The freight allocation problem with lane cost balancing constraint

Andrew Lim\textsuperscript{a}, Hu Qin\textsuperscript{b,a,*}, Zhou Xu\textsuperscript{c}

\textsuperscript{a}Department of Management Sciences, City University of Hong Kong, Tat Chee Ave, Kowloon Tong, Hong Kong
\textsuperscript{b}School of Management, Huazhong University of Science and Technology, No. 1037, Luoyu Road, Wuhan, China
\textsuperscript{c}Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

\section{Introduction}

This study is motivated by a project awarded by a buying office for one of the largest retail distributors in the world, over 2000 outlets in 33 countries in Europe, Africa and Asia. The buying office (henceforth referred to as the shipper) annually procures diverse products, from textiles and foodstuffs to major electrical appliances, from over one thousand suppliers across Asia to satisfy the demands of its parent company’s sales divisions in over twenty European countries. Long-distance ocean shipping is the main transportation mode for the shipper for the delivery of the procured products, accounting for around 95% of its total annual turnover on average. The transportation network under the purview of the shipper comprises more than 1000 shipping lanes connecting 71 loading ports in Asia and 27 discharging ports in Europe.

This study is the first to consider the lane cost balancing constraint in the context of freight allocation. We formulate the freight allocation problem with this lane cost balancing constraint as a mixed integer programming model, and show that even finding a feasible solution to this problem is computationally intractable. Hence, in order to produce high-quality solutions in practice, we devised a meta-heuristic approach based on tabu search. Experiments show that our approach significantly outperforms the branch-and-cut approach of CPLEX 11.0 when the problem increases to practical size and the lane cost balancing constraint is tight. Our approach was developed into an application that is currently employed by decision-makers at the buying office in question.

This paper examines the problem of allocating the freight quantity for all lanes to shipping companies (henceforth referred to as carriers) such that the total transportation cost is minimized. The shipper performs this freight allocation at the strategic level as part of the process of procuring transportation services from carriers. At the beginning of every fiscal year, the shipper forecasts the total quantity of freight (demand) for the coming year on each lane. The forecasted demand of each lane is given to the carriers through its web portal, who then return with their price quotations and minimum quantity commitments (MQC). The MQC is the minimum total freight quantity over all lanes for a contracted carrier; this is a standard requirement imposed by carriers to guarantee a minimum freight volume, and is in fact stipulated by the U.S. Federal Maritime Commission when transporting goods into U.S. cities (Lim et al., 2006).

With this collected information, the shipper performs additional scenario-based analyses that help to direct its negotiations with each carrier. For instance, a carrier that handles a large proportion of freight on important shipping lanes would justify major negotiations (and corresponding concessions) since a reduction in its price quotation would have a significant effect on the transportation cost. The shipper might also consider reallocating demand between lanes. With hundreds of lanes and dozens of carriers,
the number of scenarios examined is considerable, and each sce-
nario might necessitate the solution of a separate freight allocation
problem. It usually takes several iterations of this negotiation and
analysis process before the final decisions are made and freight
volume on each lane is allocated to a subset of candidate carriers;
each selected carrier is given a one-year contract, and during the
contract period the freight rates quoted by the carriers are fixed
regardless of market fluctuations.

The shipper must consider several factors when allocating
freight volume to carriers. Firstly, as a safeguard against the
inability of a carrier to fulfill its contractual obligations, the
shipper specifies a minimum number of carriers for each lane. Sec-
ondly, since the 1990s, carriers have formed strategic alliances to
realize inter-organizational cooperation to achieve operational
synergies, gain benefit from economies of scale, and increase
market power (Song and Panayides, 2002); to prevent over-reli-
ance on a particular carrier or alliance, the shipper sets quantity
limits on the freight allocated to a single carrier or alliance on
each lane.

Thirdly, the shipper wishes to limit disproportionate freight
charges on all lanes. Although all the sales divisions are members
of the parent company, they are financially independent, i.e., staff
bonuses are commensurate with the profits and performance within
their respective sectors. Hence, each sales division can be con-
sidered an independent buyer. The procurement cost paid by a
buyer primarily consists of the production cost to the suppliers,
freight cost paid to the carriers, and the commission fee charged
by the shipper. For the shipper, the commission fees represent its
only profit source, and therefore it is important to maintain good
relations with each buyer. Only the shipping rates and allocation
results are disclosed to the buyers, and a buyer operating on a par-
ticular lane can compare its freight costs with those of other lanes;
if it is found that the freight cost is disproportionate, then the
buyer has reason to be dissatisfied. At the very least, the freight
cost arranged by the shipper must significantly offset the best price
that the retailer can obtain from the carriers on the open market in
order to justify the commission. This has been a problem with the
manually allocated freight assignments in previous years.

In this study, we formulate the freight allocation problem with
the lane cost balancing constraint (FAPLBC) as a mixed integer pro-
gramming model, and show that achieving an optimal solution and
identifying a feasible solution are both NP-hard in the strong
sense; the lane cost balancing constraint increases the difficulty
of the problem. We then design and implement a customized tabu
search algorithm for this problem, which we evaluate using both
real and generated instances. This algorithm was developed into
an application that is currently employed by the shipper for its
freight allocation planning decisions.

The rest of this paper is organized as follows. We first provide
an overview of related existing research in Section 2. In Section 3,
we state the FAPLBC precisely, formulate it as a mixed integer pro-
gramming model, and examine its computational complexity.
Next, we describe our customized tabu search algorithm for this
problem in Section 4 that includes a random move heuristic for
escaping from local optima, and also discuss how the output for
this problem is presented in a useful form such that the decision-
makers at the shipper can easily interpret the data. Section 5 gives
the results of our experiments and Section 6 concludes this study.

2. Related works

The freight allocation problem can be viewed as a transporta-
tion service procurement problem. When buying transportation
services, a shipper typically adopts combinatorial auction mecha-
nisms that allow carriers to quote a price for a package of lanes so
that carriers can take better advantage of economies of scope (Caplice and Sheffi, 2003; Sheffi, 2004; Caplice and Sheffi, 2005).
Many shippers, such as Sears Logistics Services, The Home Depot,
The Limited, Inc., K-Mart Corporation and Ford Motor Company,
have successfully achieved cost savings when procuring transpor-
tation services using combinatorial auctions (Ledyard et al., 2002;
Elmaghraby and Keskinocak, 2003; De Vries and Vohra, 2003). The transportation service providers can construct effective bid-
ing packages of lanes to exploit complementation of volumes such
that the costs for both shipper and carrier can be reduced (Song and Regan, 2004; Song and Regan, 2005). In particular, as
pointed out by Song and Regan (2003), combinatorial auctions
can help small or medium-sized transportation service providers
significantly balance their transportation networks by exchanging
inefficient lanes.

Another class of procurement models mainly considers the
MQC constraint and has been used to procure transportation ser-
vices by many shippers (Lim et al., 2006). The MQC constraint has
been widely studied in the analysis of supply contracts (Bassok
and Anupindi, 1997; Chen and Krass, 2001; Bassok and Anupindi,
2008; Lian and Deshmukh, 2009). Guha et al. (2000) studied an
MQC application in which a franchise must open stores to mini-
mize their average distance from customers while ensuring a min-
imum number of customers to make each store profitable. Lim
et al. (2005) examined an extension of the k-center facility location
problem in which each supplier must cover a minimum number of
clients. A new bottleneck problem with the MQC constraint was
introduced by Lim and Xu (2006), where the task is to assign all cli-
ents to suppliers such that each selected supplier has a minimum
number of clients and the total cost of the selected suppliers is lim-
itied by a budget; the objective is to minimize the maximum dis-
tance between the clients and their suppliers.

Existing works on practical procurement problems with the
MQC constraint usually employ standard commercial software
such as LINGO to solve their models, e.g., Sadrian and Yoon
(1994) and Katz et al. (1994). In contrast, in their work on the
freight allocation problem with the MQC constraint, Lim et al.
(2006) focused on analyzing the structure of the mathematical
model and designing tailored algorithms to achieve high quality
solutions; they proposed a mixed integer programming model de-
fined by a number of strong facets and applied a branch-and-cut
scheme, a rounding heuristic and a greedy approximation method
in its solution. This work was extended by Lim et al. (2007) by
incorporating a fixed selection cost for each carrier, which makes
even finding a feasible solution is NP-hard in the strong sense.

Recently, a new transportation service procurement model was
proposed by Lim et al. (2008), where the shipper guarantees that
the shipping volume during non-peak season is commensurate with
that during peak season, where the goal is to reduce drastic
fluctuations in shipment volumes over the contract duration. This
model has been implemented in practice by a large Dutch electron-
ics manufacturer.

The problem studied in this paper also extends Lim et al. (2006),
by considering the lane cost balancing constraint. To the best of our
knowledge, the lane cost balancing constraint has not been inves-
tigated in previous research in a similar context.

3. Problem description and formulation

We model the problem faced by the shipper as follows. There
is a set of carriers \( I = \{1,2, \ldots, m\} \) and a set of lanes \( J = \{1,2, \ldots, n\} \). The projected demand for lane \( j \) is given by \( d_j \). The price quoted by carrier \( i \) to transport one twenty-foot equivalent unit of prod-
uct on lane \( j \) is \( p_{ij} \), and each carrier has a minimum quantity
commitment, denoted by \( b_i \), that defines the minimum quantity
that must be allocated to that carrier if it is selected. The task is to assign a non-negative freight quantity to each carrier-lane pair such that the demand for each lane is fulfilled at minimum cost.

The shipper requires that each lane must use a minimum number of carriers based on the characteristics of the lane such as traveling distance, annual quantity of freight and safety level of the route. By imposing a minimum number of carriers for each lane, the shipper protects against the possibility that there are no carriers with sufficient cargo hold space available for a shipment; otherwise, a long delivery delay may result since most carriers send vessels to visit a port only once a week (Agarwal and Ergun, 2008). However, simply assigning a minimum number of carriers for each lane is insufficient. For example, if we impose that the number of carriers for a certain lane must be at least 4, then the resulting allocation pattern may be 97%, 1%, 1%, 1%, which is almost the same as allocating all freight quantity to one carrier. Therefore, the shipper decided to model this requirement as a maximum carrier percentage allocation \( q_j^p \) of the freight quantity for each selected carrier on lane \( j \), i.e., the amount of freight quantity allocated to each carrier servicing lane \( j \) can be at most \( q_j^p - d_j \).

Similar applications of such quantity limit measures can be found in Sadrizadeh and Yoon (1994) and Goosens et al. (2007). The minimum number of carriers operating on lane \( j \) is denoted by \( n_j^p = \lfloor 1/ q_j^p \rfloor \).

We assume that all carriers are partitioned into \( g \) alliances, denoted by the set \( A = \{ A_1, A_2, \ldots, A_g \} \), where \( A_i \subseteq I \), \( \bigcup_{i=1}^{g} A_i = I \), and \( A_i \cap A_j = \emptyset \) for \( i \neq j \). Vessel sharing is a common form of inter-organizational cooperation among carriers in the same alliance (Song and Panayides, 2002). If the shipper allocates all freight of a certain lane to a carrier belonging to one alliance, it is likely that all selected carriers for that lane depart on the same shipping date. There are several disadvantages to having only a single shipping date: (1) there may be sailing delays or cancelations due to adverse weather or other factors on that date; (2) if the delivery to the loading port misses the departure date of the designated vessel, it must be stored at the loading port for the next available vessel, resulting in added costs; (3) a single shipping date restricts the scheduling options for both product manufacturing and in-land transportation, reducing flexibility for the suppliers and increasing their relative operation costs. Hence, we impose a maximum alliance percentage allocation \( q_j^A \) on the freight volume for alliances similar to the restriction on carriers, i.e., the amount of freight quantity allocated to all carriers in a single alliance servicing lane \( j \) can be at most \( q_j^A - d_j \).

The minimum number of alliances servicing lane \( j \) is likewise given by \( n_j^A = \lfloor 1/ q_j^A \rfloor \).

Finding a solution to the freight allocation problem can be divided into two stages: (1) select a subset of candidate carriers; and (2) allocate the freight for all lanes to these selected carriers. In the retail industry, transportation cost is a key determinant of a commodity’s final selling price. Hence, all buyers desire reduced transportation cost for their products in order to be able to offer a lower selling price to customers. For any selected carrier combination \( F \subseteq I \), let the price of the cheapest carrier servicing lane \( j \) be \( p_j^F = \min \{ p_k : k \in F \} \); from the buyers’ perspective, the lowest possible freight cost for a lane \( j \) under \( F \) is given by the theoretical lower bound \( \sum_{i \in A} p_i x_i \), which assumes that all freight for that lane is allocated to the cheapest carrier. However, this allocation is usually infeasible since it is likely to violate the MQC, carrier quantity limit, and/ or alliance quantity limit constraints.

When only the MQC and quantity limits are taken into account, an unbalanced allocation solution such as the one shown in Table 1 may result, where the actual freight cost for lane 1 is 2% more than its theoretical lower bound but the actual freight cost for lane 2 is 20% more. The cost ratio is calculated by: \((\text{actual cost-theoretical lower bound cost})/\text{theoretical lower bound cost}\). Since all buyers have complete information of the shipping rates for the selected carriers and the allocation results, the buyer associated with lane 2 is likely to be dissatisfied by the unfairness of this allocation, and this loss of goodwill may have an adverse effect when the shipper negotiates future commissions. To address this issue, the shipper can impose a \textit{lane cost balancing constraint} of the form “the cost ratio for each lane must be at most \( z \% \).” For example, when this constraint is imposed with balancing factor \( z = 10 \), then the allocation solution shown in Table 2 may be produced.

Note that the total transportation cost incurred with the lane cost balancing constraint is 1630, which is greater than the cost incurred of 1620 without the constraint. In effect, this constraint reduces the gap between the cost ratios of different lanes by shifting the transportation cost burden between buyers, resulting in an increase in freight costs as a whole. Although it would be theoretically preferable to minimize total transportation cost for the entire organization, and then to redistribute the savings to the buyers on the lanes with high freight costs, this is impractical for two reasons. Firstly, the overhead involved in the logistics of redistribution may be greater than the amount saved. Second, the allocation is based on forecasted demand that may be inaccurate due to market fluctuations, and an unbalanced allocation would exacerbate the effects of such market shifts. It is the opinion of the shipper that a fairer allocation justifies a small increase in total transportation cost. Our methodology allows the shipper to measure how much the transportation cost would increase for a representative range of cost ratio (\( z \)) values; experience with real-life problems has shown that the resulting cost increase is usually quite small.

We formulate the FAPCLB as the following mixed integer programming model:

\[
\text{MIP} \quad z = \min \sum_{j \in J} \sum_{i \in I} p_i x_{ij}, \quad (1)
\]

\[
\text{s.t.} \quad \sum_{i \in I} x_{ij} = d_j, \quad \forall j \in J, \quad (2)
\]

\[
b_i y_i \leq \sum_{j \in J} x_{ij}, \quad \forall i \in I, \quad (3)
\]

\[
x_{ij} \leq u_i y_i, \quad \forall i \in I, \quad j \in J, \quad (4)
\]

\[
\sum_{i \in I} x_{ij} \leq q_j A_k, \quad \forall A_k \in A, \quad j \in J \quad (5)
\]

\[
\sum_{i \in I} p_i x_{ij} \leq (1 + z\%) p_j d_j + (1 - y_j) M, \quad \forall j \in J, \quad i \in I \quad (6)
\]

\[
y_i \in \{0,1\}, \quad x_{ij} \geq 0, \quad \forall i \in I, \quad j \in J. \quad (7)
\]
where

\[ u_j = q^*_j - d_j \] the maximum amount of freight volume that can be allocated to a single carrier on lane \( j \);

\[ a_j = q^*_j - d_j \] the maximum total freight volume that can be allocated to carriers from a single alliance on lane \( j \);

\[ M \] a sufficiently large positive number;

\[ x_{ij} \] a nonnegative continuous variable representing the volume of freight transported by carrier \( i \) on lane \( j \);

\[ y_i \] a binary variable that is equal to 1 if carrier \( i \) is selected and 0 otherwise.

The objective function (1) minimizes the total transportation cost. Constraints (2) are the allocation constraints, which state that all demand must be allocated. Constraints (3) guarantee that each selected carrier is assigned at least \( b_i \) of freight, thereby fulfilling its MQC. Constraints (4) and (5) impose the quantity limit restrictions on carriers and alliances, respectively. Constraints (6) balance the cost ratios among all lanes such that the actual freight cost for each lane is at most \( \alpha \% \) more than its theoretical lower bound cost.

The FAPLCB is \( \mathcal{NP} \)-hard in the strong sense, which can be shown by constructing a restricted version of the FAPLCB as follows. Let the values of \( b_i = b \) for all \( i \in I \), and relax the quantity limits and lane cost balancing constraints by setting \( q^*_j = q^j = 1 \) and \( \alpha = M \). This transforms the problem into the Transportation Problem with MQC, which is \( \mathcal{NP} \)-hard in the strong sense (Lim et al., 2006).

If all prices \( p_j \) are identical, then the FAPLCB instance can be solved in polynomial time; otherwise, we can establish the following theorem.

**Theorem 1.** Finding a feasible solution to the FAPLCB is \( \mathcal{NP} \)-hard in the strong sense.

**Proof.** See Appendix A. \( \square \)

As shown in Lim et al. (2006), if the cost balancing and quantity limit constraints are relaxed, finding a feasible solution to the problem can be accomplished in polynomial time. Currently, we do not know if finding a feasible solution to the problem with only the lane cost balancing constraint relaxed is \( \mathcal{NP} \)-hard, but the following theorem suggests that the quantity limit on alliances and the lane cost balancing constraints have complicated the problem.

**Theorem 2.** If the quantity limit on alliances and the lane cost balancing constraints are relaxed, finding a feasible solution to the problem can be accomplished in polynomial time.

**Proof.** See Appendix B. \( \square \)

For ease of exposition, we use the notation \( \text{MIP}(\alpha) \) to represent the MIP model with the given \( \alpha \) value, and \( z(\alpha) \) to represent the optimal objective value of \( \text{MIP}(\alpha) \). Observe that we can relax the lane cost balancing constraints (6) by setting \( \alpha = M \); in practice, we can assume that \( \text{MIP}(M) \) has a feasible solution (which is realizable by setting the quantity limit parameters \( q^*_j \) and \( q^j \) to relatively high values). Furthermore, when \( \alpha = 0 \), then the only possible solutions to the problem must allocate all freight to the cheapest carrier on each lane, which is likely to be infeasible due to the violation of MQC or quantity limit constraints.

Let \( \alpha^- \) be the smallest value where \( \text{MIP}(\alpha^-) \) has a feasible solution and \( \alpha^+ \) be the smallest value of \( \alpha \) such that \( z(\alpha) = z(M) \). It is easy to observe that if \( \text{MIP}(M) \) has feasible solutions, \( z(\alpha) \) is a non-increasing function of \( \alpha \) in the interval \([\alpha^-, \alpha^+]\) because any feasible solution for \( \text{MIP}(\alpha) \) must also be a feasible solution for \( \text{MIP}(\alpha^-) \) for all \( \alpha > \alpha^- \). This observation implies that there exists a tradeoff between minimizing total transportation cost and minimizing \( \alpha \).

The MIP model can be modified to calculate \( \alpha^- \) as follows:

\[
\alpha^- = \text{MIP} \quad \alpha^- = \min \alpha \\
\text{s.t.} \quad (2)-(7) \\
\alpha \geq 0.
\]

We refer to this version of the problem as \( \alpha \)-FAPLCB. From Theorem 1, we can easily derive the following theorem.

**Theorem 3.** The \( \alpha \)-FAPLCB is \( \mathcal{NP} \)-hard in the strong sense.

**4. Solution methodology**

Theorems 1 and 3 imply that no polynomial-time algorithm exists to find an optimal solution to either \( \text{MIP} \) or \( \alpha \)-MIP unless \( \mathcal{NP} = \mathcal{P} \). In addition, preliminary experiments show that the commercial linear programming solver ILOG CPLEX is very slow when solving instances of this problem, especially for large instances where \( \alpha \) is close to \( \alpha^- \). Consequently, we developed a meta-heuristic approach based on tabu search (Glover and Laguna, 1997) to achieve high-quality solutions for instances of practical size; we call this approach the Random Move Tabu Search (RMTS).

**4.1. Random move tabu search**

Let \( I_1 \subseteq I \) be the set of selected carriers, i.e., \( I_1 = \{i; y_i = 1, i \in I\} \), and let \( I_0 = I - I_1 \) be the set of unselected carriers. The resultant model is given by:

\[
\text{MIP}(I_1) \quad z = \min \sum_{j \in J} \sum_{i \in I_1} p_j x_{ij} \\
\text{s.t.} \quad \sum_{j \in J} x_{ij} = d_i, \quad \forall j \in J, \\
\sum_{j \in J} x_{ij} \geq b_i, \quad \forall i \in I_1, \\
\sum_{i \in A_k \cap I_1} x_{ij} \leq a_j, \quad \forall A_k \in A, \quad j \in J, \\
\sum_{j \in J} p_j x_{ij} \leq (1 + \epsilon)M d_i \min\{p_j : i \in I_1\}, \quad \forall j \in J, \\
0 \leq x_{ij} \leq u_j, \quad \forall i \in I_1, \quad j \in J.
\]

A state in our search algorithm is defined by the set of selected carriers \( I_1 \) and its associated tabu information. We call \( I_1 \) a feasible carrier selection for the MIP model if \( \text{MIP}(I_1) \) has a feasible solution. A move in our algorithm consists of either removing a single carrier from or adding a single carrier to the current solution \( I_1 \). Given a current selection \( I_1 \), let \( z_i \) be the optimal objective value of \( \text{MIP}(I_1) \), where changing the state of carrier \( i \) transforms \( I_1 \) to \( I_i^* \). When \( \text{MIP}(I_i^*) \) has no feasible solutions, we set \( z_i = M \).
Algorithm 1.

Algorithm 1. Random move tabu search
1: Set the current iteration number \( k = 0 \); set the number of non-improving diversifications \( l = 0 \)
2: Randomly generate an initial carrier selection \( I_1 \)
3: Solve \( \text{MIP}(I_1) \) to find its optimal objective value \( z \); set the best objective value \( z_* = z \); set the iteration number when the best solution was found \( k_* = 0 \); set the attribute \( p_{t1} = 0 \)
4: Clear tabu information for all carriers by setting \( k_i = k_* - k_i - 1 \) for \( i \in I_1 \) and \( k_i = k_* - k_{00} - 1 \) for \( i \in I_0 \)
5: Increment \( p_{t1} \) for all carriers \( i \in I_1 \)
6: Compute \( z_i \) for all carriers and sort the carriers by increasing value of \( z_i \)
7: for each carrier \( i \) do
8: \( z_i = M \) then
9: Break
10: else if switching the state of carrier \( i \) is not tabu or it satisfies the aspiration criterion then
11: Switch the state of carrier \( i \) and set \( k_i = k \)
12: Break
13: end if
14: end for
15: If no valid move exists, invoke random move heuristic (Sub Section 4.2)
16: Increment \( k \)
17: If \( z_* < z \), then set \( z_* = z \), \( k = k \) and \( l = 0 \)
18: If \( k = k_* \), go to step 5
19: Set \( l = l + 1 \). If \( l < b \), generate \( I_1 \) using the diversification procedure, set \( k = k \) and go to step 4
20: Return \( z \)

The RMTS approach is outlined in Algorithm 1. The algorithm begins with a random set of selected carriers \( I_1 \) (line 2 in Algorithm 1). We solve \( \text{MIP}(I_1) \) using a commercial linear programming solver and record its optimal objective value in \( z \), representing the best solution found so far. We also keep a value \( k_i \) that stores the iteration number when the best solution was found (line 3 in Algorithm 1). Next, we evaluate all the neighbors of \( I_1 \) and sort them by increasing optimal objective value (line 6 in Algorithm 1).

We maintain two tabu lists, one for selected carriers and one for unselected carriers, with separate tabu tenure values. We use the notation \( z_i \) for the tabu tenure for the currently selected carriers. I.e., a carrier must remain selected for at least \( z_i \) iterations after it was selected by a move. Similarly, \( z_0 \) is the tabu tenure for the currently unselected carriers. Whenever the state of a carrier is changed, we record the iteration number as \( k_i \); hence, a move that changes the state of carrier \( i \) is not considered as tabu on iteration \( k \) if:

\[
 k - k_i > z_i \quad \text{if} \quad i \in I_1; \quad \text{or} \quad k - k_i > z_0 \quad \text{if} \quad i \in I_0.
\]  

If this condition is not satisfied, then we check if the aspiration criterion is fulfilled, which allows tabu moves to be performed if the solution that results is better than the current best solution, i.e.,

\[
 z_* < z. \]  

If either condition (14) or (15) is satisfied, we perform the move (line 10 in Algorithm 1); otherwise, the next move is considered in order of lowest objective value. If a solution that is better than the previous best solution is found, then \( z_* \) and \( k_* \) are updated. However, if all moves are invalid (line 15 in Algorithm 1), then we invoke a random move heuristic described in the next subsection.

This procedure is repeated until \( p_{t1} \) consecutive iterations have occurred where the best solution is not updated; \( p_{t1} \) is user-defined. At this point, we clear all tabu information and restart the algorithm with an initial solution that is produced by the following diversification procedure (line 19 in Algorithm 1). We maintain an attribute \( p_{t1} \) that records the number of iterations where carrier \( i \) is a member of the selected set, hence, \( k - p_{t1} \) is the number of iterations where carrier \( i \) was not selected. Given the current carrier partition \( (I_1, I_0) \), we select the \( \min\{p_{t1}(i), b_i/2\} \) carriers with the greatest \( p_{t1} \) from \( I_1 \) and the same number of carriers with the greatest \( k - p_{t1} \) from \( I_0 \), and switch the states of these carriers; the premise is to restart the search from a solution that is very different from those previously examined. The RMTS terminates when \( p_{t1} \) consecutive restarts are unable to improve on the best solution.

### 4.2. Random move heuristic

A move that changes the state of carrier \( i \) is considered invalid if it is tabu and does not fulfill the aspiration criterion, or if the resultant model \( \text{MIP}(I_1) \) has no feasible solutions. At some stage in the search, the RMTS algorithm may encounter the situation where all prospective moves are invalid (line 15 in Algorithm 1). When this occurs, the simplest option would be to randomly allow an invalid move to be performed. However, preliminary experiments show that this naive approach is likely to cause the search to be trapped in an infeasible region of the search space. Hence, we designed the following random move heuristic to help direct the search towards promising, feasible solutions.

The random move heuristic makes use of two auxiliary linear programming models derived from \( \text{MIP}(I_1) \) to rank invalid moves. The first model seeks to minimize the extent of the violation of the lane cost balancing constraint:

\[
 \text{LP1}(I_1) \quad \theta = \min \sum_{j \in J} \mu_j, \quad \text{s.t.} \quad (9) - (11), (13), \quad \mu_j \geq \sum_{i \in I_1} p_{ij} x_{ij} - (1 + \alpha%) d_j \min\{p_{ij} : i \in I_1\}, \quad \forall j \in J, \quad \mu_j \geq 0, \quad \forall j \in J. \tag{16}
\]

Aside from the objective function, the main difference between \( \text{LP1}(I_1) \) and \( \text{MIP}(I_1) \) is constraint (16), which measures the sum over all lanes of the disparity between the freight cost for a lane and its theoretical lower bound. Combined with the non-negativity constraint for \( \theta \), it is apparent that if \( \text{MIP}(I_1) \) has a feasible solution, then \( \text{LP1}(I_1) \) must have an optimal objective value of \( \theta = 0 \).

The second auxiliary model modifies \( \text{LP1}(I_1) \); it removes the lane cost balancing constraints, and its objective is to minimize the amount by which the MQC constraints are violated:

\[
 \text{LP2}(I_1) \quad \omega = \min \sum_{i \in I_1} v_i, \quad \text{s.t.} \quad (9), (11) \text{ and } (13), \quad v_i \geq \sum_{j \in J} x_{ij}, \quad \forall i \in I_1, \quad v_i \geq 0, \quad \forall i \in I_1. \tag{17}
\]

If \( \text{LP1}(I_1) \) has a feasible solution, it implies that all MQC constraints are fulfilled, and therefore \( \text{LP2}(I_1) \) must have an optimal objective value of \( \omega = 0 \). Since both of these auxiliary models are linear programming models with only continuous variables, they can be solved efficiently using a commercial linear programming solver.

Let \( \theta \) and \( \omega \) be the objective values for \( \text{LP1}(I_1) \) and \( \text{LP2}(I_1) \), respectively, where \( \theta_i \) is the result of switching the state of carrier \( i \) in the current selection. A value of \( \theta = M \) or \( \omega = M \) signifies that the corresponding model has no feasible solution in our implementation. The premise of our heuristic is to select a non-tabu move where the probability of a move being chosen is dependent on its \( \theta_i \) value; if \( \text{LP1}(I_1) \) is infeasible for all non-tabu moves, then
the probability of selection is based on its $c_i$ value instead. The MIP model may have many disjointed feasible regions; our random move heuristic permits the search process to enter into infeasible regions, and uses the auxiliary models to identify solutions that are close to feasible regions and examines them with greater probability. Hence, our tabu search process aims to conceptually visit several disjointed feasible regions by crossing infeasible regions within a reasonable number of iterations.

Algorithm 2. Random Move Heuristic

1: Initialize the set of non-tabu carriers $I' \leftarrow \{i: k - k_i > z_i \text{ if } i \in I_1 \text{ or } k - k_i > z_0 \text{ if } i \in I_0\}$
2: Calculate $c_i$ for all carriers $i \in I'$ by solving LP1($I_1$)
3: if all $\theta_i = M$ then
4: Calculate $c_i$ for all carriers $i \in I'$ by solving LP2($I_1$)
5: if all $c_0 = M$ then
6: Invoke local search procedure Algorithm 3)
7: Return
8: else
9: Sort the carriers by increasing value of $c_i$
10: end if
11: else
12: Sort the carriers by increasing value of $\theta_i$
13: end if
14: for each carrier $i \in I'$ do
15: Select carrier $i$ with probability $\tau$
16: if carrier $i$ is selected then
17: break
18: end if
19: end for
20: Switch the state of carrier $i$ and set $k_i = k$

Our random move heuristic is given in Algorithm 2. We first sort all non-tabu carriers by their $\theta_i$ values in ascending order; if this is not possible because LP1($I_1$) is infeasible for all these carriers, then we sort them by their $c_i$ values in ascending order. The first carrier has a probability of $\tau$ of being selected, where $\tau$ is a user-defined parameter; if it is not selected, then the next carrier will be selected with probability $\tau$, and so on, until a carrier is selected (the last carrier is selected if all previous carriers were not selected). Hence, the probability that the $k$th carrier is selected is the geometric probability $(1 - \tau)^{k-1} \tau$. The heuristic toggles the state of the selected carrier and updates its tabu information.

Algorithm 3. Local Search when LP2($I_1$) is Infeasible for All Moves

1: Let $n^f = \max_{i \in I_1} \{n_i^f\}$ and $n^b = \max_{i \in I_1} \{n_i^b\}$
2: Let $C(I_1) = \text{the number of carriers in set } I_1$; let $A(I_1) = \text{the number of alliances in set } I_1$
3: Clear tabu information for all carriers
4: while LP2($I_1$) is infeasible do
5: if $C(I_1) < n^f$ then
6: Randomly select a carrier $i$, i.e., update $I_1 \rightarrow I_1 \cup \{i\}$
7: else if $A(I_1) < n^b$ then
8: Randomly choose carrier $i_1 \in I_1$ from an alliance that has the largest number of carriers in $I_1$
9: Randomly choose carrier $i_2 \in I_1$ from an alliance that does not have carriers in $I_1$
10: Switch the states of carriers $i_1$ and $i_2$
11: end if
12: Update $k_i$ of the affected carriers
13: end while

On rare occasions, it is possible that LP2($I_1$) is infeasible for all non-tabu moves (line 5 in Algorithm 2). By the definition of LP2($I_1$), this can only occur if there are insufficient carriers to fulfill the demand for a lane, or the carriers belong to too few alliances to satisfy the minimum number of alliances for a lane. To resolve this issue, we employ the local search procedure given in Algorithm 3. First, all tabu information is cleared. Let $n_c$ be the largest minimum number of alliances for any lane, i.e., $n_c = \max_{i \in I} \{n_i^c\}$. While LP2($I_1$) is infeasible, we increase the number of carriers if it is smaller than $n_c$ by randomly selecting carriers one at a time. If the number of selected carriers is sufficient to fulfill the carrier minimum for all lanes but LP2($I_1$) is still infeasible, we will deselect a selected carrier from the alliance with the greatest representation, and select a carrier from an unrepresented alliance. This procedure repeats until the LP2($I_1$) model is feasible.

4.3. Application output

It is difficult in practice to determine an appropriate value for the balancing factor $\alpha$, which is an input parameter in our model. The difficulty mainly stems from the fact that the amount of change to the total transportation cost that results from a given change in $\alpha$ is unknown. Therefore, in order to present the information in a format that is most useful for the decision makers, the output of the application developed based on this research is in the form of a graph that shows the total transportation costs for a range of $\alpha$ values.

Recall that we can obtain the lower and upper bounds for $z(\alpha)$ by solving the MIP model with $\alpha = \alpha^-$ and $\alpha = \alpha^+$, respectively. Since MIP($M$) is an instance of MIP, we can solve it using RMTS to obtain a value for $z(M)$. We can also make a small modification to the RMTS algorithm to solve $\alpha$-MIP by changing the objective function to minimizing $z$, which provides a value for $\alpha$. We use this value of $\alpha$ to solve the MIP($\alpha$) model to get $z(\alpha^*)$, which is the first sample point in our graph.

The user provides a value $s$ representing the step size for the sample points. We generate the rest of the graph by solving MIP($\alpha + ks$) for successive values of the integer $k$ starting from $k = 1$, until $z(\alpha + ks) < z(M)$. We use the carrier selection in the best solution for MIP($\alpha + ks$) as the starting selection when solving MIP($\alpha + (k + 1)s$), which guarantees that $z(\alpha + (k + 1)s) \leq z(\alpha + ks)$ for all $k$. These solution values are plotted on the x-y-plane to form an allocation plan tradeoff graph, which provides a convenient pictorial representation for the decision makers to interpret the data. An example of such a graph is given in Fig. 1.

5. Computational experiments

To test the performance of our RMTS algorithm, we were provided with one real-world instance by the shipper. In addition, we generated a large number of random instances based on typical values encountered in practice. As a comparison, we also applied the branch-and-cut search scheme implemented by ILOG CPLEX 11.0 under the default options to solve both the MIP and $\alpha$-MIP models.

For the RMTS algorithm, the various parameters were given the following values after some preliminary experiments: the tabu tenure for selected carriers $\zeta_1 = 4$, the tabu tenure for unselected carriers $\zeta_0 = 10$, the maximum number of non-improving iterations $\beta_1 = m$ (i.e., the number of carriers), the maximum number of non-improving restarts $\beta_2 = 3$ and the probability parameter for the random move heuristic $\tau = 0.3$. The component linear
programming models in the RMTS approach were solved using ILOG CPLEX 11.0.

All algorithms were coded in JAVA and executed on an Intel Xeon (R) 2.66 GHz server with 3 GB of memory. Computation times reported are in CPU seconds on this server.

5.1. Experiment on real-world instance

We developed and implemented a decision support system (DSS) based on our approach, which we used to conduct the freight allocation for year 2008 for the shipper; this is a real-world instance that contains 26 candidate carriers and 378 lanes. The $x^*$ value found by the RMTS algorithm for this instance is 14.7. Given the step size $s = 0.5$, the MIP($x^* + k \times 0.5$) models were solved successively until $z = 21.7$, whose objective value found by the RMTS algorithm was 296084401. Fig. 1 shows the output produced by our DSS (marked with ‘+’) at the final round of contract negotiations; for comparison, we also provide the optimal solutions found by CPLEX (marked with ‘x’). Our RMTS algorithm produced this output after about 1 h of processing, taking between 144.6 and 381.5 s per solution. In contrast, CPLEX required about 6 h in total.

Our generation procedure is given as follows:

1. Lanes are classified into three categories depending on their demands into high-volume lanes ($[1000,2000]$), medium-volume lanes ($[100,1000]$) and low-volume lanes ($[10,100]$), which account for 10%, 60% and 30% of the total number of lanes, respectively. We uniformly randomly generate the demand $d_j$ for each lane $j$ within the given range in these proportions.

2. Set the maximum carrier percentage allocation $q_j^c$ and maximum alliance percentage allocation $q_j^a$ values for each lane $j$ based on their demand $d_j$ according to Table 3.

3. Generate the MQC values $b_i$ for all carriers $i \in I$ from $U[0.1D,0.16D]$, where $D = \sum_d d_j$ and $U[a,b]$ is the uniform distribution in the interval $[a,b]$.

4. Generate the freight price $p_j$ for each carrier-lane pair as follows. First, select a mean price $p_{ij}$ for lane $j$ from $U[20,100]$. Next, set the price $p_{ij}$ for all carriers $i$ to a random value from $U[0.8p_i, 1.2p_i]$. Finally, adjust the price to reflect the fact that a high MQC usually results in a lower freight price: compute a percentage $r$ from $U[0, (b_i - 0.13D)/0.26D]$ when $b_i > 0.13D$ or $U[(b_i - 0.13D)/0.26D, 0]$ when $b_i < 0.13D$, and then modify the price by setting $p_{ij} = (1 - r)p_{ij}$.

5. Assign each carrier to an alliance, where the number of carriers per alliance is an integer from $U[1,6]$.

<table>
<thead>
<tr>
<th>$d_j$</th>
<th>$q_j^c$</th>
<th>$q_j^a$</th>
<th>$q_j^{\gamma}$</th>
<th>$q_j^{\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[1000,2000]$</td>
<td>20%</td>
<td>5</td>
<td>40%</td>
<td>3</td>
</tr>
<tr>
<td>$[500,1000]$</td>
<td>30%</td>
<td>4</td>
<td>50%</td>
<td>2</td>
</tr>
<tr>
<td>$[200,500]$</td>
<td>40%</td>
<td>3</td>
<td>60%</td>
<td>2</td>
</tr>
<tr>
<td>$[0,200]$</td>
<td>60%</td>
<td>2</td>
<td>80%</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 1. Allocation plan tradeoff graph.
We compared the RMTS algorithm with the branch-and-cut search scheme implemented in CPLEX on these generated instances. The results for \( x \)-MIP (finding the minimum \( x \) with feasible solutions) are given in Table 4. The columns “RMTS time” and “RMTS obj” show the computation times and the \( x \) values found by RMTS, respectively. We first ran RMTS on these instances, and then set the time limit for CPLEX to be up to ten times the running time of RMTS. The columns “Gap (%)” \((x \in \{1,2,4,6,8,10\})\) give the gap between the best solutions obtained by CPLEX at \( x \) times of “RMTS Time” and the solutions obtained by RMTS, defined as follows:

\[
\text{Gap} = \frac{\text{CPLEX} - \text{RMTS}}{\text{RMTS}} \times 100\%.
\]  

In the process of negotiating for transportation services with carriers, numerous scenarios are analyzed by the shipper that might require the solutions of different freight allocation problems, so the speed of the solution approach is an important factor. The values in column “Gap1 (%)” show that when CPLEX is given the same amount of time required by RMTS to produce its solution, the \( x \) value found by CPLEX is larger than that found by RMTS for all instances. However, CPLEX was able to find better or equal solutions for 3 small instances with 50 lanes when given 4 times the computation time required by RMTS; when given 10 times the computation time, CPLEX found equal or superior solutions for 14 out of the 20 instances with 100 lanes or fewer. On the other hand, the solutions produced by RMTS were far superior to those...
found by CPLEX for all instances with 300 or more lanes even when CPLEX is given 10 times the computation time used by RMTS, with a gap of up to 40.1%; this suggests that RMTS is superior to CPLEX when solving a problem instance within a reasonable amount of computation time. Although the RMTS algorithm found feasible solutions for all instances using significantly less computation time.

We also conducted some additional experiments to gauge how the value of $x$ influences the performance of the two algorithms on the MIP model. We found that the as the value of $x$ increases, the performance of CPLEX increases more rapidly than that of the RMTS algorithm; for large values of $x$ (where the lane cost balancing constraint has a comparatively smaller effect on the final solution), CPLEX outperforms RMTS even for large instances. We can therefore conclude that when solving the MIP model with tight lane cost balancing constraints, i.e., when $x$ is close to $x^*$, the RMTS algorithm outperforms CPLEX for larger instances.

Finally, we conducted some additional experiments to discover the effect of different minimum carrier percentage allocation $q^*_C$ and minimum alliance percentage allocation $q^*_A$ values on the total cost. These experiments were performed on the ten instances with 100 lanes. For each instance, we increase both the $q^*_C$ and $q^*_A$ values by $0\%$, $5\%$, $7\%$, $10\%$, $12\%$, $15\%$, $17\%$, and $20\%$, and solve the corresponding MIP$(x + 3)$ model optimally using CPLEX. We also solve each instance with both carrier and alliance quantity limit constraints removed, which has the lowest cost.

Table 7 gives the gap between the optimal solutions for each of the above cases and the corresponding instance with the quantity limit constraints removed, which is a percentage value calculated in a manner similar to Eq. (17). We see that for these instances, the full quantity limit constraints (given by column “0%”) increase the total transportation cost by 8.76% on average, and less stringent constraints incur lower costs.

### 6. Conclusions

In this work, we addressed a new freight allocation problem which incorporates a lane cost balancing constraint. This constraint requires that the freight cost of all lanes are within a certain percentage of the theoretical lower bound, thereby avoiding the scenario where buyers pay unreasonably high transportation costs on some lanes. We show that with the lane cost balancing constraint, even finding a feasible solution is NP-hard in the strong sense. We developed a customized tabu search algorithm with a random move heuristic to solve this problem in practice. Experiments show that it is superior to the branch-and-cut approach implemented in ILOG CPLEX 11.0 for instances of practical size.
although finding more efficient exact algorithms to solve this problem could be a direction for future research.

This algorithm was implemented as the engine of a decision support system. It is currently employed by the shipper for its freight allocation decisions, and is generally regarded as an effective and useful tool in the decision-making process. In 2008, the decision makers chose to use a freight allocation plan that balances lane costs to a large extent even though it was not the best plan in terms of total transportation cost; they believed that the slight increase in transportation cost is justified by the retention of goodwill as a result of the fairer allocation. This is an example of a practical transportation research problem where the most cost-effective solution is not necessarily the best when the less tangible human considerations of fairness and goodwill are taken into account.

Acknowledgments

We would like to thank the co-ordinating editor Prof. Roman Slowinski and the anonymous reviewers for their insightful comments and suggestions, which have significantly improved our paper. We also thank the project leader Chan Hou Che who facilitated the communication and development process with the client company. The third author was partially supported by the Hong Kong Research Grant Council (RGC) Direct Allocation from the Hong Kong Polytechnic University, RGC No. PolyU 5328/08E, Project Account Code A-SA24.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at: doi:10.1016/j.ejor.2011.08.028.

References