



Integrated hub location and flow processing schedule problem under renewable capacity constraint

Mohammad Rostami¹ · Armin Jabbarzadeh^{2,3}

Received: 24 December 2020 / Revised: 14 May 2021 / Accepted: 29 May 2021 / Published online: 12 June 2021
© SBMAC - Sociedade Brasileira de Matemática Aplicada e Computacional 2021

Abstract

Transport and logistics networks are more complex than ever before. Complex networks and limitation of available capacities make flow processing schedule (FPS) a challenging problem to ensure customer satisfaction and profitability. This paper addresses an integrated hub location and flow processing schedule problem (HLFPSP) to determine hub locations, the allocated traffic flows and the optimal FPS at hubs under capacity constraints. Specifically, in an air transport network, the optimal FPS leads to optimal flight scheduling. The problem is formulated as a mixed-integer linear programming (MILP) model to minimize total tardiness costs (at operational level) and hub construction costs (at strategic level), simultaneously. The developed model is utilized to solve small-size hub location-scheduling problems. To provide a good solution in a reasonable time, the Lagrangian relaxation algorithm is employed. Based on the data from 80 airports in Turkey, the application of problem in real world is showed and the efficiency of the proposed solution methods is evaluated. Finally, a number of sensitivity analyses and managerial insights are provided to enrich the computational results.

Keywords Hub location · Flow processing schedule · Total tardiness · Mixed-integer linear programming · Lagrangian relaxation algorithm · Heuristic algorithm

Mathematics Subject Classification 90B06 · 90B35

1 Introduction

Hubs are special facilities that act as switching, transshipment and sorting points in many-to-many distribution systems [see O'Kelly (1987) and Ernst and Krishnamoorthy (1996)]. Instead

Communicated by Hector Cancela.

✉ Mohammad Rostami
rostami_m@shahroodut.ac.ir

¹ Department of Industrial Engineering and Management, Shahrood University of Technology, Shahrood, Iran

² Department of Industrial Engineering, Iran University of Science and Technology (IUST), Tehran, Iran

³ Department of Systems Engineering, École de Technologie Supérieure (ÉTS), University of Quebec, Montreal, Canada

of serving each origin–destination pair independently, hub facilities concentrate flows to take advantage of economies of scale. Flows from the same origin with different destinations are consolidated on their route to the hubs and then split and re-consolidated with flows of other origins for similar destinations [see Alumur and Kara (2008)]. Hub location problems (HLP) examine potential location of facilities through which flows of passengers or freights are to route from origins to destinations. Literature of hub location problems aims to find suitable location for the hubs to enhance the network performance. Reviewing the HLP literature shows two main trends over the past couple of decades. First is the increasing use of hub-and-spoke network structure in research and practice, which can be witnessed by geometric growth of published papers in the field. And second, is more complex logistics network requiring sophisticated algorithms to solve the models, which has promoted the increasing application of heuristic solution methods [see Farahani et al. (2013), Campbell and O’Kelly (2012) and Alumur and Kara (2008)].

In recent years, diversification of services has been considered by transportation companies. These diverse and sometimes exclusive services are their competitive advantages. One of these services is to provide express service for special customers to deliver passengers or freights to their destinations. To do this, tight due dates are set for some flows, and the company will be subject to penalties commensurate with the amount of tardiness if the cargo is delivered after due date. High-speed delivery with no latency makes a competitive environment for the firms (Karimi and Setak 2018). Some of them, like FedEx, apply a money back guarantee if quoted delivery time is missed even by as little as 60 s. Another example is related to aviation compensation laws in the European Union. In the European Union, Flight Compensation Regulation 261/2004 states that flight delays for over 3 h entitle passengers to a compensation from €250 up to €600 per passenger from the airline (2004). Therefore, it is very important to pay attention to delivery times in a transportation network.

The distribution network topology significantly impacts distribution time and cost (Karimi and Setak 2018). Hub location problem is a well-known network design problem which effectively handles the time-definite service (Campbell 2009). In a hub network, the delivery time is a function of two decisions. The first is the routing of the flow under the hub network, which directly affects the distance and travel time. The second is the flow processing schedule at each hub, which can affect the waiting time of the flow for in-hub services, especially in congested hubs. According to Levin (2007) study, one of the main reasons for flight delays is congestion in air traffic. Transportation companies use hub-and-spoke network structure to reduce transportation costs through consolidation of flows. Such a decision may cause flow congestion in some hubs, and therefore there is a need to properly schedule the processing of flow at each hub to minimize delivery times. On the other hand, introducing flow processing schedule into hub location problems is necessary to efficiently design hub networks, as integrated decisions may have an effect on the optimal hub locations.

Based on the definitions, HLP is a strategic decision and the problem is mostly solved independently from requirements of tactical and operational planning. However, it should be noted that decisions at the strategic level will directly affect decisions at the operational planning level, and therefore their separate planning may lead to decisions that are far from optimal [see Badri et al. (2013)]. Suppose that the optimal location of the hubs and the flow allocation be determined without considering FPS. So, such a decision may negatively affect the total tardiness. The total tardiness may be reduced by changing the location of the hubs or by allocating flows or both. In a real case, we know that some FPS features such as due dates and release times may change on a much shorter time scale, but the amount of flow (the number of jobs) that must be scheduled in each hub (machine environment) as well as its capacity (number of parallel machines) are affected by strategic planning, i.e. hub location and

flow allocation to them. Therefore, creating a trade-off between these two levels of planning seems necessary. It should be noted that in this study, these two levels of decisions are made in the same time scale. While, in classic HLP, it is not possible to calculate and minimize the tardiness times of passengers or freights, simultaneously. This paper addresses this gap and considers both hub location and flow processing schedule problems. This issue has not been addressed in the literature to date. However, there are research in the literature on considering operational decisions in HLP that indicate the importance of considering integrated decisions in this scope [see Zhang et al. (2017), Karimi and Setak (2018), Masaeli et al. (2018) and Charisis et al. (2020)].

Operational characteristics related to in-hub services stimulate the primary motivation of this study. Besides that, changing the approach to resource planning in the hub network is another motivation of this research. Operations at the hubs are limited by their capacity, which also makes the flow processing plan more challenging. To formulate the capacitated hub location problem, the common approach is to consider the hub capacity over a given time period, which is calculated based on net capacity times the length of time period [e.g. da Graça Costa et al. (2008), Correia et al. (2013) and Kumar and Sivakumar (2013)]. In this approach the capacity is nonrenewable, meaning that the available capacity of a hub gets updated when a flow is allocated to the hub but it is not re-calculated when the transiting flow leaves the hub. In the literature, a large number of research have used this approach because their planning is at a strategic level. The other approach, renewable capacity, is to investigate the hub capacity per unit of time. So, the capacity is utilized as long as the flow resides at the hub and becomes available when the flow leaves. What is happening in the real world is closer to the second approach. For example, if we consider the number of runways as a limited capacity resource in the hub, this resource will only be used when the aircraft occupies the runway. As soon as the aircraft leaves it, the available capacity of the resource increases. Although it is necessary to consider the latter approach for flow processing schedule at each hub, it is more complicated and less employed in the literature. In fact, another contribution of this paper is to incorporate the capacity constraint based on renewable capacity approach in the model.

The formal path of this research field goes back to the paper published by Toh and Higgins (1985) on the application of hubs in airport business. Motivated by Toh and Higgins's idea on hub location, O'Kelly (1986a, 1986b, 1987) published the first mathematical modeling and solution methods for HLP. After O'Kelly, several modeling efforts and solution methods have been presented in the literature. To learn more about the various models of the hub location and classification review of that kind of problems, you can see Alumur et al. (2020) research.

Planning the airport operations is one the main applications of FPS problem. It seems that the literature on this subject is not very rich. Brueckner and Zhang (2001) provided a comprehensive economic analysis of scheduling decisions in airline networks. McWilliams (2005) investigated the scheduling problem in freight consolidation terminals in parcel delivery industry. The unloading inbound trucks were scheduled for a fixed number of cross-docks with the goal of minimizing time span for transfer operation. The truck scheduling was done through a simulation-based genetic algorithm to search for new solutions. McWilliams et al. (2005, 2008, 2010) proposed different solution methods including genetic algorithm and simulated annealing to schedule the trailers to the unload docks in the parcel delivery industry. Lu et al. (2015) designed a timetable for managing airport which is a new access service provided by the airport with the purpose of attracting more passengers. Zhang et al. (2017) considered an integrated plane assignment and hub location problem for the air-cargo delivery service. They present two different MIP model to integrate hub location, flow allocation

and plane scheduling and also developed a two-stage hybrid algorithm for solving large-size instances.

Recently, some researchers have investigated flow shipment scheduling in a hub location problem. This problem is introduced by Masaeli et al. (2018) for the first time. Based on their definition, shipment scheduling seeks to determine the number of vehicles dispatched from the hubs at different times and does not pay attention to the flow processing schedules at each hub. It was supposed that the capacity of the hubs is unconstrained while the capacity of vehicles is limited. It is also assumed that shipment schedules are to be defined only within the inter-hub links. Karimi and Setak (2018) developed a bi-objective model for flow shipment scheduling problem in an incomplete hub location-routing network. They present mathematical programming models in deterministic and stochastic environments when flow arrivals follow a piecewise linear form. Also, Charisis et al. (2020) introduced the problem of locating leasing hubs and their optimal leasing schedule. They proposed a linear programming model to achieve the optimal solution.

As can be seen, the problems of hub location and flow processing schedule under renewable capacity constraint are not formulated and solved simultaneously, while the importance of tactical and operational requirements cannot be denied. This paper investigates the hub capacity from renewable capacity approach that is directly influenced by operational scheduling (e.g. incoming and outgoing time) of the flows. If a general pattern for the flows between different origins and destinations can be estimated, that estimation can be used for better decision making regarding the hub locations. The main purpose of this paper is to develop an integrated hub location and flow processing schedule (HLFPS) problem to minimize the total construction and operational cost. This model uses the information of flow patterns to firstly solve the hub location problem and then generate the preliminary schedule for the processing of flows.

The developed model investigates the choice of hub airports, where the flights scheduling depends on airport runway constraints for aircraft servicing (such as refueling, maintenance, etc.). Considering the limitations on the number of runways, a mixed-integer linear programming model is presented for flow processing schedule to minimize total tardiness and total cost of hub construction. The problem is large scale and an optimal solution cannot be achieved in a reasonable time. Considering the high number of nodes, a Lagrangian relaxation (LR) method is employed to solve the problem for a good lower bound. In addition, performance of the proposed solution method is compared with an optimal solution and an upper bound heuristic algorithm for small-size problems.

The remaining parts of this paper are organized as follows. The problem is introduced and a mixed-integer linear programming model is developed in Sect. 2. To solve the problem in large-size settings, Sect. 3 proposes a solution algorithm based on the Lagrangian Relaxation where a heuristic method is used to obtain the upper bound at each iteration. Computational results are elaborated in Sect. 4 and paper concludes with discussion on the results and suggestions for future research.

2 Problem formulation

To formulate the HLFPS problem, we consider a network with N nodes. Some of the nodes can be selected to act as hub and to provide the following two major functions: (1) switching, sorting, or connecting (SSC) function, and (2) consolidation/break-bulk (CB) function. The SSC function facilitates redirections of the flows and allows many origins and destinations to

be connected with fewer links than a fully connected network. The CB function allows flows to be aggregated and disaggregated to decrease the total cost through economies of scale. Each flow is recognizable based on its origin–destination. They have a predetermined release time and the due date for arrival of the flow at destination is known. The flow processing preemption is not allowed and hubs have limited capacity for simultaneous processing of flows. If a flow is assigned to a hub with fully occupied capacity, it has to wait in a queue that may lead to delayed arrival at destination. A fully connected network is very expensive to run, because not all the direct routes can exploit the economies of scale and cover the associated costs. In addition, the required functions for preparation of the flows (e.g. sorting, consolidating, and break-bulking) require that the flows pass through at least one and at most two hubs.

This problem is a single allocation hub location problem where each demand point is attached only to one hub. The planning horizon is T , over which the flows' scheduling is done. The objective function is to minimize the total tardiness costs plus the homologized costs of hub construction. Homologized cost of hub construction is obtained by dividing the cost to establish a hub by the length of the planning horizon. Please remember that the transportation cost is included as indirect cost in the total tardiness cost function. Table 1 provides the definitions of the parameters and variables.

Based on the defined parameters and decision variables, the HLFPS Problem is formulated as follows:

$$\text{Min } f = \sum_k f_k z_k + \beta \sum_{ij} T_{ij}. \tag{1}$$

Subject to:

$$\sum_k \sum_l x_{ijkl} = 1 \forall ij (i \neq j), \tag{2}$$

$$\sum_l x_{ijkl} \leq z_k \forall k, ij (i \neq j), \tag{3}$$

$$\sum_k x_{ijkl} \leq z_l \forall l, ij (i \neq j), \tag{4}$$

$$y^t_{1ijk} \leq \sum_l x_{ijkl} \forall k, t, ij (i \neq j), \tag{5}$$

$$y^t_{2ijl} \leq \sum_k x_{ijkl} \forall l, t, ij (i \neq j), \tag{6}$$

$$\sum_k \sum_t y^t_{1ijk} = 1 \forall ij (i \neq j), \tag{7}$$

$$\sum_l \sum_t y^t_{2ijl} + \sum_l x_{ijll} = 1 \forall ij (i \neq j), \tag{8}$$

$$\sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y''_{1ijk} \right) + \sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y''_{2ijk} \right) \leq C_k z_k \forall k, t, \tag{9}$$

$$\sum_t t y^t_{1ijk} \geq \sum_l (x_{ijkl} (r_{ij} + t_{ik} D_{ik})) \forall k, ij (i \neq j), \tag{10}$$

$$\sum_t t y^t_{2ijl} + M(1 - x_{ijkl}) \geq \sum_t t y^t_{1ijk} + v_k p_{ij} + \alpha t_{kl} D_{kl} \forall ij (i \neq j), k, l (k \neq l) \tag{11}$$

$$T_{ij} \geq \sum_k \sum_t t y^t_{1ijk} + \sum_k \sum_l (x_{ijkl} (v_l p_{ij} + t_{lj} D_{lj})) - d_{ij} \forall ij (i \neq j), \tag{12}$$

Table 1 Parameters and decision variables

	Description	Values taken
<i>Indices</i>		
I	Index of origin nodes	$1, 2, \dots, N$
J	Index of destination nodes	$1, 2, \dots, N$
k, l	Index of hubs	$1, 2, \dots, N$
T	Index of time	$1, 2, \dots, T$
<i>Parameters</i>		
r_{ij}	Release time of the flow from origin i to destination j	
p_{ij}	Processing time of the flow from origin i to destination j at each hub	
R_{ij}	Size of the flow from origin i to destination j	
M	A positive big number	
D_{ij}	The distance between node i and j	
t_{ij}	The transportation time for an unit of distance from node i to node j	
d_{ij}	Due date of the flow from origin i to destination j	
f_k	Fixed homologized cost to establish a hub at node k	
C_k	Capacity of hub at node k	
v_k	Velocity of flow processing at hub k	
β	Time to cost coefficient	
α	Discount coefficient for transportation time between two hubs	
<i>Decision variables</i>		
T_{ij}	Tardiness of the flow from origin i to destination j	Nonnegative
z_k	Equal to 1 if a hub is located at node k	0, 1
x_{ijkl}	Equal to 1 if flow from origin i to destination j goes through the hubs located at nodes k and l	0, 1
y^t_{1ijk}	Equal to 1 if the process of flow from node i to node j starts at time t at hub k (as the first hub)	0, 1
y^t_{2ijl}	Equal to 1 if the process of flow from node i to node j starts at time t at hub l (as the second hub)	0, 1

$$T_{ij} \geq \sum_l \sum_t t y^t_{2ijl} + \sum_k \sum_l (x_{ijkl} (v_l p_{ij} + t_{ij} D_{lj})) - d_{ij} \forall i, j (i \neq j), \quad (13)$$

$$\begin{cases} z_k, x_{ijkl}, y^t_{1ijk}, y^t_{2ijl} \in \{0, 1\} \forall i, j, k, l, t \\ T_{ij} \geq 0 \forall i, j \end{cases} \quad (14)$$

The objective function minimizes hub construction cost and total tardiness costs. The β coefficient is tardiness penalty costs per unit time which include special services to passengers, delay penalties fees are paid to them and other imposed transportation costs. It should be noted that these costs can be different for each route; however, to simplify the model, here, it is assumed that for all routes, these costs are fixed. Constraint (2) ensures the flow passes through minimum one and maximum two hubs from origin to destination. Constraints (3) and (4) state that if no hub is formed at a node, then that node does not perform the hub functions. Constraints (5) and (6) necessitate starting the flow processing only at the hubs. Constraint (7) guarantees that each flow is allocated to a hub (called “the first hub”), and

the hub function is done only once. Constraint (8) is to assure one-hub or two-hub route is selected. So, if the flow is planned to be routed on a one-hub route, the binary variable y_{2ijl}^t becomes 0 for all other hubs (i.e. there is not a second hub in the chosen route). Constraint (9) controls the capacity limitation and prevents the flows from being allocated to the hubs more than the available capacity at any given time. The capacity used for one-hub routes is calculated based on the first component in Eq. (9) and the second component controls the used capacity for two-hub routes. Constraints (10) and (11) indicate that the start time of processing for each flow at the first (or second) hub should not be earlier than the arrival of that flow to the hub. In constraint (11), the beginning of processing at the second hub is not calculated for one-hub routes, because according to relation (8) this is always equal to 0. Constraint (12) and (13) are to calculate the tardiness for each flow at its destination. If the flow ij is travelled through a one-hub route, then the start time for process at the second hub is zero and its tardiness will be calculated based on constraint (12). If the arrival time for a flow is earlier than its due date, then its tardiness will be 0 (i.e. the early arrival of the flow at its destination would not incur any penalty for the system). Finally, constraint (14) defines the decision variables of the problem.

3 Lagrangian relaxation algorithm

3.1 Lagrangian relaxation model

The Lagrangian relaxation (LR) is one of the well-known methods for calculating lower bounds in combinational minimization problems. By eliminating complicated constraints of the problem and adding them with Lagrange multiplier to the objective function as penalties, this method attempts to facilitate the solution process. The Lagrangian relaxation problem is solved recursively and the multipliers are updated at each iteration. The sub-gradient optimization technique is used to update multipliers mostly. For a review on the LR methods and techniques, see Guignard (2003).

Constraints (3) and (4) can be eliminated from the developed model, and can be added to the objective function using the multipliers of $\mu_{l,ij}$ and $\lambda_{k,ij}$. Also, the solution process can be facilitated through relaxing constraint (9) and adding it to the objective function with the multiplier γ_{kt} . This procedure gives the updated Lagrangian goal function as follows:

$$\begin{aligned}
 L(\lambda, \mu, \gamma) = & \text{Min} \sum_k f_k z_k + \beta \sum_{ij} T_{ij} \\
 & + \sum_{ij} \sum_k \lambda_{k,ij} \left(\sum_l x_{ijkl} - z_k \right) + \sum_{ij} \sum_l \mu_{l,ij} \left(\sum_k x_{ijkl} - z_l \right) \\
 & + \sum_k \sum_t \gamma_{kt} \left(\sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y_{1ijk}^u \right) + \sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y_{2ijk}^u \right) - C_k z_k \right).
 \end{aligned} \tag{15}$$

S.t:

Constraints (2), (5)–(8), (10)–(14).

To simplify the above model, we define t_{ijk}^1, t_{ijl}^2 as two new variables, where $t_{ijk}^1 = \sum_t t y_{1ijk}^t$ indicates the start time of the processing of the flow ij at hub k as the first hub; and

$t_{ijl}^2 = \sum_t t y_{2ijl}^t$ shows the start time of the processing of the flow ij at hub l as the second hub. Accordingly, the model $L(\lambda, \mu, \gamma)$ can be rewritten as below:

$$\begin{aligned}
 L(\lambda, \mu, \gamma) = & \text{Min } \sum_k f_k z_k \\
 & + \beta \sum_{ij} \max \left(\sum_l t_{ijl}^2 + \sum_k \sum_l (x_{ijkl} (v_l p_{ij} + t_{lj} D_{lj})) - d_{ij}, \sum_k t_{ijk}^1 \right. \\
 & \left. + \sum_k \sum_l (x_{ijkl} (v_l p_{ij} + t_{lj} D_{lj})) - d_{ij}, 0 \right) \\
 & + \sum_{ij} \sum_k \lambda_{k,ij} \left(\sum_l x_{ijkl} - z_k \right) + \sum_{ij} \sum_l \mu_{l,ij} \left(\sum_k x_{ijkl} - z_l \right) \\
 & + \sum_{ij} \sum_k \sum_{l \neq k} x_{ijkl} \left(R_{ij} \left(\sum_{t=t_{ijk}^1}^{t_{ijk}^1 + v_k p_{ij} - 1} \gamma_{kt} + \sum_{t=t_{ijl}^2}^{t_{ijl}^2 + v_l p_{ij} - 1} \gamma_{lt} \right) \right) \\
 & + \sum_{ij} \sum_k x_{ijkk} \left(R_{ij} \sum_{t=t_{ijk}^1}^{t_{ijk}^1 + v_k p_{ij} - 1} \gamma_{kt} \right) - \sum_k \sum_t \gamma_{kt} C_k z_k
 \end{aligned} \tag{16}$$

S.t:

$$\begin{cases}
 (2), (5), (6), (7), (8), (10), (11) \\
 t_{ijl}^2, t_{ijk}^1 \in \text{Integer} \\
 z_k, x_{ijkl}, y_{1ijk}^t, y_{2ijl}^t \in \{0, 1\}
 \end{cases}$$

It should be taken into account that the problem $L_z(\lambda, \mu, \gamma)$ can be divided into two separate sub-problems: (1) a problem in the space of the variable z , and (2) a problem in the space of variables x and t .

The sub-problem related to the variables z is modeled as follows:

$$L_z(\lambda, \mu, \gamma) = \text{Min } \sum_k z_k \left(f_k - \sum_{ij} \lambda_{k,ij} - \sum_{ij} \mu_{k,ij} - C_k \sum_t \gamma_{kt} \right) \tag{17}$$

S.t :

$$z_k \in \{0, 1\}$$

The solution of $L_z(\lambda, \mu, \gamma)$ is simply found if only the variables z_k are equal to 1 where $f_k - \sum_{ij} \lambda_{k,ij} - \sum_{ij} \mu_{k,ij} - C_k \sum_t \gamma_{kt} \leq 0$. The remaining z_k are equal to 0.

For the sub-problem associated with x and t , it is possible to convert it for each flow ij into new sub-problems. Hence, modeling of the sub-problem related to the variables of x and t for each ij will be as follows:

$$L_{x,t}(ij, \lambda, \mu, \gamma) = \text{Min } \beta \sum_k \sum_l x_{ijkl} \left(\lambda_{k,ij} + \mu_{l,ij} + \underbrace{\left(\max(t_{ijk}^1 + v_k p_{ij} + t_{kj} D_{kj} - d_{ij}, t_{ijl}^2 + v_l p_{ij} + t_{lj} D_{lj} - d_{ij}, 0) + R_{ij} \left(\sum_{t=t_{ijk}^1}^{t_{ijk}^1 + v_k p_{ij} - 1} \gamma_{kt} + \sum_{t=t_{ijl}^2}^{t_{ijl}^2 + v_l p_{ij} - 1} (\gamma_{lt} | l \neq k) \right) \right)}_{\varphi_{ijkl}} \right), \tag{18}$$

S.t:

$$\sum_k \sum_l x_{ijkl} = 1, \tag{19}$$

$$t_{ijk}^1 \geq \sum_l (x_{ijkl} (r_{ij} + t_{ik} D_{ik})) \forall k, \tag{20}$$

$$t_{ijl}^2 + M(1 - x_{ijkl}) \geq \sum_k (x_{ijkl} (t_{ijk}^1 + v_k p_{ij} + \alpha t_{kl} D_{kl})) \forall k, l \tag{21}$$

$$x_{ijkl} \in \{0, 1\} \forall k, l.$$

The problem above is primarily examined for two situations. First, for all k and l ($k \neq l$) and for all possible states where $t_{ijl}^2 \geq t_{ijk}^1 + v_k p_{ij} + \alpha t_{kl} D_{kl}$, the best solution for t_{ijl}^2 is chosen to minimize φ_{ijkl} . Then, the best combination of (t_{ijk}^1, t_{ijl}^2) with the minimum φ_{ijkl} is chosen as the optimized combination for t_{ijk}^1 and t_{ijl}^2 . In the second situation, per all $k = 1$, the amount of t_{ijk}^1 ($t_{ijk}^1 \geq r_{ij} + t_{ik} D_{ik}$) is chosen for which φ_{ijkl} is minimum.

Through determining the best values of φ_{ijkl} for each k and l , for each flow from origin i to destination j , we can solve the second sub-problem (Eqs. (18)–(21)) using the simple rule below. The structure of lagrangian relaxation method can be summarized in the following results:

Proposition 1 *It is possible to say:*

$$L(\lambda, \mu, \gamma) = L_z(\lambda, \mu, \gamma) + \sum_{ij} L_{x,t}(ij, \lambda, \mu, \gamma). \tag{22}$$

3.2 Solution procedure

To obtain the best possible lower bound for the problem, the problem $L(\lambda, \mu, \gamma)$ must be maximized for all different values of λ, μ and γ .

$$Z_D = \max L(\lambda, \mu, \gamma) \tag{23}$$

$$\lambda, \mu, \gamma \geq 0.$$

To do this, we implement the iterative sub-gradient method. In this algorithm, LB^m indicates the lower bound obtained at iteration m . UB is the best upper bound obtained for this main problem. Also, the coefficient α^m will be halved if no improvement is obtained for the lower bound solution after 10 iterations. In this algorithm, $g(\lambda^m, \mu^m, \gamma^m)$ is the sub-gradient

of $L(\lambda^m, \mu^m, \gamma^m)$ at iteration m , equal to the following:

$$g(\lambda^m, \mu^m, \gamma^m) = \begin{pmatrix} \left(\sum_i x_{ijkl}(m) - z_k(m) \right), \\ \left(\sum_k x_{ijkl}(m) - z_l(m) \right), \\ \left(\sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y_{ijk}^u(m) \right) + \sum_{ij} R_{ij} \left(\sum_{u=t-v_k p_{ij}+1}^t y_{ijk}^u(m) \right) - C_k z_k(m) \right) \end{pmatrix}. \tag{24}$$

The outputs of this algorithm are z_D and sUB which indicate the lower and upper bounds of the master problem, respectively. The pseudo-code related to this algorithm is as follows:

Step 1. (Initial setting)

- Set $m=0, Z_D = -\infty, \lambda^0 = \mu^0 = \gamma^0 = 0$ and $\alpha^0 = 2$.

Step 2. (Calculation)

Repeat

- Solve $LB^m = L(\lambda^m, \mu^m, \gamma^m)$,
- **If** $LB^m > Z_D$
Then $Z_D \leftarrow LB^m$,
- Calculate UB for this iteration,
- Calculate step size $s^m = \frac{\alpha^m (UB - LB^m)}{\|g(\lambda^m, \mu^m, \gamma^m)\|^2}$
- Update Lagrange multiplier vector;
 $(\lambda^{m+1}, \mu^{m+1}, \gamma^{m+1}) = \max(0, (\lambda^m, \mu^m, \gamma^m) + s^m g(\lambda^m, \mu^m, \gamma^m))$,
- $m=m+1$,

Until $\frac{UB - LB^m}{UB} < 0.005$ or $m=500$ or $Time_{max}=3600$.

Step 3. End.

3.3 The upper bound heuristic

In the Lagrangian relaxation method, the quality of upper bounds obtained from each iteration is of great importance to achieve the best lower bound. In this section, a heuristic algorithm is presented to find an upper bound for the master problem at each iteration. At each iteration, after solving the problem $L(\lambda^m, \mu^m, \gamma^m)$, the resulted solution $x_{ijkl}(m)$ for each flow ij is used to generate a possible upper bound for the master problem.

For those nodes which are used as hubs for at least one flow from i to $j, z_k(m) = 1$ holds. In this case, we need to be sure that both hub construction and routes will be feasible for all flows. So, it is crucial to evaluate the possibility of an established schedule for each hub provided that the constraints associated with the capacity are satisfied at any time. Inspired by the heuristic algorithm developed by Rostami et al. (2014) for resource-constrained project scheduling problems, a two-stage algorithm is used to solve the problem. In the first stage, the priority of scheduling at each hub k is determined by a specific rule. In the second stage,

according to such priorities, a simple algorithm is applied to schedule the processing on the flows allocated to each hub k .

At the first stage of the algorithm, three types of priorities are generated for the flows. The first priority is based on the flow arrival time to a given hub (it means: $r_{ij} + t_{ik} D_{ik}$ (or $r_{ij} + t_{ik'} D_{ik'} + v_{k'} p_{ij} + \alpha t_{k'k} D_{k'k}$)). The first-in-first-served (FIFS) rule implies that each flow reaching the hub earlier will be given higher processing priority. Hence, the idle probability of the hub is decreased; however, with regard to due dates, this rule may cause tardiness. It should be considered that, the arrival time of a flow to hub k as the second hub is estimated when the flow has passed through hub k' as the first hub. Since the real time of processing the flow ij at the hub k' is not clear, this is considered as equal to the arrival time of that flow to hub k' .

The second prioritization rule, earliest-due-date (EDD), implies that any flow with closer due date have higher priority for processing. Such prioritization can increase the waiting time for flows with farther due dates, although they might have reached the hub earlier. The third prioritization rule is based on arrival time, processing time, and due date. The highest priority will be given to a flow with the least sum of arrival time, processing time and due date.

At the second stage of the algorithm, a certain schedule is generated for each type of the priorities provided before. To create feasible schedules regarding the capacity of hub k , the start time of processing each flow ij is considered as the earliest possible time in the planning horizon, if the available capacity is enough for processing the flow ij over the time interval of $[t_{ijk}^1$ (or t_{ijk}^2), $t_{ijk}^1 + v_k p_{ij}$ (or $t_{ijk}^2 + v_k p_{ij}$)]. It should be noted that the earliest possible time to start processing the highest priority flow is equal to the arrival time of that flow to the hub. For the flows with lower priority, the probability of immediate processing at the time of their arrival to the hub is relatively lower, because the higher priority flows have already been scheduled for processing. Hence it is expected that the lowest priority flows spend more time waiting to be processed. When all the flows are scheduled, a scheduling plan with the minimum sum of tardiness will be chosen as the optimized schedule at hub k . Below is the formal definition for obtaining a feasible schedule at hub k :

Stage I: (priority setting)

- Flow sorting type 1 (fst=1): FIFS rule,
- Flow sorting type 2 (fst=2): EDD rule,
- Flow sorting type 3 (fst=3): combinatorial rule

Stage II: (Flow scheduling)

For fst=1 to 3 **do**

For each flow ij **do**

Step1. **If** processing flow ij at time t ($t \geq$ arrival time flow ij)
 does not exceed hub capacity in interval $[t, t + v_k p_{ij}]$

Then schedule flow ij and modify hub capacity

Step2. Otherwise increase t to the nearest capacity
 modification and go to Step1.

End do.

End do.

Having the cost parameters for hub construction, routes for different flows and schedule of the flows at the hubs, the original objective function can be calculated, that its solution is the upper bound heuristic at each iteration.

4 Computational results

This section evaluates and compares the performance of three solution methods; namely, optimization with CPLEX solver in GAMS, Lagrangian relaxation algorithm and upper bound heuristic algorithm. The developed MIP model was solved utilizing CPLEX solver in GAMS, on a system with Intel Core i7, 3.1 GHz CPU and 8 GB RAM. The Lagrangian Relaxation algorithm and the upper bound heuristic algorithms were coded with C# and ran on the same system.

The test problems were generated based on the data used by Alumur et al. (2009) for airports flows in Turkey. Some of the required data in our model that were not included in the former study by Alumur et al. (2009), have been generated randomly. The main intention was to generate a set of data close to the reality. The parameter p_{ij} was randomly generated based on a uniform distribution function on $[0.1, 1]$, r_{ij} from a uniform distribution function on $[1, 18]$, d_{ij} from a uniform distribution function on $[r_{ij} + 3, r_{ij} + 6]$, t_{ij} from a uniform distribution function on $[0.0025, 0.0035]$, c_k from a uniform distribution function on $[1, 8 + \lfloor n/20 \rfloor]$, v_k from a uniform distribution function on $[0.85, 1.15]$, f_k from a uniform distribution function on $[20, 60]$, and R_{ij} is considered to be 1 for all flows. Further, the coefficients of $\beta = 5$ and $\alpha = 0.3$ were included and the planning horizon considered to be $T = 24$, respectively.

Table 2 exhibits the results of small-size problems consisting of up to 20 nodes including all origins, destinations and potential hubs. The optimal solution of small-size problems can be computed with CPLEX in a logical time (less than 3600 s). It should be noted that the logical time can be different depending on the type of problem; however, in this problem, because of planning at the operational level (scheduling daily flights), we have to solve the problem less than 1 h (3600 s). In Table 2, the first section presents the values of objective functions obtained from the CPLEX, LP-relaxation (LPR) method, Lagrangian relaxation (LR) algorithm and the upper bound (UB) heuristic for the individual problems. The second section indicates the CPU running time per second for different methods, while the third section provides the gap between the lower bounds and the optimal solution. Table 3 presents the results of large size problems, containing up to 80 nodes. For these problems, the model

Table 2 Results for small-size problems

# of nodes	Objective function value				CPU running time (s)			Gap ^d	
	CPLEX	LPR ^a	LR ^b	UB ^c	CPLEX	LPR	LR	LPR	LR
5	60.52	60.14	60.52	60.52	2.150	0.832	0.183	0.0062	0
7	138.62	134.70	138.62	138.62	17.231	5.883	0.635	0.0282	0
10	66.64	63.94	66.64	66.64	84.713	29.038	2.172	0.0405	0
12	78.32	75.39	78.32	78.32	293.615	114.286	4.390	0.0374	0
15	71.20	67.85	71.13	71.31	844.126	362.962	9.732	0.0470	0.0025
17	69.85	64.23	69.85	69.85	1533.027	519.541	16.136	0.0805	0
20	120.25	108.25	120.18	120.50	3584.276	1142.557	28.315	0.0998	0.0026

^aLP-relaxation method

^bLagrangian relaxation algorithm

^cUpper bound heuristic

^dThe gap between lower bound and optimal solution with CPLEX

Table 3 Results for large-size problems

# of nodes	Objective function value			CPU running time (s)		Gap ^d	
	LPR ^a	LR ^b	UB ^c	LPR	LR	LPR	LR
25	112.14	129.30	129.75	2293.480	61.026	0.1357	0.0035
30	102.35	123.90	123.90	3549.216	104.692	0.1739	0
35	–	151.90	152.40	–	199.112	–	0.0033
40	–	156.34	158.24	–	382.206	–	0.0120
45	–	153.74	154.40	–	533.419	–	0.0043
50	–	161.98	162.34	–	810.795	–	0.0022
60	–	182.16	183.73	–	1482.565	–	0.0086
70	–	213.68	218.28	–	2670.140	–	0.0211
80	–	202.91	210.04	–	3600	–	0.0339

^aLP-Relaxation method

^bLagrangian relaxation algorithm

^cUpper bound heuristic

^dThe gap between lower bound and upper bound

cannot be solved with CPLEX in a reasonable computational time (3600 s), so the gap between lower bounds and upper bounds heuristic are computed.

Based on the presented results, an increase in the number of nodes results in an increase in the CPU running time; HLFPS has higher increases as compared to the LPR, and LPR higher than LR. Commercial solver cannot solve the problems with more than 20 nodes in a reasonable time, and the LP-relaxation method is not effective for problems with more than 30 nodes. As witnessed by the results in Table 2, the LR algorithm can find nearly global optimum solution in all small-size problems. For all the small-size problems, the LR has a better performance than the LPR and can find solutions of higher quality. For problems with 25, 30, 35, 45 and 50 nodes, the LR algorithm terminated by the first termination criterion, whereas problems having 40, 60 and 70 nodes terminated over the second termination criterion, and problems with 80 nodes terminated through the third criterion (Table 3). In conclusion, it can be claimed that the solution of LR algorithm deviates less than 4% from the optimal solution. Figures 1 and 2 illustrate the CPU running time of the discussed solution methods for different problem sizes. The rapid growth of the CPU running time for HLFPS and LPR is evident in Fig. 1 for small-size problems, while the LR has a slow increasing rate. For large-scale problems, LPR has a sharp jump in running time from 25-node to 30-node problem, and it is not effective from that point forward. Similar to the small-size problems, LR shows a slow growth in running time for large-size problems (Fig. 2).

The objective function in the developed HLFPS, is sensitive to the introduced coefficient parameters (α and β). α is the coefficient factor for travel time discount between hubs and β is the time–cost conversion coefficient. To evaluate the sensitivity of the developed HLFPS against each of them, all the small-size problems (due to the optimal solutions have been obtained only for this category of problems) introduced in Table 2 have been solved by the CPLEX for different values of α and β (see Figs. 3, 4, 5, 6).

Figure 3 shows how average number of hubs is influenced by α and β . According to this figure, the number of the constructed hubs is higher when the cost of the tardiness gets higher weight (higher β). This can be because any increase in the number of hubs leads to a

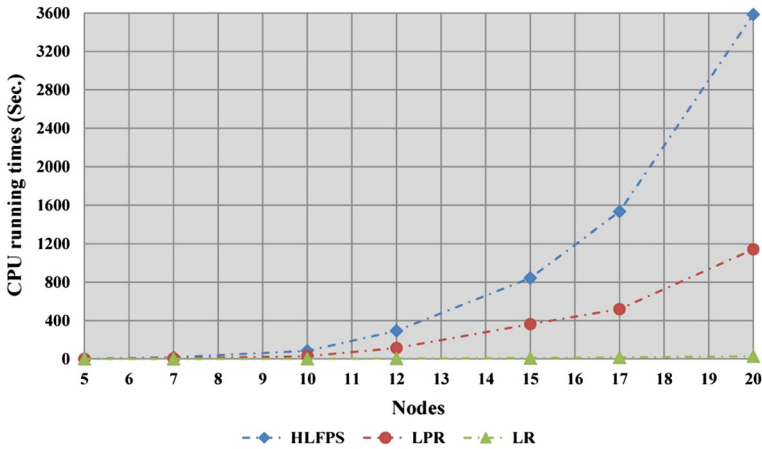


Fig. 1 Investigating the effect of problem dimensions on CPU running time based on different methods (small-size problems)

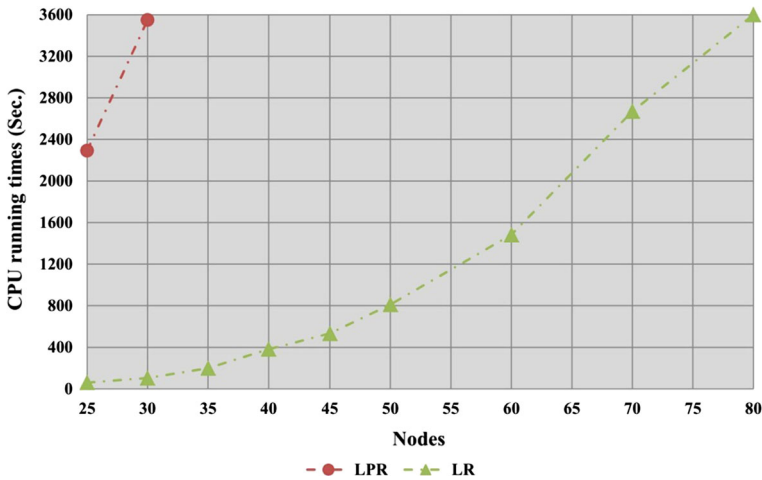


Fig. 2 Investigating the effect of problem dimensions on CPU running time based on different methods (large size problems)

decrease in the traffic of flows available at each hub, thereby decreasing the tardiness. This figure also indicates that for higher α , the model has less tendency to create 2-hub routes, and as a result, the average number of constructed hubs is reduced. Figure 4 displays the changes in the average percentage of two hub routes in the network, affected by two coefficient factors of α and β . As it is evident, for higher β , number of 2-hub routes will be increased compared to the total routes generated to reduce the traffic and hence to decrease the sum of tardiness. Furthermore, by increasing α , the model shows reduced tendency to build 2-hub routes. Figure 5 compares the changes in the average cost of hub construction (SC). The graph shows when the second part of the objective function (the sum of tardiness) has higher weight, constructing more hubs becomes necessary to increase the capacities, so the traffic

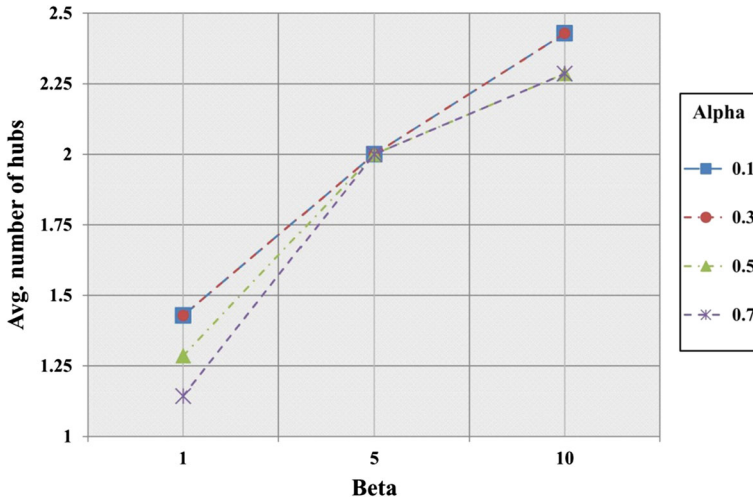


Fig. 3 Investigating the effect of alpha and beta values on the average number of hubs constructed

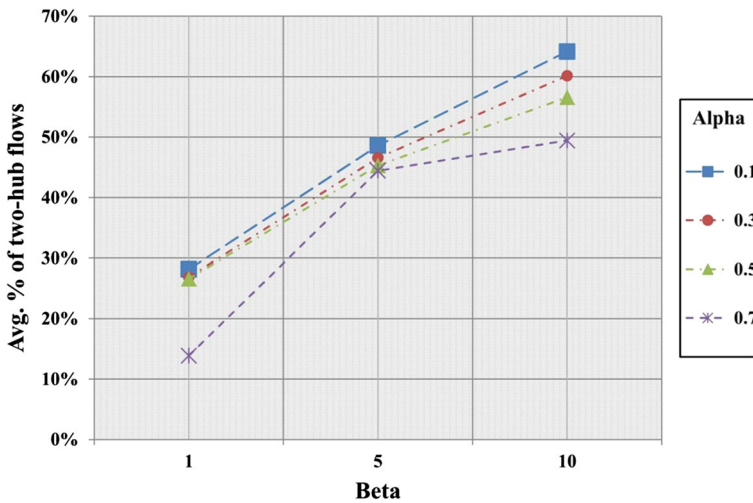


Fig. 4 Investigating the effect of alpha and beta values on the percentage of 2-hub routes

volume of the flows is decreased and, therefore, the sum of tardiness is reduced. It is also obvious that higher α may result in reduced tendency to build 2-hub routes, thereby reduced costs for hub construction. Figure 6 assesses the changes in average total tardiness (TT); an increase in β results in a decreased TT due to increased hub construction and decreased traffic at each hub. However, by increasing α and decreasing number of 2-hub routes, the flow traffic will be increased at each hub and some level of tardiness is expected in the network.

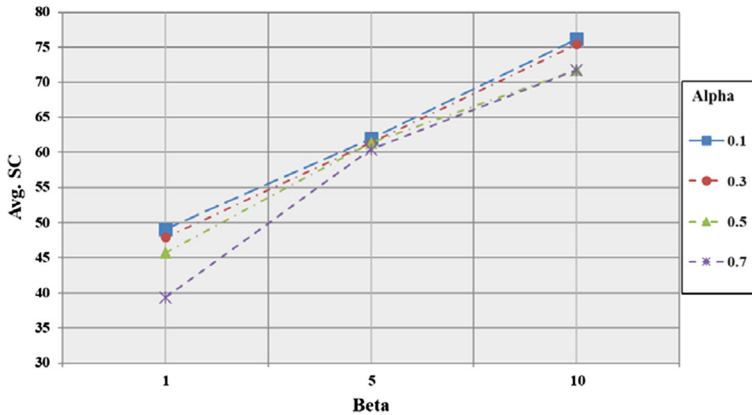


Fig. 5 Investigating the effect of alpha and beta values on the average construction costs

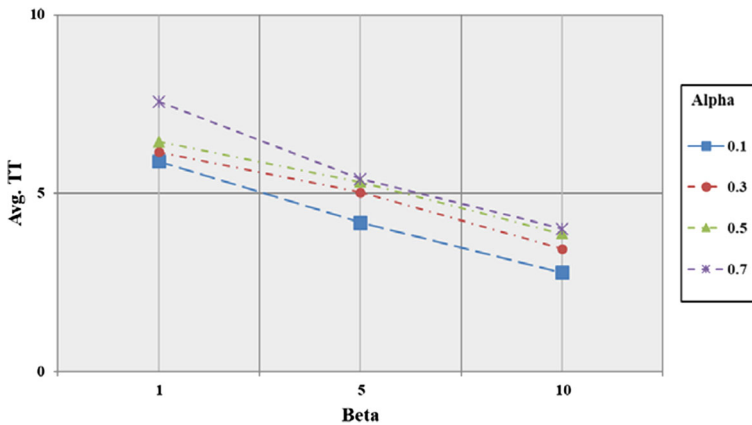


Fig. 6 Investigating the effect of alpha and beta values on the average total tardiness time

5 Conclusion

This paper revisits the hub location problem and proposes an integrated model to schedule the processing flows and select the hub locations. Another contribution of this work is to address a novel approach for modeling the capacity constraint through introducing renewable capacity for the hubs, meaning that the available capacity is dynamically changing when a flow enters or leaves the hub. So, instead of net capacity of the hubs over a period of time, the real-time capacity of the hubs is monitored and scheduling of the process is being done accordingly. The integrated hub location and flow processing schedule problem is formulated based on a mixed-integer linear programming model and a heuristic algorithm is proposed to solve the large-scale problems. The model was solved for a group of small and large size problems to evaluate the performance of different solution methods. The Lagrangian relaxation method obtains very close to optimal solution for small-size problem, where the problem can be solved with GAMS for optimal solution. For large-scale problems, Lagrangian relaxation

method can solve the problems of up to 80 nodes in a reasonable time. Moreover, based on the calculated upper bound for the optimal solution, the obtained solutions deviate from the optimal solution less than 4%. This paper assumes a pattern of the flows can be estimated in the system. The model is solved based on this estimated pattern to suggest location of the hubs to minimize the total cost. This work can be extended by incorporating uncertainty in the model. Particularly, future research can be done to propose a robust design considering risk and uncertainty in the model parameters.

References

- Alumur S, Kara BY (2008) Network hub location problems: the state of the art. *Eur J Oper Res* 190:1–21
- Alumur SA, Campbell JF, Contreras I, Kara BY, Marianov V, O’Kelly ME (2020) Perspectives on modeling hub location problems. *Eur J Oper Res* 291(1):1–17
- Alumur SA, Kara BY, Karasan OE (2009) The design of single allocation incomplete hub networks. *Transp Res Part B* 43:936–951
- Badri H, Bashiri M, Hejazi TH (2013) Integrated strategic and tactical planning in a supply chain network design with a heuristic solution method. *Comput Oper Res* 40:1143–1154
- Brueckner JK, Zhang Y (2001) A model of scheduling in airline networks: how a hub-and-spoke system affects flight frequency, fares and welfare. *J Transport Econ Policy* 35(2):195–222
- CAMPBELL, J. F. (2009) Hub location for time definite transportation. *Comput Oper Res* 36:3107–3116
- Campbell JF, O’Kelly ME (2012) Twenty-five years of hub location research. *Transp Sci* 46:153–169
- Charisis A, Spana S, Kaisar E, Du L (2020) Logistics hub location-scheduling model for innercity last mile deliveries. *Int J Traffic Transport Eng* 10:169–186
- Correia I, Nickel S, Saldanha-da-gama F (2013) Multi-product capacitated single-allocation hub location problems: formulations and inequalities. *Networks Spatial Econ* 14:1–25
- da Graça Costa M, Captivo ME, Clímaco J (2008) Capacitated single allocation hub location problem—a bi-criteria approach. *Comput Oper Res* 35:3671–3695
- Ernst AT, Krishnamoorthy M (1996) Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Locat Sci* 4:139–154
- Farahani RZ, Hekmatfar M, Arabani AB, Nikbakhsh E (2013) Hub location problems: a review of models, classification, solution techniques, and applications. *Comput Ind Eng* 64:1096–1109
- Guignard M (2003) Lagrangean relaxation. *TOP* 11:151–200
- Karimi H, Setak M (2018) A bi-objective incomplete hub location-routing problem with flow shipment scheduling. *Appl Math Model* 57:406–431
- Kumar PR, Sivakumar AI (2013) Simulated annealing-based algorithm for the capacitated hub routing problem. *Int J Serv Oper Manag* 14:221–235
- Levin A (2007) Airline glitches top cause of delays. *USA Today*
- Lu J, Yang Z, Dong X, Zhu X (2015) Design of timetable for airport coach based on ‘time–space’ network and passenger’s trip chain. *Transport* 33:1–9
- Masaeli M, Alumur SA, Bookbinder JH (2018) Shipment scheduling in hub location problems. *Transport Res Part B* 115:126–142
- McWilliams DL Simulation-based scheduling for parcel consolidation terminals: a comparison of iterative improvement and simulated annealing. In: *Proceedings of the 37th conference on Winter simulation, 2005. Winter Simulation Conference, 2087–2093*
- McWilliams DL (2010) Iterative improvement to solve the parcel hub scheduling problem. *Comput Ind Eng* 59:136–144
- McWilliams DL, Stanfield PM, Geiger CD (2005) The parcel hub scheduling problem: a simulation-based solution approach. *Comput Ind Eng* 49:393–412
- McWilliams DL, Stanfield PM, Geiger CD (2008) Minimizing the completion time of the transfer operations in a central parcel consolidation terminal with unequal-batch-size inbound trailers. *Comput Ind Eng* 54:709–720
- O’Kelly ME (1986a) Activity levels at hub facilities in interacting networks. *Geogr Anal* 18:343–356
- O’Kelly ME (1986b) The location of interacting hub facilities. *Transp Sci* 20:92–106
- O’Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. *Eur J Oper Res* 32:393–404
- Regulation FC (2004) Flight compensation regulation. In: UNION, E. P. A. C. O. T. E. (eds) *Regulation (EC) No 261/2004*

- Rostami M, Moradinezhad D, Soufipour A (2014) Improved and competitive algorithms for large scale multiple resource-constrained project-scheduling problems. *KSCE J Civ Eng* 18:1261–1269
- Toh RS, Higgins RG (1985) The impact of hub and spoke network centralization and route monopoly on domestic airline profitability. *Transport J* 24:16–27
- Zhang C, Xie F, Huang K, Wu T, Liang Z (2017) MIP models and a hybrid method for the capacitated air-cargo network planning and scheduling problems. *Transport Res Part E* 103:158–173

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.