Dynamic Truckload Routing, Scheduling and Load Acceptance for Large Fleet Operation with Priority Demands

Yongjin Kim¹, Hani S. Mahmassani²*, Patrick Jaillet³

1. Center for Northeast Asian Transportation, Logistics & Transport Economics
   The Korea Transport Institute
   2311 Daehwa-dong, Ilsan-gu, Koyang-city, Kyonggi-do, 411-701, Korea
   Tel. 82.31.910.3149
   Fax. 82.31.910.3226
   Email: yongjin@koti.re.kr

2. Department of Civil and Environmental Engineering
   The University of Maryland at College Park
   College Park, MD 20742-3021
   Telephone: (301) 405-0221
   Fax: (301) 405-2585
   E-mail: masmah@wam.umd.edu

3. Department of Civil and Environmental Engineering
   Massachusetts Institute of Technology
   Room 1-290
   77 Massachusetts Ave.
   Cambridge, MA 02139
   Tel: (617) 452-3379
   Fax: (617) 452-3294
   E-mail: jaillet@mit.edu

* Corresponding Author

Text = 5,713
Figures and Tables =1,750
Total = 7,463

Submitted for presentation at the 83rd Annual Meeting of the Transportation Research Board, January 2004, Washington, D.C. and publication in Transportation Research Record
Dynamic Truckload Routing, Scheduling and Load Acceptance for Large Fleet Operation with Priority Demands

Yongjin Kim, Hani S. Mahmassani and Patrick Jaillet

ABSTRACT

This paper focuses on dynamic load acceptance/rejection decisions in dynamic (online) truckload routing and scheduling problems with time-windows when the carrier provides two classes of service. The delivery service is classified into two types -priority and regular service- with different price rate corresponding to various customer requirements in terms of time sensitivity vs. price sensitivity. For this problem, a Mixed Integer Programming formulation is provided to solve a static version of the problem. Then, a dynamic acceptance/rejection policy is proposed and tested as a part of dynamic operation policy, in which the decision should be made upon arrival of the demand or at least in a short time. This acceptance decision based on system state (vehicles’ location, status and their scheduled demands) at a decision moment represents the approximation of the expected number of vehicles being able to serve future (still unrealized) priority demand. Simulation experiments conducted to evaluate the relative performance show that the proposed policy significantly improves overall profit relative to the benchmark policies.
This paper focuses on dynamic load acceptance/rejection decisions in dynamic (online) truckload routing and scheduling problem with time-windows under situations that a carrier provides two classes of service. In this problem, the demands are requested dynamically and a decision maker needs to decide whether accept or reject a demand upon arrival of the demand and assign it to a feasible vehicle. It is assumed that a carrier provides two types of delivery service corresponding to various customer requirements in terms of time sensitivity vs. price sensitivity. The customers’ choice between these two types of service is also revealed dynamically as the routes are executed along with other demand information including origin, destination locations, and time-window. In addition, over-saturated demand situations are assumed: the demand exceeds the system’s average capacity. Thus, the decision maker has opportunities to select profitable demands.

Problem context including the explanation of the two demand types and formal statement of the problem is provided in the first section. Next, a dynamic dispatching system is briefly explained along with Mixed Integer Programming model. A dynamic acceptance/rejection policy is described in the following section. In this section, general concept of the acceptance/rejection decision in the dynamic routing and scheduling problem is provided. Furthermore, a new decision criterion is developed and a dynamic operational policy implementing this criterion in a dynamic problem setting is presented. Finally, numerical results are reported, followed by concluding comments.

**PROBLEM CONTEXT**

Two classes of service are provided in order to respond to the following two types of customers: A fraction of customers requires time sensitive ‘priority’ or express and on-time delivery service. In other words, they are willing to pay a premium for on-time and earlier delivery. Thus, the time-window width of this type is relatively narrow compared to that of the ‘regular’ type demand. The other customers are more sensitive to price, and request the ‘regular’ low-price service. This class of demands has wider and flexible time-windows.

In order to presents the demand characteristics, two types of time-windows are specified as in Figure 1, in which the solid line represents the penalty incurred, and $\tau^-$ and $\tau^+$ denote the earliest and latest pickup times of a demand, respectively. Relatively narrow Type I time-
window is employed to represent the priority demand property, namely to ensure the on-time and express delivery. In this type, no penalty is charged if a demand is picked up at any time within the specified time interval. In contrast, Type II time-window is employed for regular demands. In this case, one more component, $\tau^{cr}$, denoting a “critical” time is introduced such that no penalty is incurred if the demand is picked up in $(\tau^{c}, \tau^{cr})$. Otherwise, if the demand is picked up after the “critical time”, a penalty proportional to the amount of “over time” (from $\tau^{cr}$ to the actual pickup time) is assessed. In this way, Type II favors earlier delivery after the “critical time”. Note that, both type of the time-windows are hard time windows. Thus, the demand must be picked up before the latest pickup time $(\tau^+)$. 

The objective of the fleet management of this problem is to find a dynamic operation policy to maximize profit over the demands requested during the specified finite time horizon $[0, T]$. The profit is equivalent to earned revenue minus required cost. Since it is assumed that the revenue is proportional to the haul-length and depends on the demand type, the revenue is managed through a load-acceptance decision. The cost is composed of fixed and variable components. The cost associated with the vehicle is a fixed cost because it is assumed that the fleet size is fixed. This assumption stems from the fact that the fleet size is determined by a long-term plan and cannot be modified in the short run. The variable operating cost is incurred through the traveled distance, including loaded and empty movements. Once a demand is accepted, the associated loaded distance is fixed. On the other hand, the empty movement required to serve an accepted demand is still variable until the demand is picked up because the dispatcher can modify the routing schedule. Therefore, one of the main responsibilities of the dispatcher is to minimize the variable costs particularly the empty movements by appropriate assignment (routing schedule) decisions. Note that, the assignment decisions have significant impact on the fleet capability of accepting demands.

The objective function of the problem is defined as follows. Three variables are defined to record the outcomes of decisions under a certain policy $\pi$. First, for the acceptance/rejection decision, let $D_i^\pi$ denote the decision of whether to accept or reject the requested demand $i$ when a certain policy $\pi$ is applied as follows:
\[ D_i^\pi = \begin{cases} 1 & \text{if request } i \text{ is accepted} \\ 0 & \text{if request } i \text{ is rejected} \end{cases} \]

Second, let \( \phi_i^\pi \) denote the transportation cost required to serve load \( i \). This value is based on the empty traveled distance to pick up the load. Finally, \( \delta_i^\pi \) represents the pickup time of demand \( i \). The last two decision variables, \( \phi_i^\pi \) and \( \delta_i^\pi \) are determined from the routing schedule that a fleet of vehicles actually follow under policy \( \pi \). In addition, two parameters \( R_i \) and \( \beta \) are defined; \( R_i \) denotes the reward from load \( i \), which is proportional to the loaded distance \( (l_i) \) and price rate \( (r) \) as well as the type of the demand (\( \xi_i = 1 \) regular demand, \( \xi_i = 0 \) priority demand) such that \( R_i = r \times l_i \), in which different unit prices per unit-loaded distance are applied \((r_{priority} > r_{regular})\). Let \( \beta \) denote the transportation cost per unit distance. Then, the total profit obtained under policy \( \pi \) is defined as

\[
V^{\pi^*} = \max_{\pi \in \Pi} \sum_{i \leq T} D_i^\pi (R_i - \beta(\phi_i^\pi + l_i) - \gamma l_i \xi_i (\delta_i^\pi - \tau_i^\pi))
\]

Where \( V^{\pi^*} \) captures the fleet’s total revenue and transportation cost including traveled empty distance and loaded distance under policy \( \pi \) over the requested demand in \{0, T\}. Note that \( A_i \) represents the arrival time of demand \( i \).

**Assignment Decision**

One of the main decisions required of the dispatcher is assignment decision (routing and scheduling). In this paper, the vehicle assignment decision follows the dynamic operating policy developed in (1-3). The basic scheme to construct a routing and scheduling of the fleet is to solve successive deterministic local snapshot problems repeatedly, as close to optimality as possible. The process of updating the routing schedule is triggered by new load arrivals and reassignments of loads. In other words, various assignment techniques are adaptively applied depending on the state of the system in order to construct an initial schedule and/or improve the schedule by re-optimizing existing routes while keeping the response time to the customer within a tolerable
range. In these various assignment procedures, successive local snapshot problems are set and solved using a Mixed Integer Programming model. For two-class demand situation, a new formulation is required. The following is developed based on the formulation presented in Yang et al. (4-5) and Kim et al. (3) in which detailed discussion of basic formulation structure is presented.

This formulation corresponds to a routing and scheduling problem involving two types of demands with time-window constraints. There are $K$ vehicles ($1, \ldots, K$) and $N$ demands at time $t$ in a local problem. The objective function is to find the least-cost set of cycles that involve all the nodes of ($I, \ldots, K, K+I, \ldots, K+N$) where node $k$ ($k = 1, \ldots, K$) represents vehicle $k$ and node $K+i$ ($i = 1, \ldots, N$) corresponds to demand $i$. The binary decision variable $x_{uv}$ ($u, v = 1, \ldots, K+N$) indicates whether arc $(u, v)$ is selected in one of the cycles.; $x_{k,K+i}$ ($k = 1, \ldots, K; i = 1, \ldots, N$) indicates whether vehicle $k$ serves demand $i$ first, $x_{K+i,k,j}$ ($k = 1, \ldots, K; i,j = 1, \ldots, N$) indicates if demand $j$ is served immediately after demand $i$, $x_{K+i,K+i} = 1$ ($k = 1, \ldots, K$) means that demand $i$ is rejected, and $x_{k,k}$ = 1 represents that vehicle $k$ will be idle. Furthermore, the continuous variable $\delta_i$ represents the pickup time of demand $i$.

The input parameters are extracted from the information collected at a decision epoch, and are updated considering the vehicle status. Considering vehicle status, the time that vehicle $k$ will become available is denoted by $\nu_k$. The distance from vehicle $k$’s updated location to demand $i$’s origin is denoted by $d_{0i}^k$. The distance from demand $i$’s destination to demand $j$’s origin is expressed by $d_{ij}$. In addition, $l_i$ represents the loaded distance of demand $i$. Finally, $M$ and $B$ are a constant large enough to let constraint (5) be nonrestrictive when $x_{K+i,K+j} = 0$.

**MIP Formulation**
The formulation presented here is general one, in that it allows acceptance/rejection of load requests. However, the dynamic operation policy applied here does not allow rejecting any feasible demands once a local problem is set. Thus, \( \rho \) is specified as a large constant to preclude any rejection. A variable \( \omega_i \) is introduced to represent the overtime, the elapsed time from the critical time \( \tau^c_i \) to the pickup time \( \delta_i \) for a priority demand \( \xi_i = 0 \). Due to constraints (5) as well as the objective function, \( \omega_i \) takes a positive value only if a regular demand \( \xi_i = 1 \) is served after the corresponding critical time \( \tau^c_i \). Otherwise, the value of \( \omega_i \) is zero.

The penalty is proportional to the overtime \( \omega_i \) as well as haul length of demand \( i \). \( \gamma \) denotes a scaling factor to determine the penalty charged to the objective function.

\[
\text{Min } \sum_{k=1}^{K} \sum_{i=1}^{N} d_{0i}^k x_{k,i,K+i} + \sum_{j=1}^{N} l_j x_{K+i,K+i} + \sum_{j=1}^{N} d_{ij} x_{K+i,K+j} + \gamma \sum_{i=1}^{N} l_i \omega_i \\
\text{subj to}
\]

\[
\sum_{v=1}^{K+N} x_{uv} = 1 \quad \forall \ u = 1, \ldots, K + N \tag{1}
\]

\[
\sum_{v=1}^{K+N} x_{uv} = 1 \quad \forall \ u = 1, \ldots, K + N \tag{2}
\]

\[
- \sum_{k=1}^{K} (d_{0i}^k + \nu^k) x_{k,i,K+i} + \delta_i \geq 0 \quad \forall \ i = 1, \ldots, N \tag{3}
\]

\[
(l_i + d_{ij}) x_{K+i,K+i} - M x_{K+i,K+i} - \delta_i + \delta_j \geq -M + l_i + d_{ij} \quad \forall \ i, j = 1, \ldots, N \tag{4}
\]

\[
B+ \omega_i - \delta_i \geq B \xi_i - \tau^c_i \quad \forall \ i = 1, \ldots, N \tag{5}
\]

\[
\omega_i \geq 0 \quad \forall \ i = 1, \ldots, N \tag{6}
\]

\[
x_{uv} \in \{0,1\} \quad \forall \ u, v = 1, \ldots, K + N \tag{7}
\]

\[
\tau^c_i \leq \delta_i \leq \tau^c_i \quad \forall \ i = 1, \ldots, N \tag{8}
\]
ACCEPTANCE DECISION POLICY

Conceptual Framework

A general optimal acceptance rule in a dynamic and stochastic knapsack problem (6) and distribution problem (7) can provide good insight into the acceptance decision of the dynamic fleet management problems. This concept can be applied to this problem as follows. Upon arrival of demand \( j \), the current locations and status of the vehicles at time \( t \) under a certain policy \( \pi \) are represented by \( \Gamma^\pi_t = \{ \Gamma^\pi(k,t), k = 1, ..., K \} \), and the vehicle routing schedule of the fleet under policy \( \pi \) is denoted by \( Q^\pi_k = \{ q^{k,\pi}(t), k = 1, ..., K \} \). Let \( q^{k,\pi}_j(t) \) denote the \( j \)th scheduled load in vehicle \( k \)'s job queue under policy \( \pi \) at time \( t \). Then the queue of vehicle \( k \) is described by a vector \( q^{k,\pi}(t) = \{ q^{1,\pi}_1(t), q^{2,\pi}_1(t), ..., q^{1,\pi}_K(t) \} \). The above two vectors, \( \Gamma^\pi_t \) and \( Q^\pi_t \) fully describe the current (at time \( t \)) dynamics of the operational system under policy \( \pi \). In addition, when the demand \( j \) is accepted, let \( \overline{Q}^\pi_t \) represent the updated routing schedule including demand \( j \) under policy \( \pi \) at time \( t \). Note that the locations and status of vehicles do not change. In other words, it is assumed that the solution execution time is ignored. With these notations, the optimal acceptance rule is defined as follows:

\[
D^*(\Gamma^\pi_t, Q^\pi_t, j) = \begin{cases} 
1 & \text{if demand } j \text{ is feasible} \\
0 & \text{if demand } j \text{ is not feasible}
\end{cases}
\]

\[
= \begin{cases} 
\text{if } R_j > E\left[ \sum_{i} D^\pi_i (R_i - \beta(\phi^\pi_i + l_i) - \gamma \xi^l_i (\delta^\pi_i - \tau_i))^+ \mid \Gamma^\pi_t, Q^\pi_t \right] - \\
\text{or } R_j \leq E\left[ \sum_{i} D^\pi_i (R_i - \beta(\phi^\pi_i + l_i) - \gamma \xi^l_i (\delta^\pi_i - \tau_i))^+ \mid \Gamma^\pi_t, Q^\pi_t \right] - \\
E\left[ \sum_{i} D^\pi_i (R_i - \beta(\phi^\pi_i + l_i) - \gamma \xi^l_i (\delta^\pi_i - \tau_i))^+ \mid \Gamma^\pi_t, \overline{Q}^\pi_t \right]
\end{cases}
\]
If load $j$ is accepted, the expected profit is estimated over the demands that will be requested during the time horizon $(A_j, T)$ not counting load $j$, given state of the system represented by vehicles’ status and locations as well as the updated routing schedule including demand $j$:

$$
E\left[ \sum_{i} A_j < A_s \leq T \sum_{i} D_i^s (R_i - \beta(\phi_i^s + l_i) - \gamma \xi_i l_i(\delta_i^s - \tau_i)^+) \big| \Gamma_i^s, \bar{Q}_i^s \right]
$$

Otherwise, if the load is rejected, in this case the state of the system is identical to the state just before the demand comes in, the expected total profit is:

$$
E\left[ \sum_{i} A_j < A_s \leq T \sum_{i} D_i^s (R_i - \beta(\phi_i^s + l_i) - \gamma \xi_i l_i(\delta_i^s - \tau_i)^+) \big| \Gamma_i^s, Q_i^s \right]
$$

Hence, the expected marginal total profit ($EMP$) by rejecting the load $j$ is:

$$
EMP(j) = E\left[ \sum_{i} A_j < A_s \leq T \sum_{i} D_i^s (R_i - \beta(\phi_i^s + l_i) - \gamma \xi_i l_i(\delta_i^s - \tau_i)^+) \big| \Gamma_i^s, Q_i^s \right] - E\left[ \sum_{i} A_j < A_s \leq T \sum_{i} D_i^s (R_i - \beta(\phi_i^s + l_i) - \gamma \xi_i l_i(\delta_i^s - \tau_i)^+) \big| \Gamma_i^s, \bar{Q}_i^s \right]
$$

The $EMP(j)$ represents the expected marginal profit that the capacity of the system, which is held by rejecting demand $j$, would produce by serving future demands. In other words, the carrier may lose the expected profit by accepting load $j$, because accepting demand $j$ may prevent the system from accepting future demands that would generate more profit. If the system has a few loads in the queue so that the system has enough room for accepting future demands, the expected marginal profit by rejecting demand $j$ is negligible. In contrast, as the system is saturated with the demands that have been accepted but not-yet-picked up, the expected marginal profit by rejecting demand $j$ tends to increase. Therefore, the optimal acceptance decision criteria for demand $j$ is as follows: At first, the load $j$ should be feasible. In addition, the revenue earned
from accepting the load $j$ should be greater than the expected marginal profit by rejecting demand $j$.

This acceptance decision rule makes an intuitive sense. However, it is extremely difficult to estimate the expected total profit over the future (unrealized) demands with a given system state, which is described by locations and status of vehicles along with associated scheduled demands. Furthermore, even if it is possible to estimate the expected profit, the computation time of the estimation process is another restriction under this dynamic operating environment since the acceptance decision on a requested demand should be informed to the customer in a short time. Therefore, the acceptance decision policy for this dynamic fleet management problem should be able to make a decision in a short time, and should take into account the system state upon arrival of a load.

In this problem of interest, customers can choose the service type out of two mutually exclusive alternatives. The priority service is more expensive than the regular service in a way that the price rate (unit price per unit distance) of a priority demand ($r_{\text{priority}}$) is much higher than the regular demand’s unit price ($r_{\text{regular}}$). Furthermore, although a long-haul regular demand can generate the same revenue as a short-haul priority demand, the priority load is more valuable to the carrier than the regular load since the priority demand consumes less resource of the fleet. Short haul-length corresponds to less operating cost, and the short service (delivery) time provides the opportunity that the driver can serve additional demands. Thus, the dispatcher favors priority demands.

Therefore, the basic approach to maximize the overall profit is to accept as many priority demands as possible while controlling the acceptance decision of the regular demands. This control process aims to manage the state of the system, under which the carrier can accept future priority demands without the risk of underutilizing the transportation resources. In addition, this acceptance decision policy may improve the efficiency of the reassignment procedure by keeping total number of demands in the system under a critical point (2).

One of the key features of a demand that affects the system ability to accept a demand is the time-window width of the demand. Thus, even under the identical system state, the probability that the system can accept an unknown priority demand is different from that of a regular demand. Therefore, rejecting one regular demand under consideration cannot guarantee that the system can accept a future priority demand. Furthermore, relative location of a requested demand
with respect to the locations of the fleet of vehicles in the system is another important factor that affects the feasibility of the demand. However, the location of a future priority demand is not known when a decision maker has to make an acceptance/rejection decision on the requested regular demand in hand. Therefore, the system ability to accept a (unrealized) priority demand is represented in terms of the probability rather than the number of discrete slots.

**Feasibility Index**

This section presents the ‘Feasibility Index’ \( FI \), which represents a system state upon arrival of a regular arrival in terms of the approximation of the expected number of vehicles that can serve an unrealized (future) priority demand. Due to the complexity of the probability estimation (the probability that the system can accept a future priority demand), approximation algorithm is developed based on several assumptions.

First, it is assumed that the time-window width of a priority demand is known in advance. This assumption is applicable because the service provider, carrier, can pre-specify the delivery service characteristics.

Second, it is assumed that the arrival time of a future priority demand is equivalent to the time that \( FI \) is estimated. In other words, a priority demand is requested immediately after arrival of a regular demand. This assumption stems from the fact that, until the next demand comes in, the ability of the system accepting a future priority demand tends to improve gradually in the course of time, since the vehicles serve the accepted demands in continuous basis. Thus, \( FI \) provides the lower bound of the expected number of vehicles that can accept a future priority demand.

Third, the estimation process exploits an insertion heuristic method to explore the feasible routing schedule involving the future priority demand. Note that feasibility of a demand should be assured by feasible routing schedule including the demand. Even though this insertion heuristic based feasibility check process does not explore all possible routing schedules, it provides the lower bound. The feasibility check process employed in this thesis, the initial assignment procedure allowing resequencing of the demands in a vehicle queue, would demonstrate better performance than the acceptance decision rule based on the \( FI \). In other words, acceptance decision policy based on the initial assignment, in some cases, would be able to accept a demand that is rejected by the acceptance rule based on insertion heuristics.
Finally, the $FI$ estimation process involves probabilistic information about haul-length of a priority demand and the required empty distance to serve it. The probability density function of the haul-length can be obtained from the simulation experiment historical data. Furthermore, a threshold ($\beta$) is pre-specified representing the maximum allowable empty distance required to serve a demand. The empty distance value is essential when estimating $FI$ using the insertion heuristic method. The longest possible empty distance, in this problem setting 141.4 miles, can be applied to provide the lower bound. This value, however, assumes an extremely rare case, in which all vehicles locate at a corner of the geographic region under consideration and the new demand’s origin location is the opposite corner. The possibility of this event is negligible. Furthermore, $FI$ value estimated using the largest value significantly underestimates the expected number of vehicles that can serve a future priority demand. Therefore, average plus one standard deviation of the empty distance required to serve a demand (obtained from the simulation experiment historical data) would be used as the threshold ($\beta$).

The estimation process based on the assumptions stated above is as follows.

$$FI(k) = FI(k \mid \Gamma^\pi(k,t), q^{k,\pi}(t))$$

$$= \begin{cases} 
\text{if } s^{k,\pi} = 1 \text{ (loaded)} & \max(\{FI(k,(q_i^{k,\pi}, q_{i+1}^{k,\pi})), \text{ for } i = 1, \ldots, I - 1\}) \cdot FI(k, q_i^{k,\pi}) \\
\text{if } s^{k,\pi} = 2 \text{ (empty)} & \max(FI(k, q_i^{k,\pi}), \{FI(k,(q_i^{k,\pi}, q_{i+1}^{k,\pi})), \text{ for } i = 1, \ldots, I - 1\}) \cdot FI(k, q_i^{k,\pi}) \\
\text{if } s^{k,\pi} = 3 \text{ (idle)} & FI(k) 
\end{cases}$$

where $I$ represents queue length ($I = |q^{k,\pi}(t)|$), and for notational simplicity current time $t$ is omitted and $FI(k, (q_i^{k,\pi}, q_{i+1}^{k,\pi})) = FI^L(k, (q_i^{k,\pi}, q_{i+1}^{k,\pi})) \cdot FI^E(k, (q_i^{k,\pi}, q_{i+1}^{k,\pi}))$

$$FI(t \mid \Gamma^\pi, Q^\pi) = \sum_{k=1}^{K} FI(k \mid \Gamma^\pi(k,t), q^{k,\pi}(t))$$

The first step is to estimate, $FI(k)$, the probability that single vehicle $k$ can accept a future priority demand, in which the current location, status ($\Gamma^\pi(k,t)$) of vehicle $k$ and associated
scheduled demands \((q^{k,\pi}(t))\) are known to the dispatcher. This process is based on the given system state. There are two main features of the system that affects the \(FI(k)\). The first feature is the updated location of vehicle \(k\) considering the vehicle status. The other feature is the set of times, \(\{(\tau^+_i - t), i = \{ q^{k,\pi}(t) \}\}\), where \(t\) represents the decision epoch. This set of times is obtained from the latest pickup times \((\tau^+_i)\) of the demands, which are schedule in vehicle \(k\)'s queue \(i = \{ q^{k,\pi}(t) \}\). Considering those features, estimation process of \(FI(k)\) begins with the location-based Feasibility Index of vehicle \(k\) \((FI^L(k))\). Suppose that an idle vehicle locates in a square region, and the time-window width of a priority demand is known in advance as \(\tau\) \((\tau^+_i - \tau^-_i)\). In this case, the vehicle can serve a requested demand only if the vehicle can reach the origin of the requested demand within \(\tau\). In addition, a pre-specified threshold \(\beta\) is applied to limit the maximum empty distance allowed to serve the demand. Thus, \(\tau' = \text{min}(\tau, \beta)\) is the updated threshold. In other words, the \(FI^L(k)\) for an idle vehicle is the probability that origin of a priority demand is generated in the area where the vehicle can travel within the time-window of the demand and the area is also restricted by the threshold \(\beta\).

If an idle vehicle located at a square area as shown in the Figure2, the vehicle can serve a future demand only if the origin of the load is generated within overlapped shaded area of the rectangular and the circle. Since the origin of a demand is uniformly distributed over the square region in this problem setting, the location-based Feasibility Index of vehicle \(k\), \(FI^L(k)\) is represented by the size of the shaded area over the size of the rectangular area.

If a vehicle is not idle, the location and \(FI^L(k)\) of the vehicle should be updated considering the scheduled demands associated with the vehicle. For example, when a vehicle (with only one demand in its queue) is loaded status, namely on the way to the destination of a load, the vehicle will be available at the destination of the current load. Therefore updated \(\tau' = \text{min}(\tau, \beta)\) should be reduced taking into account available time of the vehicle, and \(FI^L(k)\) should also accommodate the update location of vehicle \(k\). The Figure 3 depicts the reduction of \(\tau'\) to \(\tau''\) where solid line represents the scheduled movement of vehicle \(k\). In this figure, it is assumed that load \(a\) is already picked up and should be delivered to its destination. Of course, the threshold \(\beta\) is already considered when \(\tau\) is updated to \(\tau'\). Note that, calculation of the \(FI^L(k)\) assumes that a new priority demand with time-window width \(\tau\) comes into the system immediately after current
decision process. This process provides the lower bound of \( \tau' \), since as the next demand arrival time is late, the value of the \( \tau' \) increases.

If a vehicle has more than two demands in its queue (or one demand and empty status), the probability that a priority demand can be inserted into existing service schedule should be taken into consideration along with the location based estimation process. In this case, the time-windows (particularly the latest times) of the existing demands in the queue (\( \tau^+_i, i = \{ q^{k-\tau}(t) \} \)) are important factors that affect the Feasibility Index. This type of FI is called time-window based Feasibility Index (\( FIT(k) \)). For example, a vehicle has scheduled three loads (\( a, b, c \)) sequentially as shown in the Figure 4. The possible insertion slots for a new demand are (\( a, b \)), (\( b, c \)) and end of the queue, where (\( a, b \)) denotes the slot between demands \( a \) and \( b \).

The maximum time-space room between loads \( a \) and \( b \) can be obtained by shifting the scheduled pick up times of load \( b \) and \( c \) up to the points that their time-windows allow. Let \( S(a,b) \) be the maximum allowable time-space between demand \( a \) and \( b \).

\[
S(a,b) = TO(b) - TD(a) + \min\{ \tau^+_b - TO(b), \tau^+_c - TO(c) \}
\]

\(TO(j)\): scheduled time at the origin of load \( j\)
\(TD(j)\): scheduled time at the destination of load \( j\)

In order to insert a demand between demand \( a \) and \( b \), the maximum time-space \( S(a,b) \) should be greater than the haul-length of the (future, unrealized) demand plus required empty distance to serve it. For the haul-length, a random variable (\( \eta \)) is defined following a probability density function (pdf) obtained from the historical data. With respect to the required empty distance to serve the demand, average plus one standard deviation value from the historical data would be used as the maximum allowable empty distance (\( \beta \)).

\[
FIT(k, (a,b)) = Pr\{ S(a,b) > \eta + \beta \}
\]

Where \( FIT(k, (a,b)) \) denotes the probability that a priority demand can assigned between demands \( a \) and \( b \). In addition to this time-window related probability, the previously presented
location based probability $F I^E(k, (a,b))$ is also taken into account with updated $\tau$ considering updated vehicle location (in this case, at the destination of the load $a$).

$$F I(k, (a,b)) = F I^E(k, (a,b)) * F I^E(k, (a,b))$$

This procedure is applied into every available slot including the case of appending the new demand at the end of the queue, $F I^E(k, c)$ such that the vehicle’s updated location is the destination of demand $c$. Calculation process of $F I^E(k, c)$ regards the demands $a, b$ and the empty movement between the two demands as a single long demand and consider updated location of the vehicle. Thus, the feasibility index of vehicle $k$, $F I(k)$ is as follows.

$$F I(k) = \max \{ F I(k, (a,b)), F I(k, (b, c)), F I(k, c)\}$$

This estimation procedure explores all available slots of vehicle $k$, to which a new demand can be inserted depending on the vehicle status. Note that if a vehicle is empty moving status, the slot between the vehicle and the first demand should be considered. This slot represents the en-route diversion case.

Finally, feasibility index of the system is the sum of $F I(k)$’s for all vehicles in the fleet, which represents the expected number of vehicles that can accept a future priority demand with empty movements less than $\beta$.

**FI Based Acceptance Decision Policy**

Whenever a demand arrives, the feasibility-check process presented in (1,2) is applied. If the demand is feasible and is a priority demand, it is accepted outright. Otherwise, if the demand is a regular demand, initial routing schedule involving the new load is constructed. Then, $F I$ is estimated based on this updated state of the system. Only if the $F I(t | \Gamma^x_i, Q^x_i)$ is greater than specified threshold ($F I^*$), the requested demand is accepted. Otherwise, if the $F I(t | \Gamma^x_i, Q^x_i)$ is less than $F I^*$, the load is rejected in order to increase $F I$. 
NUMERICAL RESULTS

The simulation framework for the numerical experiment is as follows: It is assumed that a company consisting of 100 trucks, which receives demands for 10 hours a day and works until the fleet completes delivery service for all accepted demands. It is assumed that the trucks have a constant speed of 50 mph. The geographical area, where the company works, is assumed to be a 100 miles by 100 miles square region. The origins and destinations of the demands, which are requested over time independently, are distributed uniformly and independently over the region. The time-window of a demand is specified such that the earliest pickup time is the same as the demand arrival time, and the latest pickup time depends on the demand type. For the priority demands, type I time-windows are employed with fixed 1.5 hour length \((\tau^*_i - \tau^*_i)\) for express and on-time delivery service. The regular demand employs the Type II time-window with total 4 hours duration \((\tau^*_i - \tau^*_i)\), of which one hour is reserved for over time \((\delta - \tau^*_i)\). If the pick up of regular demand \(i\) occurs after its critical time \((\delta > \tau^*_i)\), then a penalty in proportion to \((\delta - \tau^*_i)\) and the haul-length \((l_i)\) is charged. The maximum penalty is charged when the demand is picked up at the latest pickup time of the demand, \(\delta_i = \tau^*_i\). In this case, 20% of the revenue earned from demand \(i\) is the maximum penalty, which is specified by the scaling parameter, \(\gamma\).

The situation analyzed in this paper assumes a demand rate that produces an over saturated system. Therefore, the demand arrival rate is specified so that about 30% of the demands are rejected even with an efficient dispatching algorithm. This results in the average inter-arrival time of successive demands being around 18 seconds.
Two scenarios are tested with different distribution ratios between two demand types. In the first scenario, 6.25% of the requested demands are the priority demands. On the other hand, the second scenario evaluates the case that 25% of the requested demands are the priority demands. The experiments simulate 5 days operating of the fleet where 1265.3 (±14.2) demands including both types are requested in a day, and 77.3(±4.8) and 310.3 (±11.2) demands are priority demands, respectively.

Two benchmark policies are tested to evaluate the performance of the $FI$ acceptance decision policy. The first benchmark policy accepts as many demands as possible, regardless of the demand type i.e., feasibility-based-acceptance rule. The second benchmark is the simple filtering process presented in (2). In other words, whenever a new demand arrives, the number of demands in the system (the accepted demands that have not been picked up yet) is counted. Then, only if the number is below a predefined threshold is the demand accepted.

To find the optimal $FI^*$ threshold value, calibration process with various $FI^*$ values are conducted. Figure 5 illustrates the results with various $FI^*$ values in both scenarios. The optimal threshold values showing best performance are 7 and 6 for 25% and 6.25% of priority demand portions, respectively.

Table 1 shows the numerical results with $FI^*$ values for the two scenarios. As seen in the table, in terms of the total profit (column 5), the performance of the dispatching system significantly improves with the proposed $FI$ acceptance decision policy compared to both of the benchmarks in both scenarios. The sixth and the last columns present the total number of served demands and served priority demands respectively along with the acceptance ratios. In spite of the similar acceptance ratio with respect to the total number of accepted demands, the more accepted priority demands improves the total profit significantly. For example, in the 25% priority demand scenario, although the FI-based acceptance decision policy accepts less total number of demands (850.6, 67.2%) compared to feasibility-based acceptance decision policy (900.2, 71.1%), the high acceptance ratio of the priority demand (86.1%) produces much greater overall profit ($38,438) than the feasibility-based acceptance decision policy case ($23,307). Furthermore, it is notable that the over time cost (column 3) is successfully reduced by controlling the number of demands in the system with filtering process and FI based acceptance decision policy.
The evolution of the total number of demands in the system over time within a day is shown in 6. For the presentational simplicity, only one day’s performance of the high portion of priority demands (25%) scenario is presented, the shaded line represents the total number of demands in the system over time with the feasibility-based-acceptance policy, the solid line with the simple filtering, then bolded line with the $FI$-based acceptance policy. After the 8th hour, the company does not accept any new demands, and focuses on serving the demands that has been accepted. The solid line shows the effect of the filtering process with a threshold of 250 demands, so that all demands received when the system is at this level are rejected outright until 6.5th hour. With respect to the $FI$-based acceptance decision policy, as the portion of priority demands out of the total demands in the system increase, $FI$ value at time $t$ decreases due to the decrease of the average time-window width over the demands in the system. Therefore, in order to maintain the $FI(t)$ value at a certain level, the total number of demands in the system decreases by rejecting regular demands. Note that, these acceptance decision policies were applied until the 6.5th hour in order to hold as many demands as possible at the 8th hour to increase the number of demands to be served in a day.

CONCLUSION

An MIP formulation is presented, which models the static version of the truckload truck routing and scheduling problem with time-window constraints under situation that demands are classified into two types corresponding to various customer requirements. Then, for oversaturated demand condition, a dynamic acceptance/rejection decision policy considering system state at a decision moment is developed as a part of dynamic dispatching system utilizing the above model. In this acceptance decision policy, developed decision criterion, $FI$ represents the approximation of the expected number of vehicles being able to serve future (still unrealized) priority demand based on the current system state, in which the maximum required empty distance parameter ($\beta$) is pre-specified. This acceptance decision policy evaluates the system state in a short time and provides a decision criterion for acceptance and rejection decision.

Conducted simulation experiment results show that the $FI$ based acceptance decision policy significantly improves the total profit over the benchmark policies. This improvement is mainly caused by the increased portion of accepted priority demands out of the accepted demands even
though the total number of accepted and served demands is similar. Furthermore, this policy significantly reduces the delay (over time) cost with respect to the regular demand service.

Even though the acceptance decision policy is implemented to the specific problem setting, the insight from this work can be expanded to wide range of dynamic and stochastic fleet management problems.

The recommended future research would consider multiple pickups and deliveries for less-than truckload delivery service considering the truck capacity constraints. Moreover, explicit consideration of variable travel time of the vehicles due to unpredictable events such as network congestion and/or accidents, are also recommended for future study.

REFERENCES


LIST OF TABLES AND FIGURES

Six Figures and one table

FIGURE 1 TWO TYPES OF TIME-WINDOW CONFIGURATIONS

FIGURE 2 LOCATION BASED $FI$ OF IDLE VEHICLE $k$

FIGURE 3 UPDATE PROCESS OF $\tau'$ TO $\tau''$ WHEN VEHICLE $k$ IS LOADED STATUS WITH ONE DEMAND, LOAD $A$, IN IT’S QUEUE

FIGURE 4 MAXIMUM ALLOWABLE TIME-SPACE BETWEEN LOADS $a$ AND $b$

FIGURE 5 PERFORMANCE WITH VARIOUS $FI^*$ THRESHOLDS

FIGURE 6 COMPARISON OF VARIOUS ACCEPTANCE DECISION POLICIES WITH RESPECT TO THE TOTAL NUMBER OF DEMANDS IN THE SYSTEM OVER TIME

TABLE 1 NUMERICAL RESULTS OF THE $FI$ BASED ACCEPTANCE DECISION POLICY
FIGURE 1 TWO TYPES OF TIME-WINDOW CONFIGURATIONS
FIGURE 2 LOCATION BASED $FI$ OF IDLE VEHICLE $k$
FIGURE 3  UPDATE PROCESS OF $\tau'$ TO $\tau''$ WHEN VEHICLE $k$ IS LOADED STATUS WITH ONE DEMAND, LOAD $a$, IN IT'S QUEUE
FIGURE 4 MAXIMUM ALLOWABLE TIME-SPACE BETWEEN LOADS a AND b
FIGURE 5 PERFORMANCE WITH VARIOUS $FI^*$ THRESHOLDS
FIGURE 6 COMPARISON OF VARIOUS ACCEPTANCE DECISION POLICIES WITH RESPECT TO THE TOTAL NUMBER OF DEMANDS IN THE SYSTEM OVER TIME
TABLE 1 NUMERICAL RESULTS OF THE FI BASED ACCEPTANCE DECISION POLICY

(a) 6.25% of demands are priority demand

<table>
<thead>
<tr>
<th>Acceptance decision policy</th>
<th>Total Revenue (std.)</th>
<th>Total Cost (std.)</th>
<th>Over Time Cost (std.)</th>
<th>Total Profit (std.)</th>
<th>Total # of Served Demands (std.)</th>
<th># of served Priority demands (std.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feasibility based acceptance</td>
<td>$56,663 (618)</td>
<td>$32,312 (229)</td>
<td>$5,340 (363)</td>
<td>$19,011 (549)</td>
<td>896.8 (13.2)</td>
<td>25.8 (33.4)</td>
</tr>
<tr>
<td>Simple Filtering</td>
<td>$57,295 (412)</td>
<td>$31,937 (449)</td>
<td>$1,249 (205)</td>
<td>$24,109 (484)</td>
<td>920.8 (10.1)</td>
<td>35.2 (45.6)</td>
</tr>
<tr>
<td>FI</td>
<td>$59,879 (460)</td>
<td>$31,820 (184)</td>
<td>$979.2 (168)</td>
<td>$27,080 (453)</td>
<td>897.2 (11.2)</td>
<td>66.3 (85.6)</td>
</tr>
</tbody>
</table>

(b) 25% of demands are Priority demands

<table>
<thead>
<tr>
<th>Acceptance decision policy</th>
<th>Total Revenue (std.)</th>
<th>Total Cost (std.)</th>
<th>Over Time Cost (std.)</th>
<th>Total Profit (std.)</th>
<th>Total # of Served Demands (std.)</th>
<th># of served Priority demands (std.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feasibility based acceptance</td>
<td>$56,663 (618)</td>
<td>$32,312 (229)</td>
<td>$5,340 (363)</td>
<td>$19,011 (549)</td>
<td>896.8 (13.2)</td>
<td>25.8 (33.4)</td>
</tr>
<tr>
<td>Simple Filtering</td>
<td>$57,295 (412)</td>
<td>$31,937 (449)</td>
<td>$1,249 (205)</td>
<td>$24,109 (484)</td>
<td>920.8 (10.1)</td>
<td>35.2 (45.6)</td>
</tr>
<tr>
<td>FI</td>
<td>$68,829 (1090)</td>
<td>$30,038 (535)</td>
<td>$352 (98)</td>
<td>$38,438 (804)</td>
<td>850.6 (13.2)</td>
<td>267.2 (86.1)</td>
</tr>
</tbody>
</table>