same phenomenon was also observed in scheduling demand-responsive transport services [17]. In general, the system cost F increases with the increasing DOD. The system performance can be improved by 1.2% when all the online requests turn into static requests. This suggests that the insertion heuristic is myopic compared to that of the offline routing model, which can provide an optimal solution with all the requests information available. Hence, lowering the DOD may be a more attractive policy for transit operators to improve the system efficiency.

4.2.2. No-Shows.

When a vehicle arrives at the pick-up site and finds that the customer did not show up, the schedule must be quickly adjusted and some wasted costs may be incurred. In our study, the simulations were carried out by varying the no-show rate from 0 to 20% with a random distribution. Table 3 shows that with an increasing no-show

TABLE 2: Simulation results under different DOD values.

DOD (%)	R (min)	A (min)	I (min)	Rej (%)	O_p ($)	F ($)
0	15.76	0	0.46	11.55	3.53	9.01
25	15.69	0.13	0.49	12.20	3.54	9.05
50	15.61	0.30	0.52	12.76	3.55	9.08
75	15.60	0.44	0.53	12.70	3.54	9.11
100	15.60	0.47	0.53	12.64	3.54	9.12

TABLE 4: Simulation results with different cancellation rates.

Cancellation (%)	R (min)	A (min)	I (min)	Rej (%)	O_p ($)	F ($)
0	15.76	0	0.46	11.55	3.53	9.01
5	15.60	0	0.56	11.55	3.64	9.13
10	15.47	0	0.68	11.55	3.77	9.27
15	15.31	0	0.77	11.55	3.91	9.41
20	15.16	0	0.87	11.55	4.07	9.56

TABLE 3: Simulation results with different no-show rates.

No-show (%)	R (min)	A (min)	I (min)	Rej (%)	O_p ($)	F ($)
0	15.76	0	0.46	11.55	3.53	9.01
5	15.70	0	0.50	11.55	3.67	9.15
10	15.66	0	0.56	11.55	3.82	9.32
15	15.60	0	0.59	11.55	3.99	9.49
20	15.55	0	0.65	11.55	4.19	9.69

rate, the riding time R declines constantly. However, the idle time I gradually increases. In general, the passengers' in-vehicle time, which is equal to the sum of the riding time and the idle time, remains stable. This finding shows that although the rerouting could save the travel distance for not serving the drop-off points of the no-show customers, it could not significantly reduce the in-vehicle time of the passengers. Even worse, passengers who experience a prolonged idle time may develop a negative perception towards the flex-route transit service. We also observe that the operating cost O_p increases with the growing number of no-show passengers. This is because the salvaged operation time could not compensate for the decreased ridership. As a consequence, the system cost gradually increases with an increasing no-show rate.

4.2.3. Cancellations. In this study, only a real-time cancellation is considered. Unlike cancelling requests in advance, this kind of cancellation affects the scheduling and routing of vehicles and needs to be adjusted in real time. Table 4 indicates that the system performance indicators show a similar trend to the no-show events. The difference is that a real-time cancellation has a more significant impact on the riding time R and idle time I. This is because both the pick-up and drop-off points could be skipped when the cancellation happens while the vehicle can only skip the drop-off stops when the passengers are absent. The saved travel distance also leads to a lower operating cost O_p than their counterparts of no-show events. Overall, both the no-show and cancellation events have a great impact on the system performance, and some extra cost is incurred for both the customers and transit operators. Therefore, customers who have a poor record of no-shows or cancellations should be identified and added onto a blacklist in practice to prevent the frequent occurrence of such events.

In Figure 4, we compare the system cost without a dynamic intervention and those with a dynamic rerouting
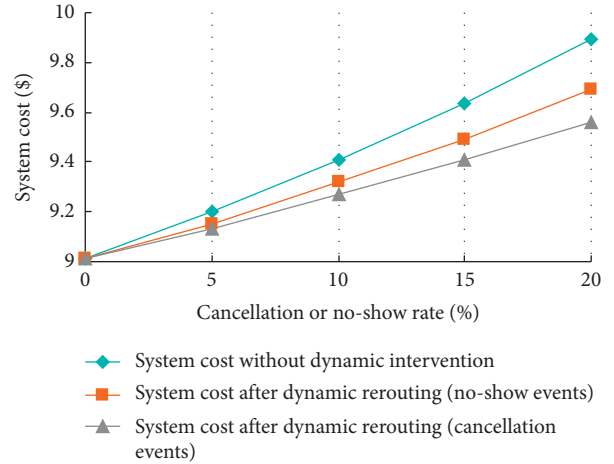

FIGURE 4: System cost with and without dynamic intervention.

scheme. We can see that our proposed rerouting scheme could save the system cost F for both no-show and cancellation events. This is because after the dynamic intervention, the unnecessary travel distance to reach the voided pick-up or drop-off points is eliminated, which leads to a lower operating cost O_p. This finding confirms that more high quality schedules can be generated under real-time adjustment.

4.2.4. Travel Time Fluctuation. The travel time that is estimated based on the average travel speed of the road segment is very likely to fluctuate in practice. Therefore, in our simulation, we assume that the travel time t_{ij} between point i and point j is normally distributed with its mean $\overline{t_{ij}}$ calculated based on the average vehicle speed, and its standard deviation σ_{ij} is defined as

$$\sigma_{ij} = \eta \overline{t_{ij}}. \tag{22}$$

We vary η from 0 to 0.2, and simulation results are shown in Table 5. It reveals that the riding time R and operating cost O_p are almost unaffected by the travel time fluctuation. However, due to the frequent early arrival and delay at both the requested stops and checkpoints, the passengers may have to bear some unwanted waiting time and idle time. These unexpected time costs can also be reflected in the system cost F. In our experiments, we assume that the time constraints of the checkpoints can be violated in the dynamic operation environment. This is not a problem for most cases, except when the checkpoint is synchronized with other transit lines. In this case, some

TABLE 5: Simulation results with different travel time standard deviations.

η	R (min)	A (min)	I (min)	Rej (%)	O_p ($)	F ($)
0	15.76	0	0.46	11.55	3.53	9.01
0.05	15.78	0.06	0.50	11.55	3.53	9.06
0.10	15.76	0.14	0.56	11.55	3.53	9.10
0.15	15.74	0.22	0.64	11.55	3.53	9.16
0.20	15.78	0.33	0.70	11.55	3.53	9.22

passengers who are not served may have to be abandoned by the designated vehicles to save the insufficient slack time. To remedy the inconvenience, the transit operator should arrange another vehicle or give some financial compensation to these passengers.

5. Conclusions

As the most widely used flexible transit services, flex-route transit is a promising option to meet the diversified travel needs in low-density areas and cover the first/last mile of a trip. Most studies have focused on fully static systems in which all the input is known beforehand and the vehicle routes do not change once they are being executed. This idealized assumption is fragile in real-life cases because some dynamic events occur during the execution of the plan and the predetermined schedule and route need to be adjusted via online manipulations.

In this article, we study the vehicle routing and scheduling problem of flex-route transit under a dynamic operating environment. A two-stage scheduling model is designed to address the prebooking requests and dynamic events. In the first stage, an offline routing model formulated as a mixed integer program is employed to build the initial routing plan. In the second stage, an online scheduling scheme is proposed to cope with the different dynamic events.

Simulation experiments were carried out to evaluate the influences of different dynamic events. The simulation results reveal that although allowing real-time requests can improve the reservation flexibility of the system, more prebooking requests can improve the system performance. This finding suggests that the transit operator should encourage more passengers to reserve their trip in advance. A possible incentive mechanism for stimulating the riders' willingness to prebook is to reduce the transit fare for advanced requests or increase the transit fare for dynamic requests. In addition, mobile-based applications can be developed to provide real-time vehicle location of the vehicles and the updated estimated arrival times.

We also find that although the dynamic intervention could reduce unnecessary travel distances, the no-show and cancellation events still have a substantial negative impact on the operating cost of the transit operators. Frequent occurrence of these events not only leads to revenue loss but also wastes the slack time, by which some rejected passengers could have been served by the transit service. Therefore, we suggest that the transit operator should identify the customers who have a poor record of no-shows or cancellations and add them onto a blacklist. In the near future, with the widespread use of mobile payment technologies, deducting the transit fare in advance may be a more efficient way to prevent these events.

Moreover, the results show that the travel time fluctuations lead to an increase in the waiting and idle time for passengers. The influence is minor compared to other dynamic events. However, it is noteworthy that punctuality at the checkpoints that are synchronized with other transit lines should be guaranteed; otherwise, the delay at these time points may result in missed connections for the passengers.

To the best of our knowledge, this is the first work that addresses the dynamic scheduling problem of the flex-route transit service. Future work will investigate the stochastic vehicle routing problems of the flex-route transit service. The inherent uncertainty of the inputs, such as demands and travel times, will be considered in the planning phase.

Data Availability

All data generated or analyzed during this study are included within this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. F. Potts, M. A. Marshall, E. C. Crockett, and J. Washington, *A Guide for Planning and Operating Flexible Public Transportation Services: TCRP Report 140*, Transportation Research Board, Washington, DC, USA, 2010.

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

[3] T. G. Crainic, F. Errico, F. Malucelli, and M. Nonato, "Designing the master schedule for demand-adaptive transit systems," *Annals of Operations Research*, vol. 194, no. 1, pp. 151–166, 2012.

[4] D. Koffman, *Operational Experiences with Flexible Transit Services: a Synthesis of Transit Practice. TCRP Synthesis*, Transportation Research Board, Washington, DC, USA, 2004.

[5] E. Bruun and E. Marx, "OmniLink," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1971, no. 1, pp. 91–98, 2006.

[6] S. R. Fittante and A. Lubin, "Adapting the Swedish service route model to suburban transit in the United States," *Journal of the Transportation Research Board*, vol. 2536, pp. 52–59, 2015.

[7] F. Malucelli, M. Nonato, and S. Pallottino, "Demand adaptive systems: some proposals on flexible transit 1," in Operations Research in Industry, T. Ciriania, E. Johnson, and R. Tadei Eds., pp.157–182, McMillan Press, London, UK, 1999.

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

[9] B. Alshalalfah and A. Shalaby, "Feasibility of flex-route as a feeder transit service to rail stations in the suburbs: case study in Toronto," *Journal of Urban Planning and Development*, vol. 138, no. 1, pp. 90–100, 2012.

[10] F. Qiu, W. Li, and J. Zhang, "A dynamic station strategy to improve the performance of flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 229–240, 2014.

[11] Y. Zheng, W. Li, and F. Qiu, "A slack arrival strategy to promote flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 442–455, 2018.

[12] L. Fu, "Improving paratransit scheduling by accounting for dynamic and stochastic variations in travel time," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1666, no. 1, pp. 74–81, 1999.

[13] L. Fu, "Scheduling dial-a-ride paratransit under time-varying, stochastic congestion," *Transportation Research Part B: Methodological*, vol. 36, no. 6, pp. 485–506, 2002.

[14] M. Schilde, K. F. Doerner, and R. F. Hartl, "Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem," *European Journal of Operational Research*, vol. 238, no. 1, pp. 18–30, 2014.

[15] Z. Xiang, C. Chu, and H. Chen, "The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments," *European Journal of Operational Research*, vol. 185, no. 2, pp. 534–551, 2008.

[16] T. Kim, *Model and Algorithm for Solving Real Time Dial-A-Ride Problem*, Ph. D. thesis, University of Maryland, College Park, MD, USA, 2011.

[17] K. I. Wong, A. F. Han, and C. W. Yuen, "On dynamic demand responsive transport services with degree of dynamism," *Transportmetrica A: Transport Science*, vol. 10, no. 1, pp. 55–73, 2014.

[18] J. Shen, S. Yang, X. Gao, and F. Qiu, "Vehicle routing and scheduling of demand-responsive connector with on-demand stations," *Advances in Mechanical Engineering*, vol. 9, no. 6, Article ID 1687814017706433, 2017.

[19] R. G. Farwell and E. Marx, "Planning, implementation, and evaluation of omniride demand-driven transit operations: feeder and flex-route services," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1557, no. 1, pp. 1–9, 1996.

[20] J. A. Lee, *Evaluating its Investments in Public Transportation: A Proposed Framework and Plan for the Omnilink Route Deviation Service*, Virginia Polytechnic Institute and State University, Blacksburg, VI, USA, 2002.

[21] Y. Zheng and W. Li, "Flex-route transit service with different degree of dynamism," in *Proceedings of the 19th COTA International Conference of Transportation Professionals*, pp. 4369–4380, Nanjing, China, 2019.

[22] S. C. Ho, W. Y. Szeto, Y.-H. Kuo, J. M. Y. Leung, M. Petering, and T. W. H. Tou, "A survey of dial-a-ride problems: literature review and recent developments," *Transportation Research Part B: Methodological*, vol. 111, pp. 395–421, 2018.

[23] D. Huang, Y. Gu, S. Wang, Z. Liu, and W. Zhang, "A two-phase optimization model for the demand-responsive customized bus network design," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 1–21, 2020a.

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

[25] F. Qiu, W. Li, and A. Haghani, "A methodology for choosing between fixed-route and flex-route policies for transit services," *Journal of Advanced Transportation*, vol. 49, no. 3, pp. 496–509, 2015.

[26] M. Wardman, "Public transport values of time," *Transport Policy*, vol. 11, no. 4, pp. 363–377, 2004.

[27] D. Huang, X. Chen, Z. Liu, C. Lyu, S. Wang, and X. Chen, "A static bike repositioning model in a hub-and-spoke network framework," *Transportation Research Part E: Logistics and Transportation Review*, vol. 141, Article ID 102031, 2020.