9th International Conference for Research in Air Transportation

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

TABLE I
SUMMARY OF LITERATURE FOR INTEGRATED AIRLINE SCHEDULING.

Ref.	SD	FA	AR	CP	Timeline	Model	Algorithm
[4]	✓	✓			Tactical	MINP	BD
[7]		✓	✓		Tactical	MIP	Diving
[5]	✓	✓			Oper.	MIP	B&B
[8]		✓	✓		Tactical	MINP	BD
[9]	✓	✓			Tactical	MIP	B&B
[10]			✓	✓	Oper.	MIP	CG
[11]		✓	✓	✓	Oper.	MIP	B&P
[12]	✓	✓			Tactical	MIP	BD
[13]	✓	✓	✓		Tactical	MIP	B&P
[6]	✓	✓	✓		Oper.	MIP	CG
[14]	✓	✓			Tactical	MIP	B&B
[15]			✓	✓	Tactical	MIP	B&B, CG
Ours	✓	✓	✓		Oper.	MINP	CG, VNS

SD: Schedule Design, **FA**: Fleet Assignment, **AR**: Aircraft Routing, **CP**: Crew Pairing; **MIP**: Mixed Integer Programming, **MINP**: Mixed Integer Nonlinear Programming; **CG**: Column Generation, **BD**: Benders Decomposition,**B&P**: Branch-and-Price; **B&B**: Branch-and-Bound, **VNS**: Variable neighborhood search.

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 classic column generation algorithm for obtaining 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

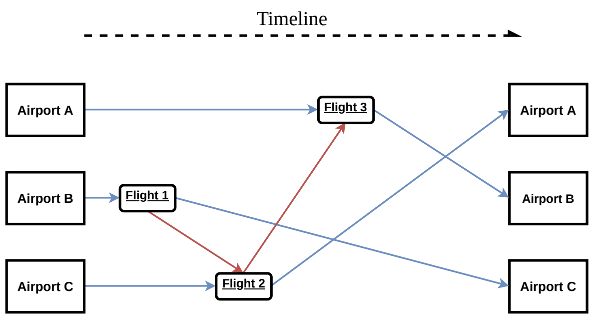

Figure 1. The structure of a connection network.

TABLE II
NOTATIONS USED FOR EXPRESSING THE MODEL

Sets:	
I	set of passenger itineraries, $i \in I$
A	set of aircraft, $a \in A$
R	set of aircraft routes, $r \in R$
F	set of flights, $f \in F$
F_O	set of optional flights, $f \in F_O \subset F$
F_M	set of mandatory flights, $f \in F_M \subset F$
I_C	set of connecting itineraries, $i \in I_C$
Parameters	
D_i	demand of passenger itinerary $i \in I$
S_a	number of seats in aircraft $a \in A$
c_r	operation cost of route $r \in R$
$fare_i$	average fare of itinerary i
α	user-defined protection parameter
$\theta^1_{r,f}$	binary indicator that takes value 1 if flight f is included in route r and 0 otherwise
$\theta^2_{i,f}$	binary indicator that takes value 1 if flight f is included in itinerary i and 0 otherwise
Decision variables:	
x^a_r	1 if aircraft a flies route r and 0 otherwise
z_f	1 if the optional flight $f \in F_O$ is flown and 0 otherwise
d_i	number of passengers taken on itinerary $i \in I$
β_f	covering probability of flight $f \in F$

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

$$PD_{f_2} = \max\{PD_{f_1} + d_{f_1} - b_{f_1,f_2}, 0\} \tag{1}$$

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_r 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_r is the number of feasible routes in connection network. The complete formulation of the airline integrated robust scheduling problem is presented below:

$$\max \sum_{i\in I} fare_i d_i - \sum_{a\in A}\sum_{r\in R} c_r x^a_r \tag{2}$$

$$s.t. \sum_{a\in A}\sum_{r\in R} \theta^1_{r,f} x^a_r = 1 \qquad \forall f \in F_M \tag{3}$$

$$\sum_{a}\sum_{r\in R} \theta^1_{r,f} x^a_r = z_f \qquad \forall f \in F_O \tag{4}$$

$$\sum_{r\in R} x^a_r \le 1 \qquad \forall a \in A \tag{5}$$

$$d_i \le D_i z_f \qquad \forall i \in I_C, \forall \alpha^i_f = 1, f \in F_O \tag{6}$$

$$\sum_{i\in I} \theta^2_{i,f} d_i \le \sum_{a\in A} S_a \sum_{r\in R} \theta^1_{r,f} x^a_r \qquad \forall f \in F \tag{7}$$

$$\boldsymbol{P}\Big(\sum_{a\in A}\sum_{r\in R} \theta^1_{r,f} x^a_r = 1\Big) = \beta_f \qquad \forall f \in F_M \tag{8}$$

$$\boldsymbol{P}\Big(\sum_{a\in A}\sum_{r\in R} \theta^1_{r,f} x^a_r = z_f\Big) = \beta_f \qquad \forall f \in F_O \tag{9}$$

$$\frac{\sum_{f\in F_M} \beta_f + \sum_{f'\in F_O} \beta_{f'} z_{f'}}{|F_M| + \sum_{f'\in F_O} z_{f'}} \ge \alpha \tag{10}$$

$$x^a_r \in \{0,1\}, y^b_p \in \{0,1\}, z_f \in \{0,1\}, 0 \le d_i \le D_i \tag{11}$$

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 α. 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 t 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)-

(10) are rewritten as Constraints (12)-(14):

$$\sum_{a \in A} \sum_{r \in R} \theta^1_{r,f} p_{r,f} x^a_r = \beta_f \qquad \forall f \in F_M \tag{12}$$

$$\sum_{a \in A} \sum_{r \in R} \theta^1_{r,f} p_{r,f} x^a_r = \beta_f \qquad \forall f \in F_O \tag{13}$$

$$\sum_{f \in F_M} \beta_f + \sum_{f' \in F_O} \beta_{f'} \geq \alpha(|F_M| + \sum_{f' \in F_O} z_{f'}) \tag{14}$$

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 x^a_r 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 μ_f be the dual variables associated with Constraints (3) and (4), η_a be the dual variables of Constraint (5), δ_f be the dual variables of Constraints (7) and π_f be the dual variables associated with Constraints (12) and (13). The reduced cost $\bar{c}^a_r$ of route *r* for aircraft *a* can be written as

$$\bar{c}^a_r = -c_r - \sum_{f \in F} \theta^1_{r,f} \mu_f - \eta_a + S_a \delta_f - \sum_{f \in F} \theta^1_{r,f} p_{r,f} \pi_f \tag{15}$$

The pricing problem is thus to find a route *r* which maximizes $\bar{c}^a_r$ 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 $\langle \bar{c}_i, t_i, r_i \rangle$ 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.

Input: Connection Network G(V, E), Aircraft *a*, Dual values $(\mu_f, \eta_a, \delta_f, \pi_f)$
Output: aircraft routes with the negative reduced cost
Initialize the source node *s*'s label set as $\{\langle \bar{c}_s, t_s, r_s \rangle\}$
for each node *i* of G in topological sort **do**
    for each label $\{\langle \bar{c}_i, t_i, r_i \rangle\}$ in node *i*'s label set **do**
        for each node *j* that is a successor of *i* **do**
            $r_j \leftarrow$ concatenate r_i, j
            $p_{r,j} \leftarrow$ simulator(r_j)
            $\bar{c}_j \leftarrow \bar{c}_i - u_j + S_a \delta_j - p_{r,j} \pi_j$
            if *i* satisfy maintenance condition **then**
                $t_j \leftarrow STA_j - STD_j$
            else
                $t_j \leftarrow t_i + STA_j - STA_i$
            end if
            if $\langle \bar{c}_j, t_j, r_j \rangle$ is valid and not dominated **then**
                Insert $\langle \bar{c}_j, t_j, r_j \rangle$ in *j*'s label pool
                Remove labels dominated by $\langle \bar{c}_j, t_j, r_j \rangle$
                Mark $\langle \bar{c}_j, t_j, r_j \rangle$'s predecessor as $\langle \bar{c}_i, t_i, r_i \rangle$
            end if
        end for
    end for
end for
Return Sink node *t*'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).

$$s.t. \sum_{e \in \delta^-(v)} x_e = \sum_{e \in \delta^+(v)} x_e \qquad \forall v \in V \tag{16}$$

$$\sum_{e \in \delta^-(v)} x_e = 1 \qquad \forall v \in V/\{s,t\} \tag{17}$$

$$\sum_{e \in \delta^-(s)} x_e = |A| \tag{18}$$

$$\sum_{e\in\delta^+(t)} x_e = |A| \tag{19}$$

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 $con(f_i, f_j)$ states whether f_i can be connected to f_j.

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 unpromising optional flights.

(1) *Cross*: Given $r_1 = \{f_{1,1}, ..., f_{1,u-1}, f_{1,u}, ..., f_{1,n}\}$, then find $r_2 = \{f_{2,1}, ..., f_{2,v-1}, f_{2,v}, ..., f_{2,m}\}$, 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:

$$r_1 = \{f_{1,1}, ..., f_{1,u-1}, f_{2,v}, ..., f_{2,m}\}$$
$$r_2 = \{f_{2,1}, ..., f_{2,v-1}, f_{1,u}, ..., f_{1,n}\}$$

(2) *Insert*: Given $r_1 = \{f_{1,1}, \ldots, f_{1,u-1}, f_{1,u}, \ldots, f_{1,n}\}$, then find $r_2 = \{f_{2,1}, ..., f_{2,v-1}, f_{2,v}, ..., f_{2,w}, f_{2,w+1}, ..., f_{2,m}\}$, 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:

$$r_1 = \{f_{1,1}, ..., f_{1,u-1}, f_{2,v}, ..., f_{2,w}, f_{1,u}, ..., , f_{2,m}\}$$
$$r_2 = \{f_{2,1}, ..., f_{2,v-1}, f_{2,w+1}, ..., f_{2,n}\}$$

(3) *Delete*: Given $r_1 = \{f_{1,1}, ..., f_{1,u}, f_{1,u+1}, ..., f_{1,n}\}$, 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 $u_l^a \in 0, 1$ to represent the decision of aircraft *a* choose LOF *l*, variables g_e to denote the number of aircraft on ground arcs or overnight arcs *e* and $\lambda_{ev}^+, \lambda_{ev}^-$ be the binary indicators that whether edge *e*/LOF starts or ends at vertex *v*, the complete model formulations are shown in Equation (20) and relating Constraints.

$$\max \sum_{i\in I} fare_i d_i - \sum_{a\in A}\sum_{r\in R} c_l u_l^a \tag{20}$$

$$s.t. \sum_{a\in A}\sum_{l\in LOFs} \theta_{lf}^1 u_l^a = 1 \qquad \forall f \in F_M \tag{21}$$

$$\sum_{a\in A}\sum_{l\in LOFs} \theta_{lf}^1 u_l^a = z_f \qquad \forall f \in F_O \tag{22}$$

$$\sum_{a\in A}\sum_{l\in LOFs} \lambda_{lv}^+ u_l^a + \sum_{e\in E} \lambda_{ev}^+ g_e = \sum_{a\in A}\sum_{l\in LOFs} (\lambda_{lv}^- u_l^a + \sum_{e\in E} \lambda_{ev}^- g_e) \qquad \forall v \in V \tag{23}$$

$$\sum_{i\in I} \theta_{i,f}^3 d_i \le \sum_{a\in A}\sum_{r\in R} S_a \theta_{l,f}^1 u_l^a \qquad \forall f \in F \tag{24}$$

$$(6), (7), (12) - (14)$$

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,

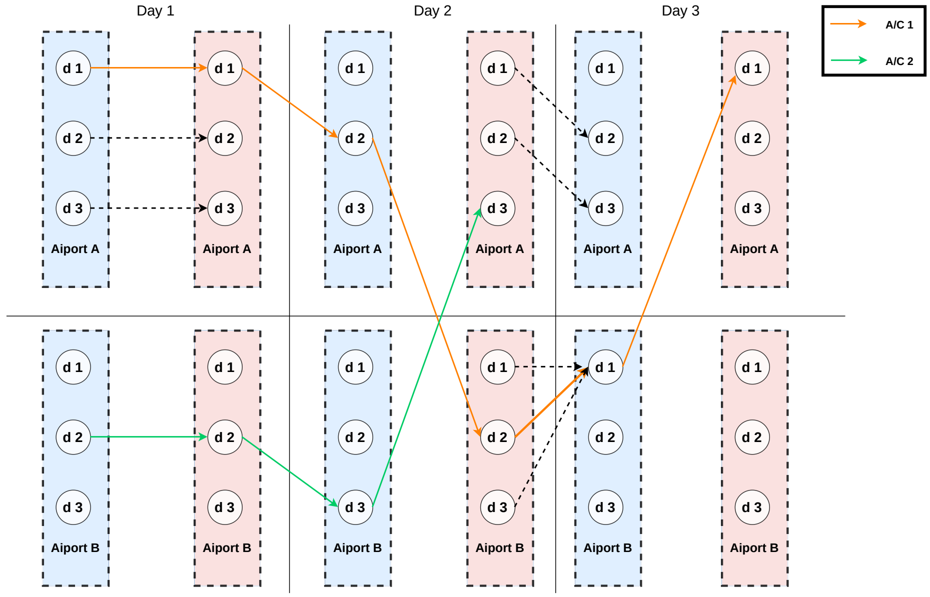

Figure 2. The structure of a LOF network.

TABLE III
CHARACTERISTICS OF TEST INSTANCES

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

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.

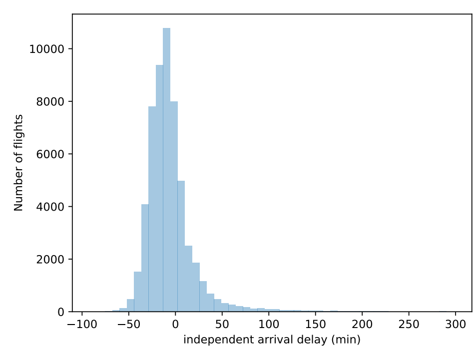

Figure 3. The independent arrival delays for China Eastern in Nov 2018.

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

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

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

TABLE V
COMPUTATIONAL RESULTS OF THE HYBRID HEURISTIC

Instances	without VNS			with VNS		
	Obj (CNY)	Time (s)	Gap (%)	Obj (CNY)	Time(s)	Gap (%)
1	1,767,842	2.68	0.75	1,771,050	1.59	0.57
2	3,782,470	9.67	0.28	3,771,555	10.94	0.56
3	5,342,091	21.64	8.83	5,825,245	26.85	0.58
4	6,905,152	49.04	0.30	6,880,234	46.45	0.66
5	11,049,320	78.81	0.94	11,106,239	73.20	0.43
6	17755275	237.75	-	17,836,340	240.41	-

-: 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 α is to investigate if propagated delays can be alleviated and correlated on-time performance can be improved when increasing the value of α. Therefore, the analysis is carried out by changing the α from 78%

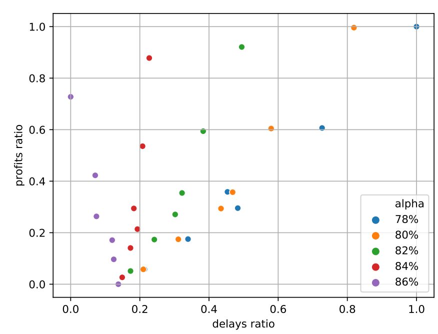

Figure 4. Impact of protection level on delays and profits.

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

α	Flight on time performance(%)			Avg delays(min)	Flights (%)
	$\leq$15 min	$\leq$30 min	$\leq$60min		
78	88.53	96.80	100.00	9.99	99.26
80	91.47	97.38	100.00	9.27	99.60
82	95.12	99.40	100.00	7.37	93.99
84	98.24	100.00	100.00	6.14	83.51
86	99.92	100.00	100.00	5.38	68.61

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 α 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 α. On the other hand, larger α 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 α are visualized in Figure 4 after normalizing the total delays and objective values for the six test instances under five α 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

[1] C. Barnhart, F. Lu, and R. Shenoi, "Integrated airline schedule planning," in *Operations research in the airline industry*. Springer, 1998, pp. 384–403.

[2] Z. Liang, F. Xiao, X. Qian, L. Zhou, X. Jin, X. Lu, and S. Karichery, "A column generation-based heuristic for aircraft recovery problem with airport capacity constraints and maintenance flexibility," *Transportation Research Part B: Methodological*, vol. 113, pp. 70–90, 2018.

[3] O. Khaled, M. Minoux, V. Mousseau, S. Michel, and X. Ceugniet, "A multi-criteria repair/recovery framework for the tail assignment problem in airlines," *Journal of Air Transport Management*, vol. 68, pp. 137–151, 2018.

[4] H. D. Sherali, K.-H. Bae, and M. Haouari, "Integrated airline schedule design and fleet assignment: Polyhedral analysis and benders' decomposition approach," *INFORMS Journal on Computing*, vol. 22, no. 4, pp. 500–513, 2010.

[5] H. Jiang and C. Barnhart, "Robust airline schedule design in a dynamic scheduling environment," *Computers & Operations Research*, vol. 40, no. 3, pp. 831–840, 2013.

[6] N. Kenan, A. Jebali, and A. Diabat, "The integrated aircraft routing problem with optional flights and delay considerations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 355–375, 2018.

[7] Z. Liang and W. A. Chaovalitwongse, "A network-based model for the integrated weekly aircraft maintenance routing and fleet assignment problem," *Transportation Science*, vol. 47, no. 4, pp. 493–507, 2012.

[8] H. D. Sherali, K.-H. Bae, and M. Haouari, "An integrated approach for airline flight selection and timing, fleet assignment, and aircraft routing," *Transportation Science*, vol. 47, no. 4, pp. 455–476, 2013.

[9] J. P. Pita, N. Adler, and A. P. Antunes, "Socially-oriented flight scheduling and fleet assignment model with an application to norway," *Transportation Research Part B: Methodological*, vol. 61, pp. 17–32, 2014.

[10] S. Ruther, N. Boland, F. G. Engineer, and I. Evans, "Integrated aircraft routing, crew pairing, and tail assignment: branch-and-price with many pricing problems," *Transportation Science*, vol. 51, no. 1, pp. 177–195, 2016.

[11] V. Cacchiani and J.-J. Salazar-González, "Optimal solutions to a real-world integrated airline scheduling problem," *Transportation Science*, vol. 51, no. 1, pp. 250–268, 2016.

[12] L. Cadarso and R. de Celis, "Integrated airline planning: Robust update of scheduling and fleet balancing under demand uncertainty," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 227 – 245, 2017.

[13] O. Faust, J. Goñsch, and R. Klein, "Demand-oriented integrated scheduling for point-to-point airlines," *Transportation Science*, vol. 51, no. 1, pp. 196–213, 2017.

[14] K. Wei, V. Vaze, and A. Jacquillat, "Airline timetable development and fleet assignment incorporating passenger choice," *Transportation Science*, 2019.

[15] A. Parmentier and F. Meunier, "Aircraft routing and crew pairing: updated algorithms at air france," *Omega*, 2019.

[16] J. M. Rosenberger, E. L. Johnson, and G. L. Nemhauser, "A robust fleet-assignment model with hub isolation and short cycles," *Transportation science*, vol. 38, no. 3, pp. 357–368, 2004.

[17] G. Froyland, S. J. Maher, and C.-L. Wu, "The recoverable robust tail assignment problem," *Transportation Science*, vol. 48, no. 3, pp. 351–372, 2013.

[18] S. Lan, J.-P. Clarke, and C. Barnhart, "Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions," *Transportation Science*, vol. 40, no. 1, pp. 15–28, 2006.

[19] L. Marla, V. Vaze, and C. Barnhart, "Robust optimization: Lessons learned from aircraft routing," *Computers & Operations Research*, vol. 98, pp. 165–184, 2018.

[20] Y. Xu, S. Wandelt, and X. Sun, "Stochastic tail assignment under recovery," in *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)*, 2019.

[21] Z. Liang, Y. Feng, X. Zhang, T. Wu, and W. A. Chaovalitwongse, "Robust weekly aircraft maintenance routing problem and the extension to the tail assignment problem," *Transportation Research Part B: Methodological*, vol. 78, pp. 238–259, 2015.

[22] C. Barnhart, D. Fearing, and V. Vaze, "Modeling passenger travel and delays in the national air transportation system," *Operations Research*, vol. 62, no. 3, pp. 580–601, 2014.

[23] M. Dunbar, G. Froyland, and C.-L. Wu, "Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework," *Transportation Science*, vol. 46, no. 2, pp. 204–216, 2012.

[24] A. Nuic, D. Poles, and V. Mouillet, "BADA: An advanced aircraft performance model for present and future atm systems," *International journal of adaptive control and signal processing*, vol. 24, no. 10, pp. 850–866, 2010.