Robust Integrated Airline Scheduling with Chance Constraints

Yifan Xu*, Sebastian Wandelt† and Xiaoqian Sun‡

School of Electronic and Information Engineering, Beihang University, Beijing, China
Email: *yifan_buaa@buaa.edu.cn, †wandelt@informatik.hu-berlin.de, ‡sunxq@buaa.edu.cn

Abstract—During planning and execution of their flights, airlines face complex decision-making processes: The operation of all aircraft according to a minimum-cost plan poses tremendous difficulties even for medium-sized airlines. Accordingly, the airline scheduling problem is historically decomposed into a collection of four subproblems: Schedule design, fleet assignment, aircraft routing, and crew pairing problems. The decomposition into subproblems reduces the required computational costs; the airline performance, however, often greatly deteriorates, given the sub-optimality of the solution or, in the worst case, infeasible solutions. With the incorporation of robustness into scheduling, solving the problem becomes further intractable. In this study, we design an integrated robust scheduling problem, which integrates the first three steps into a complex optimization problem, while considering chance constraints to ensure the overall on-time performance. An exact column generation-based algorithm and a fast hybrid algorithm that combines variable neighborhood search with column generation are developed to efficiently solve the problem. Based on a real schedule from a Chinese airline, the benefits of this fully integrated robust model and computation results are presented and validated. Our results demonstrate that the proposed hybrid algorithm can quickly derive high quality solution within an optimality gap of no more than 0.7%.

Keywords—Airline scheduling, Robustness, Chance constraints

I. INTRODUCTION

Surging travel demands and globalization, together with increased environmental concerns, forces air transportation to become more efficient and safer. The number of passengers in scheduled aviation will soon pass the five billion mark; together with the ongoing increase of the number of aircraft in the sky, stakeholders are forced to create more robust schedules; in order to reduce operation costs. Since the overall airline scheduling planning problem is highly computational intractable, it is traditionally transformed into the four sequential subproblems. First, during schedule design, flight timetables are established based on market and airline network, fixing departure and arrival times between airports. Next, at fleet assignment, flights are populated with aircraft types, with the general goal to maximize the airline’s profit. In the third step, called aircraft routing, the actual routes of individual aircraft are constructed, while satisfying maintenance feasibility and other constraints. Finally, during crew scheduling, crew pairs are created with minimum cost which obey all working and safety rules. The four stages are, naturally, highly related to each other; meaning that disturbances of a solution in one stage often induces large impacts in other stages. Therefore, within the last decade, researchers have built integrated models for airline scheduling [1]. At the same time, there has been an increasing interest in managing and mitigating disruptions and other irregular airline operations; cancellations and delay costs are on the rise and put high pressure on airlines’ profits. Such irregular operations include, but are not limited to: hardware failures, convective weather, and heavy traffic flows. Researchers have proposed reactive and proactive approaches for dealing with these issues. The reactive approach attempts to recover the schedule with less cost after the irregularity happens [2]. The proactive approach, on the other hand, builds robust schedules reducing the overall probability of disruptions [3]. It can be concluded that integration and robustness have received increasing attention in research on airline operations recently.

Concerning integrated airline scheduling models, several studies have integrated two or more these subproblems, in a bid to more accurately model the airline scheduling problem. TABLE I summarizes some of the representative studies in this field and a few selected instances are highlighted as follows. Combining passenger demand features and flight schedule adjustment, [4] presented a model that simultaneously optimized flight legs selection and fleet assignment decisions. Coping with dynamic demand fluctuation in a de-banking environment, [5] proposed a mathematical model to re-time the flights’ departure time. [6] integrated schedule design, fleet assignment, and aircraft routing to formulate a realistic model. Demand, delays and deadhead flights were considered while the maintenance requirement was only implicitly ensured.

Recoverability-based approaches usually utilize domain-specific knowledge or feedback modeling structure for airline scheduling robustness. Specifically, [16] presented a fleet assignment model that featured less hub isolation and shorter aircraft cycles which provided more flexibility once disruptions occur. [17], on the other hand, explicitly incorporated the aircraft recovery problem within the tail assignment problem (aircraft routing problem close to the day of operation). In terms of stability-based approaches, [18] developed a stochastic IP model to minimize the total expected propagated delays. [19] set up new models in extreme-value and chance-constrained paradigms to minimize delays for the aircraft routing problem which had the potential for more general
resource-allocation problems.

From the literature, advanced robustness features were still not sufficiently built into the operational integrated scheduling model formulation, allowing for aircraft maintenance constraints and other operational requirements. Moreover, the computational tractability of the integrated robust models is still a challenge for scalable solutions. Fast and reliable heuristic solution techniques are the agenda for this field.

In this paper, we study the robust integrated airline scheduling problem that combines the schedule design, fleet assignment, and aircraft routing subproblems. Through chance constraints, historical delay distribution is able to be modeled explicitly in our study. Therefore, the derived schedule is robust with respect to the user specific service level. Compared to existing literature that optimize the sum of (expected/weighted) propagated delays, e.g. [20], specifying the average service level saves the work of calibrating the cost of delay within the objective function of an integrated model and ensure that individual flights receive balanced propagated delays. To the best of our knowledge, this study is the first to embed chance constraints in an integrated robust problem. In addition, as the number of decision variables of the classic connection network based model depends exponentially on the problem size, a line-of-flights (LOF) network based model, which is similar to the WLOF network in [21] with compact size, is developed in the hybrid solution algorithm framework. This model is able to yield feasible solutions quickly even for large size problem instances.

The major contributions of our study are summarized below.

1) An airline integrated robust scheduling model with chance constraint is developed in this research. The model incorporates optional flight selection, itinerary-based passenger demand, aircraft maintenance rules and propagated delays.
2) A hybrid heuristic algorithm that combines variable neighborhood search (VNS) and column generation is presented to ease the computation burden. This algorithm needs much less computation time compared to the high-quality solutions.
3) The newly proposed LOF network based model is compact and amenable to both VNS and column generation to avoid fully enumeration. From our experimental results, the LOF based model contains a significantly smaller number of variables compared to the traditional connection network-based model used in the literature.

The remainder of this study is organized as follows: Section II introduces the mathematical models of the airline scheduling problem under consideration. Section III presents our solution algorithms. Section IV reports on a case study for a real Chinese legacy airline, summarizing computational experiments and operational insights. Section V concludes this study.

II. THE MATHEMATICAL MODEL

This section presents the specific characteristics and formulations of the airline integrated robust scheduling problem using the framework of robust optimization. We first introduce our notation and then outline the mathematical formulation.

A. Problem description

The airline integrated robust scheduling problem can be formally stated as follows: given 1) a set of aircraft with different fleet types, 2) a set of flights that need to be operated and 3) a set of passenger itineraries consisting of one or two connecting flights, determine the served flights and corresponding aircraft routes, so that the total profits are maximized and the flights are immune against high delays. The airline integrated robust scheduling problem in our study is designed to be solved around one or two weeks prior to the day of operation in order to cope with improved demand forecast, recent operation irregularity and updated information on aircraft resources (e.g., location and maintenance condition).

We define the following terms first. A **passenger itinerary** is composed of one or more consecutive flights serving one specific OD pair. In the scope of our research, we consider both non-stop itineraries and one-stop connecting itineraries that contain one and two flights respectively; itineraries with more than two flights take up a rather small proportion among one-way trips [22]. To satisfy the **maintenance requirements**, aircraft should visit maintenance stations every two to four days with a **minimum maintenance time (MMT)**. Flight delays can be divided into **propagated delays** and **independent delays**. While mechanical problems and bad weather may incur independent delays, the propagated delays are a function of aircraft routes and are incurred due to delays on some prior flights.

Because individual aircraft routes are required to be traced to monitor specific delays and maintenance requirements, connection network is used in this study. The nodes in the network represent flights and airports. A connection network

<table>
<thead>
<tr>
<th>Ref.</th>
<th>SD</th>
<th>FA</th>
<th>AR</th>
<th>CP</th>
<th>Timeline</th>
<th>Model</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>[7]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Tactical</td>
<td>MIP</td>
<td>Diving</td>
</tr>
<tr>
<td>[8]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Tactical</td>
<td>MINP</td>
<td>BD</td>
</tr>
<tr>
<td>[9]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Tactical</td>
<td>MIP</td>
<td>B&amp;B</td>
</tr>
<tr>
<td>[10]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Oper.</td>
<td>MIP</td>
<td>CG</td>
</tr>
<tr>
<td>[12]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Tactical</td>
<td>MIP</td>
<td>BD</td>
</tr>
<tr>
<td>[6]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Oper.</td>
<td>MIP</td>
<td>CG</td>
</tr>
<tr>
<td>[14]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Tactical</td>
<td>MIP</td>
<td>B&amp;B</td>
</tr>
</tbody>
</table>

example with three flights is illustrated in Figure 1. Two types of arcs are created to represent flights’ origin/destination airports (in blue lines) or connecting flights (in red lines) respectively. Two flight nodes \((f_1, f_2)\) are connected if the arrival airport of flight \(f_1\) is the departure airport of flight \(f_2\) and the turn time between these two connecting flights is larger than the minimum turnaround time (MTT). Given the buffer time \(b_{f_1, f_2}\) (turn time minus minimum turnaround time) between flights \((f_1, f_2)\), the independent arrival delays of every flight \(f\) as \(d_f\) and the propagated arrival delay \(PD_f\) can be computed iteratively as follows

\[
PD_f = \max\{PD_{f_1} + d_{f_1} - b_{f_1, f_2}, 0\}
\] (1)

B. Problem formulation

In this section, we propose a connection network based mixed integer nonlinear programming formulation for the airline integrated robust scheduling problem. The summary of the definition of sets, parameters and decision variables is reported in Table II.

Decision variable \(x^{a}_f\) represents the selection of a feasible route (several connecting flights) \(r\) in the connection network from the initial airport of aircraft \(a\). The maintenance requirements are embedded in the route \(r\) with connections \((f_1, f_2)\) that the arrival airport of flight \(f_1\) is a maintenance station and the turn time is larger than the MMT. Consequently, the number of decision variables \(x^{a}_f\) is the number of feasible routes in connection network. The complete formulation of the airline integrated robust scheduling problem is presented below:

\[
\begin{align*}
\text{max} & \quad \sum_{i \in I} fare_i d_i - \sum_{a \in A} \sum_{r \in R} c_r x^{a}_r \\
\text{s.t.} & \quad \sum_{a \in A} \sum_{r \in R} \theta^{1}_{r,f} x^{a}_r = 1 \quad \forall f \in F_M \\
& \quad \sum_{a \in A} \sum_{r \in R} \theta^{1}_{r,f} x^{a}_r = z_f \quad \forall f \in F_O \\
& \quad x^{a}_r \leq 1 \quad \forall a \in A \\
& \quad d_i \leq D_i z_f \quad \forall i \in I, \forall a \in A, \forall f \in F_M \\
& \quad \sum_{i \in I} \sum_{a \in A} \sum_{r \in R} \theta^{2}_{r,f} d_i \leq \sum_{a \in A} S_a \sum_{r \in R} \theta^{1}_{r,f} x^{a}_r \quad \forall f \in F \\
& \quad \frac{\sum_{a \in A} \sum_{r \in R} \theta^{1}_{r,f} x^{a}_r}{|F_M| + \sum_{f \in F_O} z_f} \geq \alpha \\
& \quad x^{a}_r \in \{0, 1\}, y^{f}_{p} \in \{0, 1\}, z_f \in \{0, 1\}, 0 \leq d_i \leq D_i 
\end{align*}
\] (2)

The objective function (2) is the maximization of the total profits as given by the revenue minus direct operating costs. Constraints (3) and (4) are flight covering constraints for mandatory flights and optional flights respectively. In particular, Constraints (3) are set partition constraints that require every mandatory flight being flown by one aircraft. Constraints (4) are set packing constraints that restrict at most one aircraft is assigned to fly one optional flight. Constraints (5) ensure every aircraft can execute at most one aircraft route. Constraints (6) impose restrictions on one-stop passenger itineraries if there exist optional flights in itineraries. Constraints (8) and (9) collect the coverage probability for individual flights under user defined delay threshold. By applying Constraints (10), the average probability is limited to a specific value of \(\alpha\). Constraints (11) indicate the domains of decision variables.

To model uncertainty probabilistically in Constraints (8) and (9), similarly to [19], we define \(p_{r,f}\) as the protection probability of flight \(f\) in route \(r\) is delayed by less than or equal to \(r\) min (user defined delay threshold). This probability can be obtained from a historical operation data-based simulator which is introduced in Section IV. The Constraints (8)-
are rewritten as Constraints (12)-(14):
\[
\sum_{a \in A} \sum_{r \in R} \theta_{r,f}^0 p_r f x_r^a = \beta_f \quad \forall f \in F_M \\
\sum_{a \in A} \sum_{r \in R} \theta_{r,f}^1 p_r f x_r^a = \beta_f \quad \forall f \in F_O \\
\sum_{f \in F_M} \beta_f + \sum_{f' \in F_O} \beta_{f'} \geq \alpha(|F_M| + \sum_{f' \in F_O} z_{f'})
\]

III. SOLUTION METHODOLOGY

Before presenting the solution methods, we first discuss the computational complexity of the airline integrated robust problem. As the set partition problem, which is known to be NP-Complete, is included in the model formulation, the airline scheduling problem in this study is naturally NP-Hard. Additionally, the number of decision variables grows exponentially with the increase of the connection network size. Therefore, in this section, we propose two solution algorithms: a column-generation based exact solution algorithm and a hybrid heuristic.

A. Exact solution algorithm

The exact algorithm solves the model in two steps. Firstly, the linear relaxation problem is solved to optimality using column generation where a subset of decision variables with positive reduced cost (for maximizing the problem) are explored and added iteratively. Then, Branch-and-Bound method is applied for integer solutions.

Let \( \mu_f \) be the dual variables associated with Constraints (3) and (4), \( \eta_a \) be the dual variables of Constraint (5), \( \delta_f \) be the dual variables of Constraints (7) and \( \pi_f \) be the dual variables associated with Constraints (12) and (13). The reduced cost \( \bar{c}_r^a \) of route \( r \) for aircraft \( a \) can be written as

\[
\bar{c}_r^a = -c_r - \sum_{f \in F} \theta_{r,f}^0 \mu_f - \eta_a + S_0 \delta_f - \sum_{f \in F} \theta_{r,f}^1 p_r f \pi_f
\]

The pricing problem is thus to find a route \( r \) which maximizes \( \bar{c}_r^a \) in the connection network of aircraft \( a \). We formulate this problem as a shortest path problem with resource replenishment. A multi-label shortest path algorithm, which is revised based on [23] is illustrated in Algorithm 1.

The algorithm starts by initializing the source node \( s \) with a starting label \( \langle \bar{c}_s, t_s, r_s \rangle \) to track the reduced costs, total elapsed time since last maintenance and aircraft routes respectively. Because the connection network is acyclic, the main loop checks every node in the network according to a topological order. When node \( i \) is checked, its successor node \( j \) is processed to extend \( \langle \bar{c}_j, t_j, r_j \rangle \) as new labels in node \( j \). Maintenance requirements are ensured by discarding labels exceeding limitation on the maximum elapsed time and the covering probability for node \( i \) at route \( r_i \) is computed through the simulator. Additionally, dominance rule is applied to discard unpromising labels with lower reduced cost and higher elapsed time.

Algorithm 1 Multi-label shortest path algorithm.

1: **Input:** Connection Network \( G(V, E) \), Aircraft \( a \), Dual values \( (\mu_f, \eta_a, \delta_f, \pi_f) \)
2: **Output:** aircraft routes with the negative reduced cost
3: Initialize the source node \( s \)'s label set as \( \{ \langle \bar{c}_s, t_s, r_s \rangle \} \)
4: for each node \( i \) of \( G \) in topological sort do
5: for each label \( \{ \langle \bar{c}_i, t_i, r_i \rangle \} \) in node \( i \)'s label set do
6: for each node \( j \) that is a successor of \( i \) do
7: \( r_j \leftarrow \) concatenate \( r_j, i \)
8: \( p_{r,j} \leftarrow \) simulator(\( r_j \))
9: \( \bar{c}_j \leftarrow \bar{c}_i - u_j + S_0 \delta_j - p_{r,j} \pi_j \)
10: if \( i \) satisfy maintenance condition then
11: \( t_j \leftarrow STA_j - STD_j \)
12: else
13: \( t_j \leftarrow t_i + STA_j - STA_i \)
14: end if
15: if \( \langle \bar{c}_j, t_j, r_j \rangle \) is valid and not dominated then
16: Insert \( \langle \bar{c}_j, t_j, r_j \rangle \) in \( j \)'s label pool
17: Remove labels dominated by \( \langle \bar{c}_j, t_j, r_j \rangle \)
18: Mark \( \langle \bar{c}_j, t_j, r_j \rangle \)'s predecessor as \( \langle c_i, t_i, r_i \rangle \)
19: end if
20: end for
21: end for
22: end for
23: Return Sink node \( i \)'s most negative cost label and route

B. Hybrid heuristic

Solving a large-scale airline integrated robust scheduling problem using the exact solution algorithm is very time consuming. Therefore, a hybrid algorithm that combines VNS and column generation is developed in this study to achieve high quality solutions in short time. The main procedures of the proposed methodology are as follows.

1) Create initial aircraft routing solutions for every planning day through a compact mathematical model.
2) Improve daily aircraft routing solutions with a sequential VNS.
3) Concatenate daily aircraft routing solutions through a LOF network-based model.
4) Improve the concatenated solution with VNS again.

The four steps are discussed individually in detail below.

**Step 1:** Initially, the daily aircraft routing solutions are derived from a compact routing model modified based on [15], which relies on the connection network \( G(V, E) \). The formulation of this compact model is given in Constraints (16)-(19).

\[
s.t. \sum_{e \in \Delta^-(v)} x_e = \sum_{e \in \Delta^+(v)} x_e \quad \forall v \in V \quad (16)
\]

\[
\sum_{e \in \Delta^-(s)} x_e = 1 \quad \forall v \in V \setminus \{s, t\} \quad (17)
\]

\[
\sum_{e \in \Delta^-(s)} x_e = |A| \quad (18)
\]
Constraints (16) are the flow balance constraint at every flight vertex \( v \) that restrict the incoming flow on edge \( \delta^+(v) \) equal to the outgoing flow. Flight coverage is ensured by Constraints (17) and Constraints (18), (19) denote the initial and terminating flow (available aircraft number) at the source vertex \( s \) and sink vertex \( t \) respectively.

**Step 2**: Based on the daily aircraft routing solutions, the VNS algorithm is called to improve these solutions iteratively. Three types of neighborhood structures are proposed to generate routing pairs: cross, insert and delete. These neighborhoods provide effective probing to new solutions with varying search depth. Before the discussion, several notations need to introduced: \( dep(f) \) is the departure airport of flight \( f \), \( arr(f) \) is the arrival airport of flight \( f \), and \( con(f_1, f_2) \) states whether \( f_1 \) can be connected to \( f_2 \).

Selecting a pair of aircraft \( a_1, a_2 \) randomly or by rules, the corresponding routing solution is \( r_1 = \{ f_{1,1}, ..., f_{1,n} \} \) and \( r_2 = \{ f_{2,1}, ..., f_{2,m} \} \) respectively. **Cross** and **insert** can be aircraft routes pairs, dealing with partial aircraft routes. **Delete**, however, focuses on deleting optional unpromising flights.

1. **Cross**: Given \( r_1 = \{ f_{1,1}, ..., f_{1,u-1}, f_{1,u}, ..., f_{1,n} \} \), then find \( r_2 = \{ f_{2,1}, ..., f_{2,v-1}, f_{2,v} \} \), that satisfies conditions: (1) \( con(f_{1,u-1}, f_{2,v}) \), (2) \( con(f_{2,v-1}, f_{1,u}) \), to get new solutions:

\[
\begin{align*}
r_1 &= \{ f_{1,1}, ..., f_{1,u-1}, f_{2,v}, ..., f_{2,m} \} \\
r_2 &= \{ f_{2,1}, ..., f_{2,v-1}, f_{1,u}, ..., f_{1,n} \}
\end{align*}
\]

2. **Insert**: Given \( r_1 = \{ f_{1,1}, ..., f_{1,u-1}, f_{1,u}, ..., f_{1,n} \} \), then find \( r_2 = \{ f_{2,1}, ..., f_{2,v-1}, f_{2,v}, ..., f_{2,w}, f_{2,v+1}, ..., f_{2,m} \} \), that satisfies conditions: (1) \( con(f_{1,u-1}, f_{2,v}) \), (2) \( con(f_{2,w'}, f_{1,u}) \), (3) \( arr(f_{2,v-1}) = dep(f_{2,w+1}) \), to get new solutions:

\[
\begin{align*}
r_1 &= \{ f_{1,1}, ..., f_{1,u-1}, f_{2,v}, ..., f_{2,w}, f_{1,u}, ..., f_{1,n} \} \\
r_2 &= \{ f_{2,1}, ..., f_{2,v-1}, f_{2,v+1}, ..., f_{2,m} \}
\end{align*}
\]

3. **Delete**: Given \( r_1 = \{ f_{1,1}, ..., f_{1,u-1}, f_{1,u}, ..., f_{1,n} \} \), then find to remove \( f_{1,u}, f_{1,u+1} \) that satisfies conditions: (1) \( dep(f_{1,u}) = arr(f_{1,u+1}) \) (2) \( f_{1,u} \in F_O, f_{1,u+1} \in F_O \), to get new solutions:

\[
p_1 = \{ f_{1,1}, ..., f_{1,u-1}, f_{1,u+2}, ..., f_{1,n} \}
\]

**Step 3**: To concatenate multiple daily aircraft routing solutions into a complete routing solution spanning the complete planning horizon, a LOF network-based model is proposed. An example LOF network is illustrated in Figure 2. Within the network, dashed black lines represent ground arc while colored lines indicate the movement of aircraft. Assuming an aircraft can fly at most three days without maintenance, every airport at the start/end of the day are duplicated 3 times and surrounded by blue/red boxes. Connections between two airports that cross the night need to record the increase in spanned number of days. Because Airport B is a maintenance station, aircraft visit this airport can fly another three days without maintenance.

The solutions from VNS are treated as LOFs arcs and added to the LOF graph as the solid lines in Figure 2. Given new binary variable \( u^+ \in [0,1] \) to represent the decision of aircraft \( i \) choose LOF \( I \), variables \( g_e \) to denote the number of aircraft on ground arcs or overnight arcs \( e \) and \( l^+_{ev}, l^-_{ev} \) be the binary indicators that whether edge \( e \) starts or ends at vertex \( v \), the complete model formulations are shown in Equation (20) and relating Constraints.

\[
\begin{align*}
\text{max} \sum_{i \in I} fare_i d_i - \sum_{a \in A} \sum_{r \in R} c_i u_i^a & \quad (20) \\
\text{s.t.} \quad \sum_{a \in A \in \text{LOFs}} \sum_{r \in R} \theta_1^a u_i^a = 1 & \quad \forall f \in F_M \quad (21) \\
\sum_{a \in A \in \text{LOFs}} \sum_{r \in R} \theta_1^a u_i^a = z_f & \quad \forall f \in F_O \quad (22) \\
\sum_{a \in A \in \text{LOFs}} \lambda^+_{lu} u_i^a + \lambda^-_{lu} g_e & = \sum_{a \in A \in \text{LOFs}} \lambda^+_{lu} u_i^a + \lambda^-_{lu} g_e & \quad \forall v \in V \quad (23) \\
\sum_{i \in I} \theta_i^3 f_i d_i & \leq \sum_{a \in A \in R} S_i \theta_i^1 f_i u_i^a & \quad \forall f \in F \quad (24)
& \quad + (6), (7), (12) - (14)
\end{align*}
\]

In order to generate new and better LOFs, when the number of solutions generated in Step 2 is not enough, the column generation algorithm is applied to find promising LOFs in different daily connection networks. The algorithm terminates when finding a feasible solution.

**Step 4**: In the last step, concatenated aircraft routings are jointly optimized using the three neighborhoods introduced in Step 2.

**IV. COMPUTATIONAL EXPERIMENTS**

**A. Data description**

We perform computational experiments on a domestic flight network of China Eastern Jiangsu Branch, a legacy commercial airline company in China. The flight network includes 1460 flights per week where the flights are covered by 51 aircraft (three fleet types). The experiments are carried out on five test cases which are generated from the flight schedules in December 2018. Details of these instances are shown in TABLE III.

In particular, the total number of feasible paths within the connection network and LOF network (the sum of paths in every daily network; which is an upper bound to real value since aircraft initial locations are not considered except for the first day) are illustrated in the last two columns. It can be observed that the possible number of paths in LOF network is significantly smaller than that of the connection network except for Instance 1 which spans one day schedule. Hence,
Figure 2. The structure of a LOF network.

TABLE III
CHARACTERISTICS OF TEST INSTANCES

<table>
<thead>
<tr>
<th>Instance</th>
<th>Flights</th>
<th>Aircraft</th>
<th>Airports</th>
<th>Pax itinerary</th>
<th>Optional</th>
<th>Con net paths</th>
<th>LOF net paths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>17</td>
<td>24</td>
<td>66</td>
<td>6</td>
<td>593</td>
<td>593</td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>19</td>
<td>33</td>
<td>149</td>
<td>14</td>
<td>74,145</td>
<td>1,534</td>
</tr>
<tr>
<td>5</td>
<td>245</td>
<td>21</td>
<td>38</td>
<td>261</td>
<td>85</td>
<td>7,419,165</td>
<td>2,660</td>
</tr>
<tr>
<td>3</td>
<td>315</td>
<td>20</td>
<td>40</td>
<td>341</td>
<td>33</td>
<td>879,889,428</td>
<td>3,583</td>
</tr>
<tr>
<td>4</td>
<td>386</td>
<td>24</td>
<td>43</td>
<td>401</td>
<td>38</td>
<td>&gt;1 billion</td>
<td>5,261</td>
</tr>
<tr>
<td>6</td>
<td>612</td>
<td>25</td>
<td>45</td>
<td>642</td>
<td>61</td>
<td>&gt;1 billion</td>
<td>8,556</td>
</tr>
</tbody>
</table>

LOF network-based model is more compact than connection network-based model.

The solution framework is coded in Python 3.7 and executed on a laptop with 2.5GHz Intel i7-6500U CPU in a Fedora 31 system. CPLEX 12.6.3 is used as the solver for solving linear and integer programming models. The multi-label shortest path is coded in C++ and combined with Python with Boost Python module. All experiments are carried out with single thread for the sake of fair comparison.

The ticket price information is derived from Ctrip, a Chinese online travel agency. Optional flights are selected if the total frequency of one market is not larger than two. To better represent real operation situations, BADA [24] is used to calculate the fuel consumption of different aircraft types as direct operation costs.

Besides, to estimate the covering probability, historical on-time performance data of China Eastern for November, 2018 is derived from Chinese Air Traffic Management Bureau. Based on this, independent arrival delays (total delays minus propagated delays) are calculated and used for generating 100 delay scenarios. Therefore, a simulator is applied to estimate the covering probability under these scenarios (i.e. the probability of a flight delayed more than $t$ mins in an aircraft route $r$). The historical independent arrival delays are shown in Figure 3.

B. Results

We report the computational results of the proposed airline integrated robust scheduling model next. We set the protection covering probability value $\alpha = 0.80$ under delay threshold $t \leq 15$min, which is adopted as the on-time performance assessment criterion. TABLE IV reports computation time for six instances using the exact solution algorithms. Branch-and-price is not used in this research due to its slow convergence performance. The statistics in the TABLE IV indicate a
TABLE IV
COMPUTATIONAL RESULTS OF THE EXACT SOLUTION ALGORITHM

<table>
<thead>
<tr>
<th>Instances</th>
<th>Variables</th>
<th>Constraints</th>
<th>LP obj (CNY)</th>
<th>LP time (s)</th>
<th>IP obj (CNY)</th>
<th>total time(s)</th>
<th>relaxation(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>603</td>
<td>200</td>
<td>1,791,518.45</td>
<td>0.76</td>
<td>1,781,218</td>
<td>0.93</td>
<td>0.57</td>
</tr>
<tr>
<td>2</td>
<td>2,947</td>
<td>441</td>
<td>3,833,386.95</td>
<td>13.73</td>
<td>3,792,985</td>
<td>14.47</td>
<td>1.05</td>
</tr>
<tr>
<td>3</td>
<td>6,915</td>
<td>753</td>
<td>6,551,249.73</td>
<td>67.55</td>
<td>5,859,264</td>
<td>71.65</td>
<td>10.56</td>
</tr>
<tr>
<td>4</td>
<td>16,176</td>
<td>957</td>
<td>6,977,290.76</td>
<td>348.31</td>
<td>6,925,912</td>
<td>1770.10</td>
<td>1.32</td>
</tr>
<tr>
<td>5</td>
<td>20,021</td>
<td>1,167</td>
<td>11,255,726.02</td>
<td>459.73</td>
<td>11,153,887</td>
<td>5303.95</td>
<td>0.905</td>
</tr>
<tr>
<td>6</td>
<td>&gt;128,160</td>
<td>1,851</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*: the computation time exceeds a maximum of 10,000 seconds.

TABLE V
COMPUTATIONAL RESULTS OF THE HYBRID HEURISTIC

<table>
<thead>
<tr>
<th>Instances</th>
<th>without VNS</th>
<th>with VNS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obj (CNY)</td>
<td>Time (s)</td>
</tr>
<tr>
<td>1</td>
<td>1,767,842</td>
<td>2.68</td>
</tr>
<tr>
<td>2</td>
<td>3,782,470</td>
<td>9.67</td>
</tr>
<tr>
<td>3</td>
<td>5,342,091</td>
<td>21.64</td>
</tr>
<tr>
<td>4</td>
<td>6,905,152</td>
<td>49.04</td>
</tr>
<tr>
<td>5</td>
<td>11,049,320</td>
<td>78.81</td>
</tr>
<tr>
<td>6</td>
<td>17755275</td>
<td>237.75</td>
</tr>
</tbody>
</table>

*: no corresponding statistics.

fast-increasing complexity of the problem with the growth of the instance size. From Instance 3, the exact solution algorithm spends a lot of time on generating many aircraft route variables and has tremendous difficulties to solve the corresponding integer programming model. In addition, the gap between LP optimal solutions and IP optimal solutions tends to grow significantly large in Instance 3 due to the aircraft capacity Constraints (7).

Given the benchmark solutions of the exact solution algorithm, the performance of our proposed hybrid heuristic algorithm is shown in TABLE V in terms of solution time and solution quality which is indicated by the relative gap between IP obj in TABLE IV and Obj in TABLE V. To demonstrate the impact of VNS, a heuristic composed of only LOF network-based model is formulated.

TABLE V demonstrates that VNS balances the average gap between heuristic solutions and solutions from the exact solution algorithm. Although in two cases, the heuristic without VNS derives better solutions, which is because more variables are added from the pricing problem to make the LOF model become feasible and potentially improve the objective function value for small or medium size instances, the high deviation in Instance 2 indicates its instability to be safely used as a good candidate solution algorithm. Overall, our proposed heuristic is able to derive fast solutions within a 0.7% relative gap to the benchmark accurate solution.

C. Sensitivity Analysis

The main purpose for the sensitivity analysis on the protection probability $\alpha$ is to investigate if propagated delays can be alleviated and correlated on-time performance can be improved when increasing the value of $\alpha$. Therefore, the analysis is carried out by changing the $\alpha$ from 78% to 86% on the $t = 15$min delay threshold and evaluate the on-time performance on the following times of delays: 15 min, 30 min, 60 min by solving the proposed airline integrated robust scheduling model. These boundary values usually serve as assessment criteria or good approximations for flight delays, passenger connection and crew sit time (changing aircraft). The delay scenarios are again generated

TABLE VI
AVERAGE ON-TIME PERFORMANCE FOR SIX INSTANCES

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Flight on time performance(%)</th>
<th>Avg delays(min)</th>
<th>Flights (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>88.33</td>
<td>96.80</td>
<td>100.00</td>
</tr>
<tr>
<td>80</td>
<td>91.47</td>
<td>97.38</td>
<td>100.00</td>
</tr>
<tr>
<td>82</td>
<td>95.12</td>
<td>99.40</td>
<td>100.00</td>
</tr>
<tr>
<td>84</td>
<td>98.24</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>86</td>
<td>99.92</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Figure 4. Impact of protection level on delays and profits.
by sampling the historical operation data of China Eastern for every instance. To facilitate the analysis, all flights become optional which make the model always feasible and we report the statistics regarding on-time performance, average delays and the percentage of executed flights in TABLE VI.

The table indicates that increasing the value of \( \alpha \) monotonically improves the on-time performance. Due to the fact that most aircraft routes experience low delay values, the derived average percentage of on-time flights is higher than the value of \( \alpha \). On the other hand, larger \( \alpha \) may prevent airlines from operating flights with higher arrival delays which may pose a negative impact on the profits. The relationship between delays and profits under different values of \( \alpha \) are visualized in Figure 4 after normalizing the total delays and objective values for the six test instances under five \( \alpha \) values. From the figure, 80% and 82% are suitable protection levels for our problem where the overall delays witness a large drop without sacrificing profits too much.

V. CONCLUSION

In this paper, we proposed an airline integrated robust scheduling problem that jointly selects optional flights, designs aircraft routes and reduces delay propagation. Maintenance requirement and itinerary based passenger demand are also taken into account. The high complexity of the model motivates us to design scalable decomposition technique-based solution algorithms. Two solution algorithms that builds on column generation and VNS are developed and tested on real world airline instances. Low optimality gap and improved on-time performance statistics in computation for a real Chinese airline validate our proposed models and algorithms. The presented model and solution algorithm can be used by small and medium-sized airline companies for operational scheduling decisions to improve profits and eliminate propagated delays. Regarding potential future research directions, further integrating crew scheduling is expected to achieve global optimal solutions with less operation cost and lower delays by considering crew sitting time.

ACKNOWLEDGEMENT

This study is supported by the Research Fund from National Natural Science Foundation of China (Grants No. 61861136005, No. 61851110763, No. 71731001).

REFERENCES