

Delay Averse Crew Pairing Optimization

by

Onur Kilinc

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2025

© Onur Kilinc 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Airline operations involve many complex and interdependent decisions. These include flight scheduling, fleet assignment, maintenance routing, and crew scheduling. Crew scheduling is particularly important, as crew costs are the second-largest expense in airline operations. In practice, delays are common and can cause major disruptions to crew schedules. Most crew pairing models are deterministic and aim to minimize planned costs. However, they often ignore operational uncertainties, such as delays. As a result, they may perform poorly in real operations. This is critical because crew payments are based on actual flight times. This thesis addresses the delay averse crew pairing problem. We propose an optimization model with a delay cost component to create more reliable pairings. The model is based on a duty network structure, and utilizes the pay-and-credit scheme to account for crew payments. We solve the model using Lagrangian relaxation and perform extensive testing on real data. We compare the delay averse model to the nominal case and conduct sensitivity analysis to demonstrate how changes in cost parameters and delay levels affect the solutions. Finally, we evaluate the solutions using simulation based on historical flight delay data. The results show that the delay averse model creates more reliable pairings. These pairings are better at absorbing delays and reducing total propagated delays.

Acknowledgements

I wish to express my sincere gratitude to my advisor Prof. Fatma Gzara for her invaluable guidance, unwavering support, patience, and constant encouragement throughout the completion of this thesis. I would like to sincerely thank Prof. Samir Elhedhli and Asst. Prof. Sirisha Rambhatla for generously dedicating their time to read and review this thesis.

I extend my profound gratitude to my family for their relentless and unwavering support. Despite the physical distance, their love and encouragement have been a constant source of strength.

I would like to extend my heartfelt thanks to my colleagues at the University of Waterloo for fostering such a nurturing and supportive environment throughout this journey.

Dedication

To my family, for their endless love and support.

Table of Contents

Author’s Declaration	ii
Abstract	iii
Acknowledgements	iv
Dedication	v
List of Figures	viii
List of Tables	ix
1 Introduction	1
1.1 Motivation and Objective	1
1.2 Introduction to Crew Pairing	3
1.2.1 Network Structure	3
1.2.2 Formulation	6
1.3 Contributions and Thesis Content	8

2	Literature Review	10
3	Modelling Crew Pairing Optimization	17
3.1	Arc-Based Formulation	17
3.2	Path Based Formulation	22
3.3	Delay Averse Crew Pairing Optimization	25
4	Solution Methodology	27
4.1	Lagrangian Relaxation	27
4.2	Cutting Plane Algorithm	29
5	Case Study	33
5.1	Experimental Setup	34
5.2	Computational Testing	35
5.3	Simulation	43
6	Conclusions	53
	References	55

List of Figures

1.1	Example Flight Network Lu [2014]	5
1.2	Example Duty Network Lu [2014]	6
5.1	Sensitivity Analysis Results	41
5.2	Total Propagated Delays for 75R,25IR Scenarios (70,80,90)	50
5.3	Total Propagated Delays for 50R,50IR Scenarios (70,80,90)	50
5.4	Total Propagated Delays for 25R,75IR Scenarios (70,80,90)	51
5.5	Total Propagated Delays for 100IR Scenarios (70,80,90)	51
5.6	Total Propagated Delays for 100R	52

List of Tables

1.1	Example Flight Schedule Lu [2014]	4
1.2	Example Duties Lu [2014]	5
2.1	Selected Studies on Airline Crew Planning and Recovery Methods	11
3.1	Description of Sets, Parameters, and Decision Variables	20
3.2	Description of Sets, Parameters, and Decision Variables	23
5.1	Crew pairing model parameters	34
5.2	Cutting Plane Algorithm Solution Statistics	36
5.3	Sensitivity Analysis for $\gamma = 0, 5, 10$ and 90th-99th Delay Percentiles	39
5.4	Flight coverage $\gamma = 5, 10$:	42
5.5	Total Propagated Delays for 75R,25IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD	46
5.6	Total Propagated Delays for 50R,50IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD	47

5.7	Total Propagated Delays for 25R,75IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD	48
5.8	Total Propagated Delays for 100IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD	49
5.9	Total Propagated Delays for 100R Scenario;pth represents the pth percentile of the TPGD	49

Chapter 1

Introduction

1.1 Motivation and Objective

Airline planning includes a sequence of interdependent decision-making processes, with each phase depending on the outcomes of the phase before it [Hall \[2012\]](#). The process begins with flight schedule design, where airlines develop a recurring flight plan covering a period of one to two months. Once the flight schedule is determined, the fleet assignment phase follows. In this phase, airlines determine which specific aircraft will operate each scheduled flight. The third phase, aircraft routing, establishes the sequence of flights for each aircraft, ensuring that they follow regulatory time limits for required maintenance checks.

The next step is to assign crews to these flights. These crew members must follow strict rules such as minimum connection time between flights, maximum flight time and rest time between duties. The objective of crew scheduling problem is to cover every flight with an assigned crew while keeping the total cost low. This step is very important because crew-

related costs are the second highest expense in airline operations. In airline operations, disruptions can impact flight schedules and crew assignments. For this reason, airlines implement several strategies such as using reserve crews, reassigning or swapping crew assignments (Shebalov and Klabjan [2006b], Medard and Sawhney [2007]). In our study, we consider propagated delays to deal with flight delays. That means if a crew is assigned to a flight but cannot connect to the next flight on time because of delays, the next flight’s departure is delayed until the crew becomes available. We also implement a realistic and complex pay scheme. Pay schemes are often negotiated through collective bargaining between airline management and crew unions. For instance, a recent near-strike by Air Canada pilots’ unions has shown the importance of crew payment structures Reuters [2024]. Union rules in the airline industry outline how delays or early arrivals affect crew members’ work hours, compensation, and rest requirements.

Operations Research has contributed to solving the airline crew pairing problem. Many optimization frameworks and algorithms have been used to find low-cost crew pairings Kohl et al. [2007b]. These include deterministic models, stochastic programming models, robust optimization, and bi-objective models. Each of them focuses on different aspects, such as cost, uncertainty, operational flexibility, or recovery. In our study, we develop an integer programming model with a delay risk aversion component. We modify the deterministic crew pairing formulation to consider the impact of delays. We add a delay cost component to the objective function to penalize flight connections that are less reliable. We obtain the cost components by relying on the Bureau of Transportation Statistics BTS [2024] data on historical delays for the flights. In the deterministic crew pairing problem, the pairing costs are calculated using only scheduled times, without considering any operational disruptions. We also incorporate a realistic payment structure by working on a duty network. We aim to obtain crew schedules that are more risk-averse to delays and cost-efficient. Lastly, we

develop a simulation framework to test the model’s performance under delay scenarios.

In Section 1.2, we will explain the basic definitions and structure of the airline crew pairing problem.

1.2 Introduction to Crew Pairing

The crew pairing problem focuses on assigning a sequence of flight legs to a crew. A flight leg is a single flight defined by its departure and arrival airports and scheduled departure and arrival times. Crew schedules are built in two steps: first, duties are formed as individual work shifts, and then several duties are combined to create a crew pairing.

A feasible crew pairing starts at one of the airline’s crew base airports and ends at a base, often covering a period of up to four days. Airlines typically use hub airports as crew bases, which allow crew members to rest without leading to additional accommodation or travel costs.

Crew payments are usually calculated based on both flight time and connection time between flights. Various rules and regulations determine how duties and pairings can be legally formed. These constraints ensure that crew schedules meet safety, operational, and labor requirements [Hall \[2012\]](#).

1.2.1 Network Structure

In the crew pairing problem, two types of networks are used: flight networks and duty networks. A flight network represents individual flights as nodes, with arcs indicating possible flight connections based on time and location feasibility. A duty network is constructed by

grouping sequences of flights (called duties) that can be operated by a crew. Each node in a duty network represents a feasible duty, and arcs connect duties that can be flown consecutively within a pairing. For domestic flights, the flight network is mainly used because the number of possibilities for duties is larger. On the other hand, duty networks are used for international flights. In both networks, dummy nodes are added to represent crew bases.

In flight networks, the nodes represent flight departure and arrival times and locations. The arcs between these nodes represent possible flight connections. Crew base nodes contain only location information and are used to define the starting points of pairings for each crew. For example in Figure 1.1, both the source and sink nodes are located at airport A, representing a crew base. These source and sink nodes are for the crew bases. On the other hand, regular flight leg nodes contain both time and location details. A connection between two regular flight leg nodes is added to the network if the connection is feasible in terms of time and location. For instance, two consecutive flights can be connected if the first flight arrives at the same airport from which the second flight departs and if there is enough connection time between them. We include a small flight schedule example from Lu [2014] to better illustrate this concept.

Table 1.1: Example Flight Schedule Lu [2014]

Flight	Departure Airport	Arrival Airport	Departure Time	Arrival Time
1	A	B	9:00	10:00
2	B	A	11:00	11:30
3	A	B	12:30	13:00
4	B	C	13:30	14:00
5	C	A	15:00	16:00

For this given flight schedule, as we can see from Figure 1.1, we can form the following flight network.

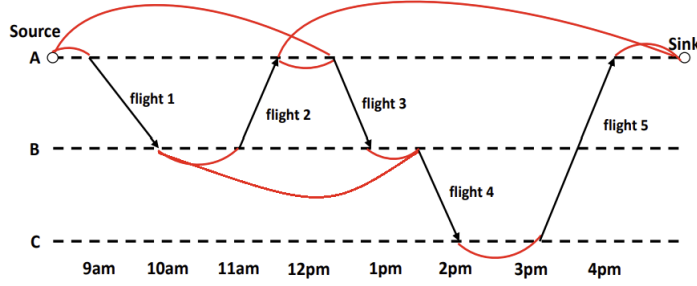


Figure 1.1: Example Flight Network [Lu \[2014\]](#)

Flight networks cannot incorporate complex network restrictions, such as the maximum allowable flight time or the minimum rest time between duties.

A set of ordered flight legs is grouped together to form duty-based networks to create legal duties. Similar to a flight-based network, there are two types of nodes: regular duty nodes and crew base nodes (source, sink). Regular duty nodes contain sets of flight legs. Two duty nodes can be connected if the first ends at the same airport where the second begins, and time constraints are satisfied. Crew base nodes represent the start and end of a pairing.

Table 1.2: Example Duties [Lu \[2014\]](#)

Flight/Duty	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
2	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
3	0	0	1	1	1	0	1	1	1	1	1	1	0	0	0
4	0	0	0	1	1	0	0	1	1	0	1	1	1	1	0
5	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1

As we can see from Table [1.2](#), these duties can be formed using the same flight schedule.

For example, Duty 3 has three flights 1, 2 and 3.

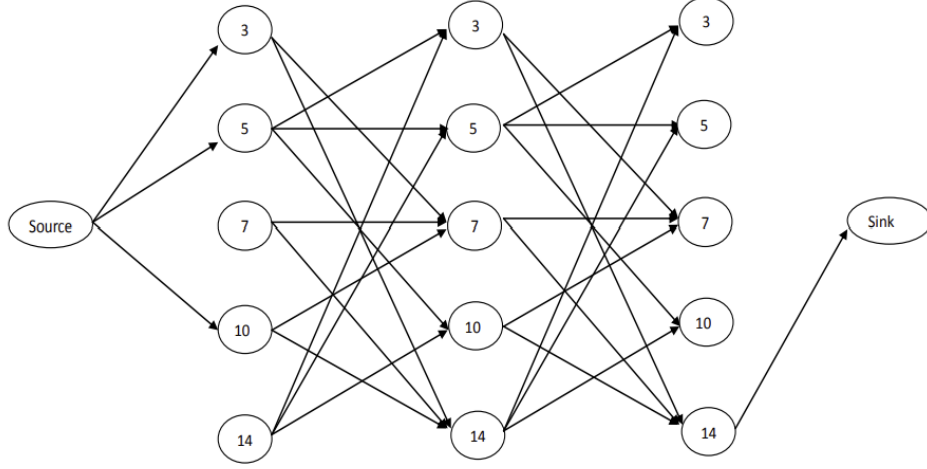


Figure 1.2: Example Duty Network [Lu \[2014\]](#)

Using this duty-based network, we can incorporate complex network restrictions, such as minimum sit time between flights, into our framework. Minimum sit time refers to the minimum required time between the arrival of one flight and the departure of the next, allowing the crew sufficient time to prepare for the next flight. Additionally, we can integrate the complex crew payment structure. For this reason, in this thesis, we primarily focus on the duty network. This allows us to distinguish between duty connections and flight connections within duties.

1.2.2 Formulation

Regulations define the minimum and maximum allowable sit time between connecting flights and limit the total duty duration. After completing a duty, crew members must take a minimum rest period before beginning a new one. The required rest time depends on the duration of the previous duty, and it is known as the 8-in-24 rule [Hall \[2012\]](#). This

limits flight time to a maximum of 8 hours within any 24-hour period. In addition, these regulations ensure a minimum guaranteed payment for each duty, regardless of the elapsed time. Incorporating these rules into the optimization model helps produce crew pairings that are both realistic and compliant with operational and legal requirements.

Crew members are paid according to the amount of flight time they complete. Their payment is based on two main factors: the total duty costs and a fraction of the pairing elapsed time. Some studies in the literature consider only the pairing elapsed time to estimate crew payments. However, in this study, both factors are incorporated to better capture the complexities of real-world crew scheduling. There are two types of costs: scheduled costs and operational costs. Scheduled costs are estimated based on scheduled flight times. On the other hand, operational costs account for the real-time variations, such as delays or unexpected changes in the schedule [Schaefer et al. \[2005b\]](#).

Set covering is well known in the literature for modeling problems where a collection of subsets (in this case, feasible pairings) must be selected to cover the entire set (all scheduled flights). Hence, the crew pairing problem is formulated as a set partitioning problem:

$$[\mathbf{P}] : \quad \min \sum_{p \in \mathcal{P}} c_p x_p \quad (1.1)$$

$$\text{s.t.} \quad \sum_{p \in \mathcal{P}} a_{ip} x_p = 1 \quad \forall i \in \mathcal{F} \quad (1.2)$$

$$x_p \in \{0, 1\} \quad \forall p \in \mathcal{P} \quad (1.3)$$

The set $\mathcal{F}$ includes all flight legs that need to be covered by flight crews. Each pairing $p \in \mathcal{P}$ is a sequence of flights or duties that a crew can operate. The parameter a_{ip} is 1 if flight i is included in pairing p , and 0 otherwise. The cost c_p usually depends on total

duration, labor cost, or other flight-related factors. The decision variable x_p equals 1 if pairing p is selected and 0 otherwise. The objective function 1.1 minimizes the total crew costs. Constraint 1.2 ensures that each flight leg is covered exactly once. Lastly, constraint 1.3 is the domain constraint.

In practice, it is not possible to enumerate all pairings in advance. The number of possible pairings can be very large. Hence, methods like column generation or Lagrangian relaxation are used to deal with large-scale problem sizes.

1.3 Contributions and Thesis Content

The crew pairing problem under uncertainty has received significant attention in the literature. Some studies adopt robust optimization methods Antunes et al. [2019], Lu and Gzara [2015], aiming to generate solutions that remain effective under worst-case scenarios. Others rely on two-stage stochastic programming Yen and Birge [2006b], where uncertainty is modeled through probability distributions. There are also deterministic approaches Schaefer et al. [2005b], Biswal [2024] that incorporate expected delay values into the planning process.

Unlike many studies that use a simplified cost structure Yen and Birge [2006b], Sun et al. [2020], this thesis applies a non-linear pay-and-credit scheme that includes both flight time and flight connection times in duty costs. A duty network is used to reflect operational constraints. In addition, a delay penalty term is added to the cost function to account for the risk of disruptions. The delay penalty depends on the buffer times within duties. Buffer time is calculated based on the minimum connection time and the arrival and departure times of consecutive flights. The penalty is then computed using historical delay data

and buffer time. This approach results in crew pairings that are more robust to schedule uncertainties.

This thesis is organized as follows. Chapter 2 provides a comprehensive review of the literature on crew pairing problems, particularly focusing on those involving uncertainty. Chapter 3 defines the mathematical models employed in this study. The chapter also introduces the delay averse model, which is designed to account for uncertainties in scheduling. Chapter 4 explains the solution methodology, including the Lagrangian relaxation and cutting plane approach. Chapter 5 presents the numerical results obtained from the proposed models. It includes a sensitivity analysis, as well as simulation experiments to evaluate solutions performance under different scenarios. Finally, Chapter 6 concludes the thesis and discusses future research directions.

Chapter 2

Literature Review

Crew scheduling problems are generally divided into two categories: crew pairing and crew rostering. Crew pairing is generating pairings without considering individual crew needs. Crew rostering assigns specific crews to flights while addressing individual requirements. Studies in crew scheduling consider simple and complex pay structures. Simple methods calculate crew payments based on flight time. More realistic approaches, like the pay-and-credit (PAC) method, consider factors such as flight time, elapsed time, rest time, sit times, and a minimum guaranteed payment, using the maximum value among them.

Table 2.1: Selected Studies on Airline Crew Planning and Recovery Methods

Paper	Problem	Pay Structure	Uncertainty	Flight Data	Recovery Method	Model Type
Schaefer et al. [2005b]	Crew Pairing	PAC	Elapsed Time, Departure Time	US	Propagated Delay	Deterministic
Lu and Gzara [2015]	Crew Pairing	Flight Time	Elapsed Time	Europe	Propagated Delay	Robust
Antunes et al. [2019]	Crew Pairing	PAC	Elapsed Time, Departure Time	US	Propagated Delay	Robust
Sun et al. [2020]	Crew Pairing	Flight Time	Departure Time, Arrival Time	US	Propagated Delay	Bi-objective
Yen and Birge [2006b]	Crew Pairing	Flight Time	Flight Time	New Zealand	Propagated Delay	SPP
Ehrgott and Ryan [2002b]	Crew Pairing	Flight Time	Flight Time	New Zealand	Propagated Delay	Bi-objective
Tam et al. [2011]	Crew Pairing	Flight Time	Flight Time	New Zealand	Propagated Delay	Bi-objective, SPP
Shebalov and Klabjan [2006b]	Crew Swapping	Flight Time	Flight Time	US	Crew Swap	Robust
Medard and Sawhney [2007]	Crew Recovery	Flight Time	Flight Time	US	Crew Swap	Deterministic
Guo [2005]	Crew Recovery	Fixed	Flight Time	Europe	Crew Swap	Deterministic
Gao et al. [2009]	Integrated Planning	PAC	Flight Time	US	Propagated Delay	Deterministic
Dunbar et al. [2012]	Integrated Planning	Flight Time	Flight Time	US	Propagated Delay	Deterministic
Weide et al. [2010b]	Integrated Planning	Flight Time	Flight Time	US	Propagated Delay	Deterministic

Moreover, crew cost is calculated based on operational cost or planned cost. Planned costs are based on scheduled departure and arrival times, while operational costs refer to the actual costs during flight operations. These costs are an important aspect of crew pairing under uncertainty, reflecting the real-world effects of operational disruptions.

[Kohl et al. \[2007b\]](#) provides an early review of airline disruption management. They identify two main strategies for handling operational uncertainties: reactive and proactive solutions. The reactive approach involves the airline operations center responding to disruptions as they occur. Examples of decisions in this approach include adjusting cruise speed, utilizing stand-by crews, performing crew swaps, or adding slack time to schedules. In contrast, the proactive approach focuses on anticipating uncertainties, using methods like robust optimization and stochastic programming to form crew schedules.

[Sun et al. \[2020\]](#) present a bicriteria optimization model and define a penalty function as a measure of robustness. They combine expected departure and arrival delays with overtime costs for crew members similar to [Antunes et al. \[2019\]](#). While [Antunes et al. \[2019\]](#) assumes that root delays and propagated delays are uncorrelated, [Sun et al. \[2020\]](#) uses a regression model to account for the correlation between these delays. This improves their delay estimations. The crew payment structure in their model is linear, with four components: deadheading (flight time of crews as passengers), connection time, unit penalty for short duty, and unit rest period. They include crew costs in their first objective. The second objective function is used to minimize the deviations between the actual flight schedule and the planned one. It consists of four components: the expected departure and arrival delays for each leg, as well as the overtime for each duty and pairing. This objective serves as a measure of robustness. They use the weighted sum method to transform their bi-objective model into a single-objective problem. They then solve the model using column generation. For their experiments, they use 2 years of flight data from a leading airline based in Hong

Kong.

To minimize the effect of delays, slack time between flights is commonly used in the literature. For instance, [Ahmadbeygi et al. \[2010\]](#) proposes a model to redistribute slack more evenly across the network. The decision variables involve adjusting the departure times of flights to balance slack distribution. Although they do not consider crew payments, their objective function minimizes the total expected propagated delay. They propose two different models: single-layer and multi-layer. The single-layer model considers interactions between two flights, while the multi-layer model accounts for the entire sequence of flights. Their case study uses data from a major US carrier.

[Schaefer et al. \[2005b\]](#) addresses the crew pairing problem under flight time uncertainties by estimating the expected cost of each pairing using a discrete event simulation. They apply a pay-and-credit scheme to calculate crew payments and make two key assumptions: planes are always available, and ground delays or push-backs are used for pairing recovery. They introduce a penalty function to improve pairing performance by considering operational factors such as short sit times, rest times, long flying times, and elapsed times, which are incorporated into the penalty function. Flights are grouped into short, medium, and long categories, and flight times are estimated by fitting a distribution for the fleet. Column generation is applied to solve the crew pairing problem by first enumerating all pairings and calculating their expected costs. The objective function is modified with a penalty for deviations from set thresholds for sit times, rest times, flying time, and elapsed time, improving the pairing cost. Due to the problem’s non-convex nature, they use a local search approach to find approximate solutions, starting with a deterministic solution and adjusting penalty parameters iteratively. The solutions obtained using this penalty method outperform the deterministic model with a smaller operational cost gap. The authors test their assumptions about plane availability and recovery methods, finding that

the availability assumption has little effect on expected costs, while the recovery method oversimplifies real-life options, especially at hub airports.

Yen and Birge [2006b] formulate a two-stage stochastic integer programming model for crew pairing. In the first stage, they construct crew pairings, and in the second stage, flight departure times are recourse decision variables. The objective is to minimize total costs, including crew salaries and expected delay costs. Their model allows only flight departure delays as a recovery action and does not account for violations of crew duty rules. Additionally, it uses a simplified crew pay structure compared to the more complex systems in North American airlines.

Ehrgott and Ryan [2002b] also proposes a bi-objective model to obtain robust yet cost-effective crew pairings. They use the epsilon constraint method to solve their bi-objective model and obtain the Pareto frontier. The first objective in their model is pairing cost, which is mainly based on flight time and is linear. The second objective is related to robustness and it penalizes short connection times. However, their approach to robustness is relatively simple and overlooks other critical operational challenges that could lead to delays.

Tam et al. [2011] compares two approaches for robust crew pairing. Building on Yen and Birge [2006b] they propose a two-stage stochastic programming model. Following Ehrgott and Ryan [2002b] they develop a bi-objective model. They develop a regression model to estimate delays, which is then used to generate scenarios for their two-stage stochastic programming model. Similarly, the regression model is used to estimate parameters for the bi-objective model. They show that the bi-objective approach works well for moderate delay costs, but their stochastic programming model is better for scenarios with higher delay costs. They use Air New Zealand data for their case study.

Shebalov and Klabjan [2006b] introduce a model based on the practical crew recovery strategy of crew swapping. Their two-stage approach first minimizes crew pairing costs, then finds a robust solution. The cost of the robust solution is within 1% of the optimal cost of a deterministic solution. Unlike studies using flight time to calculate crew costs, they use monetary values for calculating the crew payments in their computation analysis. They use a combination of Lagrangian relaxation and the column generation method to solve their model.

Gao et al. [2009] proposes an integrated approach to fleet assignment and crew scheduling that aims to enhance operational robustness by incorporating fleet and crew base purity. In this context, purity refers to limiting the number of fleet types or crew bases serving a particular station. Fleet purity ensures that no more than a specified number of fleet types operate at a given station, reducing aircraft dispersion and improving maintenance planning, scheduling efficiency, and operational robustness. Crew base purity restricts the number of crew bases serving a station, helping crews stay closer to their home bases, which reduces deadheading, minimizes locked rotations, and improves recovery from disruptions. Their study shows that implementing crew base purity leads to better crew scheduling, fewer operational disruptions, and cost savings, all while simplifying the overall planning process and enhancing computational efficiency.

Lu and Gzara [2015] develop a robust optimization model for the crew pairing problem. They use a simplified crew payment structure which is based on the time away from base. Their model is nonlinear due to the robustness-related components and is solved using Lagrangian relaxation. Using European airline flight data, they conduct a case study, incorporating robustness into both the objective function and the total allowable time away from base constraint. This study is the first attempt to introduce robustness in airline crew pairing problem. This study is notable as the first to introduce robust optimization to the

airline crew pairing problem.

[Antunes et al. \[2019\]](#) also formulate a robust optimization model for the crew pairing problem while incorporating the pay-and-credit method to calculate crew salaries. They consider propagated delays to deal with flight disruptions. These propagated delays occur when one flight’s delay causes subsequent delays for the same crew. They consider the additional operation costs of crews due to propagated delays. They obtain robust solutions. The model is reformulated as a standard set partitioning problem and solved using column generation. They validate the quality of their solutions through simulation, using Virgin America (VX) 2014 flight data in a case study.

The literature on crew pairing under uncertainty can be grouped by modeling approach, payment method, and recovery strategy. Three main modeling approaches used in crew scheduling are chance-constrained models, stochastic programming, and robust optimization. In terms of crew payments, some studies use fixed payments, while others apply a pay-and-credit scheme or calculate costs based on pairing elapsed time.

Our approach differs by adding a flight delay risk function to the objective, penalizing duties that have insufficient buffer time to absorb delays. We work on a duty network, treating flights within a duty as a sequence and handling duty connections differently. We also perform a simulation analysis to calculate total propagated delays and evaluate the performance of our solutions.

Chapter 3

Modelling Crew Pairing Optimization

In this chapter, we define the crew pairing problem under two cases: deterministic departure and arrival times and flight delays. In both cases, we adapt the pay-and-credit cost function on the duty network. Section 3.1 provides the arc-based formulation for the case of pay-and-credit crew pairing cost and deterministic arrival and departure times. Section 3.2 provides the equivalent flow-based formulation. Section 3.3 introduces the delay averse crew pairing optimization formulation where flights may experience delays due to uncertainty in departure and arrival times.

3.1 Arc-Based Formulation

Let $G = (N^0, A^0)$ represent the duty network. K is the set of all available crews, and $\mathcal{F}$ is the set of all flight legs that need to be covered. The set of nodes N^0 , has three types

of nodes: origin node, destination node, and duty nodes. Crew $k \in K$ has a designated origin node $O \in N^0$ and a destination node $D \in N^0$. The remaining nodes, denoted by $i \in N \subset N^0$, correspond to duties that crews may be assigned to. Using the duty network, we may incorporate duty restrictions, including the maximum allowable duty time, minimum rest times, and minimum and maximum sit times. A duty i can only connect to duty j if these restrictions are satisfied; otherwise, the arc linking them will not be created in the network. Therefore, the set A^0 only contains arcs between compatible duties that satisfy all such temporal and regulatory constraints. Similarly, the arcs $(i, j) \in A^0$ fall into three categories: from a crew's origin node O to a duty node j , from a duty node i to their destination node D , and between two duty nodes i and j . We define $A \subset A^0$ as the subset of arcs that connect only pairs of duty nodes in N , i.e., $(i, j) \in A$ if and only if $i \in N$, $j \in N$.

The parameter Δ_{fi} represents the relationship between flight f and duty i ; it takes a value of 1 if duty i can cover flight f and 0 otherwise. Rest time t_{ij} is described as the difference between the scheduled arrival time (SAT) of the last flight of duty i and the scheduled departure time (SDT) of the first flight duty j . The parameters $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ are weight coefficients used in the nonlinear cost structure of crew pairings. Specifically, α is applied to the total duty elapsed times, while β is applied to the total pairing elapsed time. The parameter θ represents the minimum guaranteed time for a duty.

Moreover the cost of duty i (b_i) is calculated based on three factors: total flight time within the duty, duty elapsed time, and minimum guaranteed time per duty. Duty elapsed time (e_i) is the total time spent on duty, including both the flight and sit times between flights. It is calculated as the difference between SAT of the last flight and SDT of the first flight within the duty. Total flight time (m_i) is the cumulative time spent flying during the duty. It is the summation of scheduled elapsed times (SET) of all flights in the duty.

Lastly, crews are guaranteed a minimum payment (θ). The duty cost is then calculated as a maximum of the following three components:

$$b_i = \max\{m_i, \alpha \times e_i, \theta\}, \quad i \in N \quad (3.1)$$

The crew pairing cost is computed for crew $k \in K$ as the maximum of two components. The first component is for the total duty costs covered by crew k . The second component incorporates the total costs associated with elapsed time e_i for duty i and rest time t_{ij} between duties. Overall, the second component equals the fraction of pairing elapsed time for each crew. At the beginning of the planning horizon, the cost of traveling from the base airport to any other airport is assumed to be zero. Similarly, at the end of the planning horizon, returning to the base airport does not create any cost. However, the regular cost applies if the crew visits the base airport during the pairing. We define decision variables $x_{ij}^k, \forall (i, j) \in A^0, \forall k \in K$ to indicate that crew k is assigned to duty j after duty i . Before solving the optimization model, we calculate duty costs (b_i) using the cumulative flight time (m_i), duty elapsed time (e_i) and the minimum guaranteed flight time θ as these three parameters are given. We also obtain the connection times between consecutive duties t_{ij} using the flight data. Then, we calculate the crew pairing cost for crew k as follows:

$$\max \left\{ \sum_{(i,j) \in A} b_i x_{ij}^k, \sum_{(i,j) \in A} \beta(e_i + t_{ij}) x_{ij}^k \right\}, \quad k \in K \quad (3.2)$$

Table 3.1 provides a description of the sets, parameters, and decision variables used in the arc-based formulation.

Table 3.1: Description of Sets, Parameters, and Decision Variables

Sets	
Symbol	Description
N	Set of duty nodes
A	Set of duty connection arcs
K	Set of crews indexed by k
$\mathcal{F}$	Set of flights indexed by f
Parameters	
Symbol	Description
b_i	Cost of duty $i \in N$
e_i	Elapsed time of duty $i \in N$
t_{ij}	Rest time between duty $i \in N$ and duty $j \in N$
Δ_{fi}	Binary parameter equal to 1 if duty $i \in N$ covers flight $f \in \mathcal{F}$
m_i	Total flight time in duty, $i \in N$
α, β, θ	Cost parameters
Decision Variables	
Symbol	Description
x_{ij}^k	Binary decision variable equal to 1 if duty $j \in N$ is covered after duty $i \in N$ by crew $k \in K$, 0 otherwise
z^k	Decision variable for calculating the nonlinear pairing cost for crew $k \in K$

The crew pairing problem is finding the minimum-cost set of pairings that covers all flights at least once. The objective is to minimize the crew cost while satisfying connection constraints. The following optimization model is the arc-based model:

$$\min \sum_{k \in K} \max \left\{ \sum_{(i,j) \in A} b_i x_{ij}^k, \sum_{(i,j) \in A} \beta(e_i + t_{ij}) x_{ij}^k \right\} \quad (3.3)$$

$$\text{s.t.} \quad \sum_{k \in K} \sum_{(i,j) \in A^0} \Delta_{fi} x_{ij}^k \geq 1 \quad \forall f \in \mathcal{F} \quad (3.4)$$

$$\sum_{j: (O,j) \in A^0} x_{O,j}^k = 1 \quad \forall k \in K \quad (3.5)$$

$$\sum_{i: (i,D) \in A^0} x_{i,D}^k = 1 \quad \forall k \in K \quad (3.6)$$

$$\sum_{j: (i,j) \in A^0} x_{ij}^k - \sum_{j: (j,i) \in A^0} x_{ji}^k = 0 \quad i \in N, \quad \forall k \in K \quad (3.7)$$

$$x_{ij}^k \in \{0, 1\} \quad \forall k \in K, \forall (i, j) \in A^0 \quad (3.8)$$

The objective function 3.3 is min-max because it minimizes the maximum of two terms: the total duty cost and part of the total elapsed time of each pairing. The objective is to minimize total crew cost. The objective function is convex. By summing over all duties and crews, constraint (3.4) guarantees that each flight is covered at least once. Constraints (3.5)-(3.7) are flow balance constraints and are typical in multicommodity network flow problems. They ensure that the crew pairings are connected to each other and that the pairings start and end at a crew base. Constraint (3.8) is the domain constraint.

We define an auxiliary variable z^k for crew $k \in K$, which will be utilized to linearize the objective function for crew pairing cost calculations.

$$\min \sum_{k \in K} z^k \quad (3.9)$$

$$\text{s.t. } (3.4), (3.5), (3.6), (3.7) \quad (3.10)$$

$$z^k \geq \sum_{(i,j) \in A} b_i x_{ij}^k \quad \forall k \in K \quad (3.11)$$

$$z^k \geq \beta \sum_{(i,j) \in A} (e_i + t_{ij}) \quad \forall k \in K \quad (3.12)$$

$$z^k \geq 0 \quad \forall k \in K \quad (3.13)$$

Constraints (3.11) and (3.12) are for non-linear crew pairing cost calculations where the cost of a pairing is the maximum of total duty cost or β times pairing elapsed time.

3.2 Path Based Formulation

In the path-based formulation, we introduce two additional sets. Let $\mathcal{P}$ denote the set of all feasible pairings, where pairing $p \in \mathcal{P}$ represents a sequence of duties that a crew may be assigned to. For pairing p , we define $\mathcal{D}_p$ as the subset of duties that are included in pairing p . That is, $\mathcal{D}_p \subseteq N$ contains only the duty nodes covered by pairing p .

As discussed in Section 3.1, the pairing cost is calculated as the maximum of total duty costs and pairing elapsed time. For pairing p , the pairing elapsed time T_p is the total time from the start of the first duty to the end of the last duty in the pairing.

$$c_p = \max \left\{ \sum_{i \in N_p} b_i, \beta \times T_p \right\}, \quad p \in P \quad (3.14)$$

For each pairing $p \in \mathcal{P}$, we define a binary decision variable x_p , which takes the value 1 if pairing p is selected and 0 otherwise. Let a_{fp} be a binary parameter equal to 1 if flight $f \in \mathcal{F}$ is covered by pairing $p \in \mathcal{P}$, and 0 otherwise.

Table 3.2: Description of Sets, Parameters, and Decision Variables

Sets	
Symbol	Description
P	Set of pairings, indexed by p
$\mathcal{D}_p$	Set of duties in pairing $p \in P$, indexed by i
Parameters	
Symbol	Description
c_p	Cost of using pairing $p \in \mathcal{P}$
a_{fp}	Binary parameter equal to 1 if flight $f \in \mathcal{F}$ is in pairing $p \in \mathcal{P}$
T_p	Pairing elapsed time, $p \in \mathcal{P}$
Decision Variables	
Symbol	Description
x_p	Binary decision variable equal to 1 if pairing $p \in P$ is selected, 0 otherwise
η_p	Cost of pairing $p \in \mathcal{P}$
ψ_{ip}	Cost of duty $i \in \mathcal{D}$ within pairing $p \in \mathcal{P}$

Table 3.2 provides a description of the sets, parameters, and decision variables used in the path-based formulation. Then, the path-based formulation is given by:

$$\min \sum_{p \in P} \max \left\{ \sum_{i \in \mathcal{D}} b_i, \alpha \times T_p \right\} x_p \quad (3.15)$$

$$\text{s.t.} \quad \sum_{p \in P} a_{fp} x_p \geq 1 \quad \forall f \in F \quad (3.16)$$

$$x_p \in \{0, 1\} \quad \forall p \in P \quad (3.17)$$

Each flight must be covered at least once, and constraint (3.16) ensures that all flights in the network are assigned to a crew. Constraint (3.17) is the domain constraint.

To linearize the objective function, we introduce auxiliary variables for the inner maximum expression. Let η_p represent the total cost for each pairing. Since the duty cost is a maximum of three components, we define another auxiliary variable ψ_{ip} for the duty cost within a pairing p . This will lead to the following mathematical model with flight coverage constraints.

$$\min \quad \sum_{p \in P} \eta_p \quad (3.18)$$

$$\text{s.t.} \quad (3.16), (3.17) \quad (3.19)$$

$$\eta_p \geq \beta T_p x_p \quad \forall p \in P \quad (3.20)$$

$$\eta_p \geq \sum_{i \in \mathcal{D}_p} \psi_{ip} \quad \forall p \in P \quad (3.21)$$

$$\psi_{ip} \geq m_i x_p \quad \forall p \in P, \quad \forall i \in \mathcal{D} \quad (3.22)$$

$$\psi_{ip} \geq \alpha e_i x_p \quad \forall p \in P, \quad \forall i \in \mathcal{D} \quad (3.23)$$

$$\psi_{ip} \geq \theta x_p \quad \forall p \in P, \quad \forall i \in \mathcal{D} \quad (3.24)$$

$$\eta_p \geq 0 \quad \forall p \in P \quad (3.25)$$

$$\psi_{ip} \geq 0 \quad \forall p \in P, \quad \forall i \in \mathcal{D} \quad (3.26)$$

Constraints (3.20) - (3.24) capture the min-max cost calculation. The pairing cost η_p is equal to the maximum of the right-hand side of (3.20) - (3.21). Similarly, the duty costs ψ_{ip} are calculated by constraints (3.22) - (3.24) and equals the maximum of the flight time, a fraction of the duty's elapsed time, and the guaranteed minimum pay. Lastly, there are domain constraints (3.26).

3.3 Delay Averse Crew Pairing Optimization

In Section 3.1 and 3.2, the flight arrival and departure times are assumed to be known and fixed. This deterministic case is referred to as the nominal case. In reality, there is significant uncertainty in flight operations. The actual arrival, departure, and flight times are frequently different from the scheduled ones, resulting in delays. Flight delays not only cause disruptions to flight schedules and passenger inconvenience but also impact crew costs, a significant portion of total airline costs. Crews are paid based on actual duty and pairing costs and not on scheduled ones. Crew pairings that are prone to disruption due to delays may lead to a significant increase in actual pairing cost. In this section, we incorporate delay by adding a delay penalty function to the objective function.

We define the set of flights $\mathcal{F}_i$ in duty i as an ordered set: $\mathcal{F}_i = \{1, 2, \dots, l\}$, $i \in N$, and the buffer time $(\rho_{f,f+1})$ for consecutive flights f and $f + 1$ is calculated as:

$$\rho_{f,f+1} = SDT_{f+1} - SAT_f - \tau, \quad (3.27)$$

where τ is the minimum sit time. We define φ_i as the excess delay when the delay exceeds the buffer time. We calculate a cumulative duty delay cost component φ_i as follows:

$$\varphi_i = \sum_{f \in \mathcal{F}_i} (\max \{0, (d_f - \rho_{f,f+1})\})^2, \quad i \in N \quad (3.28)$$

where d_f is the delay estimate of flight f and $\rho_{f,f+1}$ is the buffer time between consecutive flights in a duty. For example, if the first flight arrives at 12:00 PM and the following flight departs at 1:30 PM, with a minimum sit time of 60 minutes, the total connection time is 90 minutes, and the buffer time is 30 minutes. Assume that the first flight has a delay estimate of 40 minutes. In this case, the delay penalty for this duty is $\varphi_i = (\max \{0, 40 - 30\})^2 = 100$.

We introduce the delay averse crew pairing optimization model below.

$$\min \quad \sum_{k \in K} z^k + \gamma \sum_{k \in K} \sum_{(i,j) \in A} \varphi_i x_{ij}^k \quad (3.29)$$

$$\text{s.t} \quad (3.4) - (3.13) \quad (3.30)$$

The objective function (3.29) minimizes the total cost, including the scheduled crew pairing and the delay penalty. Lastly, γ is a weighting factor to capture the trade-off between scheduled cost and delay penalty cost. A higher value of γ penalizes shorter connections between flights with higher delays. With the new objective function, we minimize the scheduled crew costs and the delay related risks.

In Chapter 4, we present the solution methodology, in which we apply Lagrangian relaxation to decompose the problem into subproblems and use a cutting-plane algorithm to solve the Lagrangian master problem.

Chapter 4

Solution Methodology

Column generation is a well-established methodology for crew pairing optimization. We adopt this methodology to solve the delay averse crew pairing model of Section 3.3. In Section 4.1, we apply Lagrangian relaxation and derive the Lagrangian dual problem and the subproblem. In Section 4.2, we detail the solution of Lagrangian dual problem using cutting plane method.

4.1 Lagrangian Relaxation

We relax the flight coverage constraint 3.4 and apply Lagrangian relaxation using variables $\lambda_f \geq 0 \ f \in \mathcal{F}$,. This constraint is complicated as it couples all crews $k \in K$ by requiring that each flight be covered. By relaxing it, we can decompose the integer programming model. We obtain the following Lagrangian problem:

$$\begin{aligned}
[LR] : \quad & \min \quad \sum_{k \in K} z^k + \sum_{f \in F} \lambda_f \left(1 - \sum_{k \in K} \sum_{(i,j) \in A^0} \Delta_{fi} x_{ij}^k \right) \\
& \text{s.t.} \quad (3.5) \text{ to } (3.13)
\end{aligned} \tag{4.1}$$

The Lagrangian problem (LR) decomposes into $|K|$ subproblems, for each crew $k \in K$.

$$[SP_k] : \min \quad z^k - \sum_{(i,j) \in A} \sum_{f \in \mathcal{F}} \lambda_f \Delta_{fi} x_{ij}^k \tag{4.2}$$

$$\text{s.t.} \quad \sum_{j: (o,j) \in A^0} x_{o,j}^k = 1 \tag{4.3}$$

$$\sum_{i: (i,d) \in A^0} x_{i,d}^k = 1 \tag{4.4}$$

$$\sum_{j: (i,j) \in A^0} x_{ij}^k - \sum_{j: (j,i) \in A^0} x_{ji}^k = 0 \quad \forall i \in N \tag{4.5}$$

$$z^k \geq \sum_{(i,j) \in A} b_i x_{ij}^k \tag{4.6}$$

$$z^k \geq \beta \sum_{(i,j) \in A} (e_i + t_{ij}) x_{ij}^k \tag{4.7}$$

$$x_{ij}^k \in \{0, 1\} \quad \forall (i, j) \in A \tag{4.8}$$

$$z^k \geq 0 \tag{4.9}$$

Subproblem $SP_k, k \in K$ is a shortest path problem with a min-max cost function, and we discuss its solution in Section ??.

Furthermore (LR) provides a lower bound on the optimal solution of the delay averse crew pairing problem (DA-CPP). The best such lower bound is given by the Lagrangian dual problem:

$$[LD] : \max_{\lambda_f \geq 0} \sum_{f \in \mathcal{F}} \lambda_f + \sum_{k \in K} \left\{ \min_{z, x} z^k - \sum_{(i,j) \in A} \sum_{f \in F} \lambda_f \Delta_{fi} x_{ij}^k : (4.3) - (4.9) \right\} \quad (2.2)$$

In section 4.2, we discuss the calculation of the Lagrangian bound.

4.2 Cutting Plane Algorithm

Subproblems SP_k , $k \in K$ are identical since all crew $k \in K$ have the same origin, destination, and payment scheme. Let P denote the set of feasible pairings for subproblem k . For each pairing $p \in P$, define x_{ij}^p as a binary parameter that takes value 1 if the duty connection (i, j) is in pairing $p \in P$, and 0 otherwise.

Then, Lagrangian master problem can be formulated as follows:

$$[MP] : \max \sum_{f \in F} \lambda_f + |K|\mu \quad (4.10)$$

$$\text{s.t. } \mu \leq z^p - \sum_{(i,j) \in A^0} \sum_{f \in F} \lambda_f \Delta_{fi} x_{ij}^p \quad \forall p \in P \quad (4.11)$$

$$\lambda_f \geq 0 \quad \forall f \in \mathcal{F} \quad (4.12)$$

The Lagrangian master problem is solved iteratively using a cutting plane algorithm. The solution path found from the subproblems are added as cuts in constraint (4.11). Recall that Δ_{fi} is a binary parameter and is 1 if flight $f \in \mathcal{F}$ is covered by duty $i, \in N$. Then, we solve the Lagrangian master problem to obtain new λ values and use them to resolve the subproblems. The upper bound (UB_t) at each iteration t is calculated as follows:

$$UB_t = \sum_{f \in F} \bar{\lambda}_f + |K| \bar{\mu} \quad (4.13)$$

where $\bar{\lambda}_f$ and $\bar{\mu}$ are the optimal solution of Lagrangian master problem at iteration t . Lower bound (LB_t) at each iteration t is calculated as follows:

$$LB_t = \max \left\{ LB_{t-1}, \sum_{f \in F} \bar{\lambda}_f + \sum_{k \in K} \bar{z}^k - \sum_{k \in K} \sum_{(i,j) \in A^0} \sum_{f \in F} \bar{x}_{ij}^k \bar{\lambda}_f \Delta_{fi} \right\} \quad (4.14)$$

where LB_{t-1} is the lower bound obtained in the previous iteration. $\bar{z}^k$ and $\bar{x}_{ij}^k$ are the optimal solution of subproblem at iteration t . This cutting plane algorithm continues until the following convergence criteria is satisfied.

$$\frac{|UB_t - LB_t|}{LB_t} < \epsilon, \quad (4.15)$$

where ϵ is equal to 10^{-4} . After the cutting plane algorithm converges, we solve the path-based formulation to obtain an integer solution using the generated set of paths.

$$\min \quad \sum_{p \in \bar{P}} \eta_p \quad (4.16)$$

$$\text{s.t.} \quad \sum_{p \in \bar{P}} a_{fp} x_p \geq 1 \quad \forall f \in \mathcal{F} \quad (4.17)$$

$$\eta_p \geq \beta T_p x_p \quad \forall p \in \bar{P} \quad (4.18)$$

$$\eta_p \geq \sum_{i \in \mathcal{D}_p} \psi_{ip} \quad \forall p \in \bar{P} \quad (4.19)$$

$$\psi_{ip} \geq m_{ip} x_p \quad \forall p \in \bar{P}, \quad \forall i \in \mathcal{D} \quad (4.20)$$

$$\psi_{ip} \geq \alpha e_{ip} x_p \quad \forall p \in \bar{P}, \quad \forall i \in \mathcal{D} \quad (4.21)$$

$$\psi_{ip} \geq \theta x_p \quad \forall p \in \bar{P}, \quad \forall i \in \mathcal{D} \quad (4.22)$$

$$x_p \in \{0, 1\} \quad \forall p \in \bar{P} \quad (4.23)$$

$$\eta_p \geq 0 \quad \forall p \in \bar{P} \quad (4.24)$$

$$\psi_{ip} \geq 0 \quad \forall p \in \bar{P}, \quad \forall i \in \mathcal{D} \quad (4.25)$$

The above formulation represents the restricted path formulation, which considers a limited subset of feasible pairings $\bar{P}$. We first check the original problem's feasibility, then solve the restricted path formulation, as feasibility is not guaranteed.

Algorithm 1 Cutting Plane Algorithm

- 1: **Input:** Lagrangian multipliers λ , epsilon ϵ , weighting factor for delay penalty γ
 - 2: **Output:** Upper bound UB , lower bound LB , paths H
 - 3: **while** $\frac{UB-LB}{LB} > \epsilon$ **do**
 - 4: Compute reduced costs using $\bar{\lambda}$
 - 5: Solve Subproblem
 - 6: Obtain z^p, x^p
 - 7: Update **Master Problem**:
 - 8: Add cuts $\mu \leq z^p - \sum_{(i,j) \in A} \sum_{f \in F} \lambda_f \Delta_{fi} x_{ij}^p \quad \forall p \in P$
 - 9: Solve **Master Problem** with updated cuts and obtain $\bar{\lambda}_f, \bar{\mu}$
 - 10: Calculate $UB = \sum_{f \in F} \bar{\lambda}_f + |K| \bar{\mu}$
 - 11: Calculate $LB = \max \left\{ LB, \sum_{f \in F} \bar{\lambda}_f + \sum_{k \in K} \bar{z}^k - \sum_{k \in K} \sum_{(i,j) \in A^0} \sum_{f \in F} \bar{x}_{ij}^k \bar{\lambda}_f \Delta_{fi} \right\}$
 - 12: **end while**
 - 13: **Return** UB, LB, H
-

In Chapter 5, we apply the solution methodology to a real-world instance from Endeavor Airlines for the year 2019. We conduct simulations to further evaluate the solution.

Chapter 5

Case Study

In this chapter, we analyze the performance of the delay averse crew pairing model. We use data from the Bureau of Transportation Statistics (BTS) form flight schedules and the corresponding flight and duty network. We then apply the cutting plane algorithm to obtain bounds on the Lagrangian dual. Using the paths generated in this algorithm, we solve the restricted path-based formulation to find the optimal crew pairings. We perform extensive testing under multiple parameter settings and conduct sensitivity analysis to examine variations in the objective for the nominal and delay averse cases.

To evaluate the effectiveness of the model, we perform simulations that incorporate delay realizations and track propagated delays. The results indicate that our delay averse model significantly reduces delay propagation, leading to more stable schedules.

5.1 Experimental Setup

For our analysis, we work with a 60-crew instance for Endeavor Airlines, covering the period from December 2 to 5, 2019. This dataset is retrieved from BTS and includes detailed flight information, such as flight numbers, scheduled and actual departure/arrival times, delays, and cancellation status. We focus on the most frequent flights to ensure a representative sample of the airline’s operations. The instance consists of 243 flights across 25 airports, with John F. Kennedy International Airport (JFK) serving as the crew base.

Based on the flight schedule, we construct the flight network, ensuring that connections respect operational constraints such as minimum sit times. From this network, we generate all possible duties with respect to constraints on the maximum duty elapsed time and the minimum duty connection time. After generating duties, we apply a depth-first search (DFS) algorithm to identify and remove disconnected duties from the network. This pre-processing step significantly reduces the network size, improving computational efficiency in later optimization steps. In the end, the instance includes 479 duty nodes, with the crew base serving as the origin and destination. Duty costs are computed based on flight times and sit times between connected flights. We use the parameter values in Table 5.1 to determine the crew pairing costs. All computations are performed on a 13th Gen Intel® Core™ i9-13900H 2.60 GHz processor with 16GB RAM.

Table 5.1: Crew pairing model parameters

Parameter	Description	Value
α	Weight for pairing elapsed time	0.9
β	Weight for duty elapsed time	0.8
θ	Minimum duty length (minutes)	200
τ	Minimum sit time between consecutive flights (minutes)	60

5.2 Computational Testing

We apply Lagrangian relaxation and compute the Lagrangian bound using a cutting plane algorithm. We use different delay estimates d_f in the numerical analysis to account for varying levels of delay. We test three different values for the weighting factor used in the objective function ($\gamma = 0, 5, 10$). In Table 5.2, each row represents a different combination of γ and delay percentiles corresponding to 90th and 99th percentiles (D90 to D99), and the columns report the solution statistics. We report the upper bound (UB) and lower bound (LB) on the Lagrangian bound obtained in the final iteration of the cutting plane algorithm. IP (integer programming) denotes the objective value of the restricted path formulation. Total Time refers to the overall runtime of the cutting plane algorithm. SPT represents the total time spent solving the subproblems, and MPT indicates the time required to solve the master problem. The optimality gap (Gap) is computed as the relative difference between the IP and the Lagrangian lower bound.

Table 5.2: Cutting Plane Algorithm Solution Statistics

Instance	UB	LB	IP	Gap %	Nm of Itr	Tot Tim	SPT(s)	MPT(s)
N	228933	228933	228933	0.00	241	15.32	3.72	2.41
$\gamma = 5$								
D90	229741.2	229741.3	229741.3	0.00	246	14.02	3.46	2.55
D91	230070	230070.1	230070.1	0.00	245	13.49	3.71	2.76
D92	230240	230240	230240	0.00	228	12.93	3.43	2.66
D93	230585.3	230585.3	230585.3	0.00	255	15.06	4.21	3.04
D94	230965.6	230965.6	230965.6	0.00	240	14.33	3.99	2.87
D95	232365.3	232365.3	232365.3	0.00	237	14.53	4.01	2.97
D96	234356.6	234356.6	234356.6	0.00	248	15.60	4.37	3.14
D97	249100.4	249100.4	249455	0.001	258	16.39	4.79	3.22
D98	264975.46	264975.46	266307	0.005	233	14.68	4.16	2.93
D99	357656.31	357656.31	361269	0.01	247	16.00	4.63	3.10
$\gamma = 10$								
D90	230522.5	230522.5	230522.5	0.00	247	15.49	4.31	3.14
D91	231180.1	231180.1	231180.1	0.00	236	14.66	4.03	2.91
D92	231520	231520	231520	0.00	239	14.82	4.02	3.02
D93	232210.7	232210.7	232210.7	0.00	241	14.99	4.20	2.98
D94	232966.7	232966.7	232966.7	0.00	231	14.41	3.87	2.86
D95	235748.1	235748.1	235748.1	0.00	252	16.19	4.76	3.16
D96	238977.4	238977.4	238977.4	0.00	247	16.02	4.56	3.17
D97	268635.65	268635.65	269174	0.002	240	15.20	4.20	3.07
D98	300569.18	300569.18	302688	0.007	236	18.93	5.38	3.60
D99	486499.10	486499.10	492408	0.012	237	21.28	5.87	3.90

In most cases, the Lagrangian bounds (UB, LB) are computed to optimality. For instances with higher delay percentiles, the optimality gap ($\frac{|IP-LB|}{IP}$) is less than 0.01%. The cutting plane algorithm converges in around 250 iterations. The total solution time for each instance is under 20 seconds, with both the subproblem and master problem solve times remaining below 6 seconds.

Table 5.3 summarizes the key cost components of the crew pairing solutions. These include the total pairing cost (TC), total flight time (TFT), total connection time (TCT) between duties, total elapsed time (TET) of all duties, and the total deadheading flight time (TDFT). We also report the delay cost associated with the nominal solution. Lastly,

we present the two cost components used when calculating crew payments: Cost1, which represents the total duty costs, and Cost2, which accounts for the fraction of pairing elapsed time.

Let N^* denote the set of duty nodes and A^* the set of duty arcs used in the optimal solution. For each duty i , the duty flight time TFT_i is calculated as the sum of the flight times of all flights in that duty. The total flight time (TFT) is then:

$$TFT = \sum_{i \in N^*} \sum_{f \in \mathcal{F}_i} SET_f \quad (5.1)$$

where $\mathcal{F}_i$ is the set of flights in duty i , and SET_f is the scheduled flight time for flight f .

The total connection time (TCT) is calculated as the time difference between the arrival time of the last flight in duty i (SAT_i^l) and the departure time of the first flight in the next duty j (SDT_j^o).

$$TCT = \sum_{(i,j) \in A^*} (SDT_j^o - SAT_i^l) \quad (5.2)$$

The elapsed time of a duty i (TET_i) is the difference between the departure time of the first flight and the arrival time of the last flight in the duty, and the total elapsed time TET is:

$$TET = \sum_{(i,j) \in N^*} (SAT_i^l - SDT_i^o) \quad (5.3)$$

For deadheading ($TDFT$), if any flight appears more than once in the schedule, we include the additional flight times in the total deadheading time. This is calculated as:

$$TDFT = \sum_{f \in \text{DeadheadingFlights}} \text{SET}_f$$

The top section in Table 5.3 shows results for the nominal model, while the middle and bottom sections correspond to the delay-averse model with $\gamma = 5, 10$, and d_f ranging from 90th (D90) to 99th percentile (D99). We report percentage increases relative to the nominal case. Specifically, the middle section shows results for $\gamma = 5$ and the bottom for $\gamma = 10$. The delay estimate d_f was obtained using historical flight performance from the year 2019 by matching each flight's origin, destination, and scheduled departure time.

Table 5.3: Sensitivity Analysis for $\gamma = 0, 5, 10$ and 90th-99th Delay Percentiles

$(\gamma = 0)$										
Costs	D90	D91	D92	D93	D94	D95	D96	D97	D98	D99
TC	228933	228933	228933	228933	228933	228933	228933	228933	228933	228933
DC	626.89	1006.6	1538.9	2136.8	3244	5795.6	11892	29750	159219	545901
TFT	29106	29106	29106	29106	29106	29106	29106	29106	29106	29106
TET	36368	36368	36368	36368	36368	36368	36368	36368	36368	36368
TCT	218002	218002	218002	218002	218002	218002	218002	218002	218002	218002
TDFT	2821	2821	2821	2821	2821	2821	2821	2821	2821	2821
TST	7262	7262	7262	7262	7262	7262	7262	7262	7262	7262
Cost1	53799.8	53799.8	53799.8	53799.8	53799.8	53799.8	53799.8	53799.8	53799.8	53799.8
Cost2	228933	228933	228933	228933	228933	228933	228933	228933	228933	228933
$(\gamma = 5)$ (%)										
Costs	D90	D91	D92	D93	D94	D95	D96	D97	D98	D99
TC	0.01	0.01	0.01	0.01	0.01	0.02	0.35	0.35	0.43	0.52
DC	-75.08	-77.95	-83.36	-84.79	-87.66	-88.33	-92.23	-86.74	-95.43	-95.2
TFT	-0.46	0.81	-0.19	-0.46	-0.48	-0.45	0.31	1.46	-0.23	-0.53
TET	3.4	4.78	1.24	1.43	1.03	4.24	3.87	5.75	5.78	5.43
TCT	-0.55	-0.78	-0.19	-0.23	-0.16	-0.68	-0.24	-0.55	-0.46	-0.3
TDFT	-4.75	8.37	-1.95	-4.71	-5	-4.61	3.15	15.1	-2.34	-5.46
TST	19	21	7	9	7	23	18	23	30	29
Cost1	1.73	2.24	0.64	0.96	0.57	2.18	1.48	2.46	3.43	3.16
Cost2	0.01	0.01	0.01	0.01	0.01	0.43	0.35	0.35	0.52	0.52
$(\gamma = 10)$ (%)										
Costs	D90	D91	D92	D93	D94	D95	D96	D97	D98	D99
TC	0.01	0.01	0.01	0.01	0.01	0.02	0.35	0.35	0.43	0.53
DC	-75.07	-77.95	-83.36	-84.79	-87.66	-88.33	-92.23	-86.74	-95.43	-95.2
TFT	1.15	1.58	-0.08	-2.22	-0.48	-0.45	-0.46	0.36	-0.23	-1.37
TET	4.57	5.06	1.09	-4.75	1.03	4.24	1	3.13	5.78	1.23
TCT	-0.75	-0.83	-0.17	0.81	-0.16	-0.68	0.24	-0.11	-0.46	0.4
TDFT	11.88	16.27	-0.78	-22.94	-5	-4.61	-4.75	3.76	-2.34	-14.14
TST	18	19	6	-15	7	23	7	14	30	12
Cost1	1.82	2.29	0.52	-1.64	0.57	2.18	0.84	1.57	3.43	1.33
Cost2	0.01	0.01	0.01	0.01	0.01	0.02	0.35	0.35	0.43	0.52

The main finding of this study is that incorporating delay risk costs yields delay-averse solutions that significantly reduce overall delay. Since the total flight time remains unchanged, the model shifts from selecting duties with shorter elapsed times to those with

longer ones. This shift reflects a preference for schedules with more buffer time, which helps mitigate the impact of delays.

When we look at the results presented in Table 5.3, we observe that solving the model under different delay estimates results in only a slight increase in scheduled costs—at most 0.55% compared to the nominal case. On the other hand, the delay penalty cost is reduced by over 70%, demonstrating the effectiveness of the delay-averse approach. We also note that as duty elapsed times increase by 2–3%, the connection times between duties become shorter.

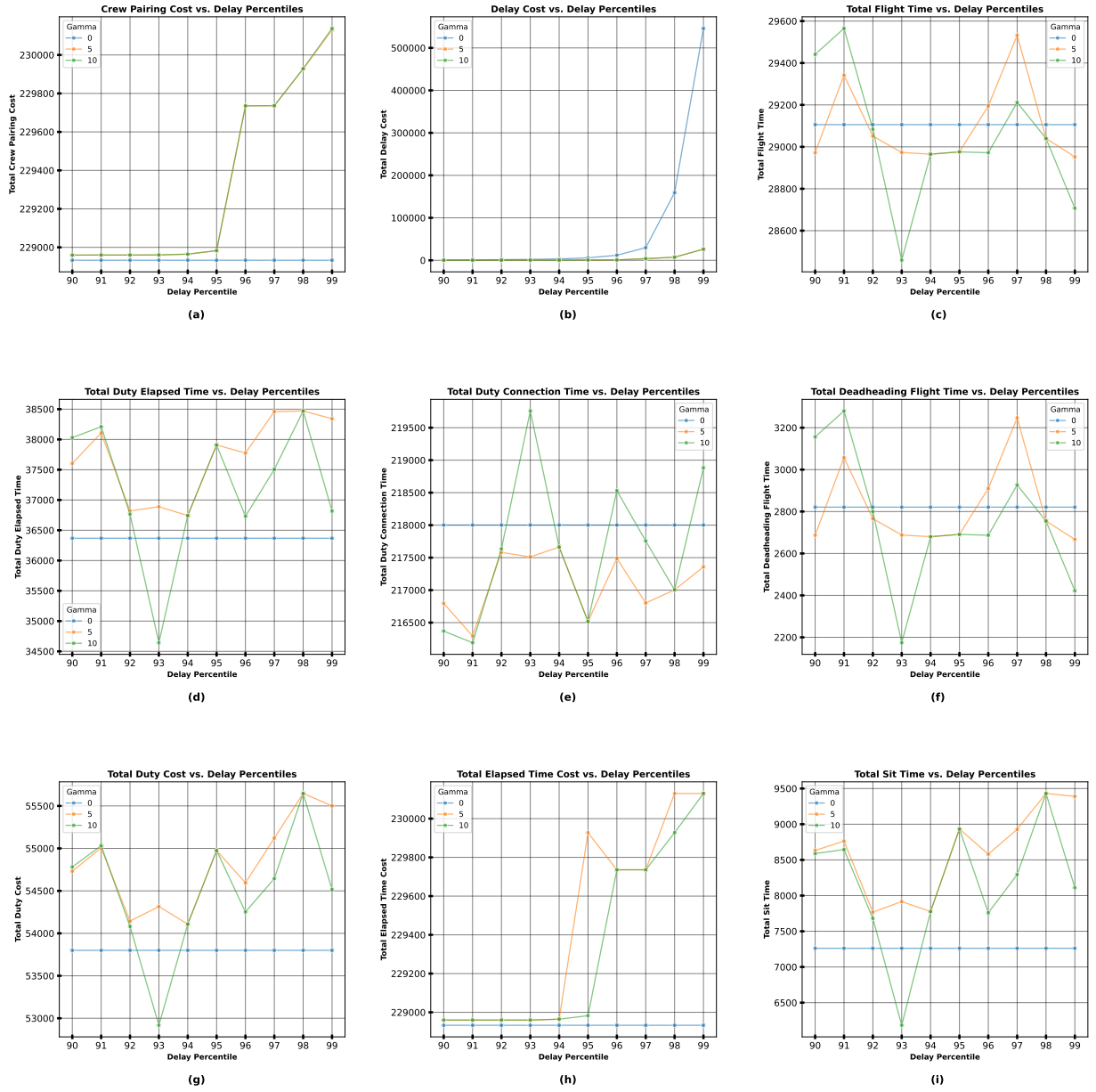


Figure 5.1: Sensitivity Analysis Results

Figure 5.1 summarizes the results of our sensitivity analysis and cost comparisons. The

x-axis represents the delay estimate percentiles used in the model, while the y-axis shows the corresponding cost components. As we vary the delay estimates, the Figure 5.1 illustrates how the scheduled cost, delay penalty, and total cost components change. This highlights the trade-offs between cost efficiency and delay robustness, and further supports the benefits of incorporating delay risk into the crew pairing model.

From Figure 5.1, we observe that as the cost associated with delay increases, certain duties are no longer selected by the model. As shown in Figure 5.1(d), total duty elapsed times are longer in the delay-averse solutions. This is because the model prioritizes pairings with longer sit times, allowing for more buffer time between consecutive flights, as illustrated in Figure 5.1(l). Consequently, pairing elapsed times also increase, as seen in Figure 5.1(h). In practice, these additional buffers help absorb delays and reduce the risk of further disruptions in the schedule. As a result, the delay-averse model produces crew schedules that are more resilient to delays compared to the nominal model.

Table 5.4: Flight coverage $\gamma = 5, 10$:

	$\gamma = 5$				$\gamma = 10$			
	Flight Coverage				Flight Coverage			
	1	2	3	4	1	2	3	4
D90	229	6	4	4	225	10	4	4
D91	225	11	3	4	222	15	1	5
D92	227	10	2	4	227	8	5	3
D93	229	7	2	5	230	8	1	4
D94	226	11	3	3	226	11	3	3
D95	228	8	3	4	228	8	3	4
D96	225	11	5	2	228	7	6	2
D97	223	11	7	2	224	12	5	2
D98	226	11	4	2	226	11	4	2
D99	227	9	5	2	227	11	4	1

For the nominal case, 227 flights are covered only once, 10 flights are covered twice, 1 flight is covered three times, and five flights are covered four times. As shown in Table 5.4, the delay averse model results in fewer deadheading assignments, even though we do not directly penalize deadheading costs. Each row corresponds to a different delay estimate, and the columns represent various values of γ .

5.3 Simulation

For each flight we match data from the BTS 2019 dataset using the scheduled arrival hour, origin, and destination. This gives us historical delay data for each flight across the year. We subtract the late aircraft delay from the arrival delay. This gives us the root delay. This approach is also used in Antunes et al. [2019]. Root delays can also be negative. This happens if a flight arrives earlier than the scheduled time or if the late aircraft delay is larger than the arrival delay. In both cases, the resulting root delay becomes negative.

We use kernel density estimation (KDE) to simulate flight delays and fit it to root delays using the SciPy library in Python. KDE is a non-parametric method to estimate the probability density function (PDF) of a random variable. In our work, we estimate the PDF of arrival delays. We do not assume any known distribution like normal or exponential for flight delays. Instead, we use KDE to get the shape of the PDF directly from the data following the approach in Biswal [2024]. This is helpful when a random variable does not follow standard distributions. From each KDE fit, we generate 100 random root delay samples. We add each root delay to the scheduled departure time (SDT) and the scheduled elapsed time (SET) to get actual arrival times (AAT).

In addition to the baseline KDE fitting, we also generate different delay scenarios. Instead of using the full delay data, we sample a subset of delays before fitting the KDE. In each

scenario, we fit the KDE to a specific upper percentile of the sampled delays, which we refer to as the p -th percentile KDE fitting. We define two groups in each scenario: R (regular delays) and IR (irregular delays). Based on these groups, we create four additional scenario groups. The first group includes 75% R and 25% IR samples, and we use the 70th, 80th, and 90th percentiles. The second group uses a 50%–50% mix of R and IR, again with the 70th, 80th, and 90th percentiles. The third group includes 25% R and 75% IR samples with the same percentiles. Finally, we consider a scenario with 100% IR delays only, using the 70th, 80th, and 90th percentiles. These scenarios allow us to analyze the performance of the solutions under different delay intensities.

We evaluate the simulations based on the realized Total Propagated Delay (TPGD), which is calculated as follows:

$$\text{TPGD} = \sum_{i \in N^*} \sum_{f \in \mathcal{F}_i} \text{pgd}_f, \quad \text{pgd}_f = \begin{cases} \tau - (\text{SDT}_{f+1} - \text{AAT}_f), & \text{if } \text{SDT}_{f+1} - \text{AAT}_f < \tau \\ 0, & \text{otherwise} \end{cases} \quad (5.4)$$

Here, pgd_f is the propagated delay for each flight f in the pairings, SDT_{f+1} is the scheduled departure time of the next flight $f + 1$, AAT_f is the actual arrival time of the current flight f , and τ is the minimum sit time between flights.

From the simulation runs, we can draw several key observations regarding the performance of the delay averse approach. Specifically, we observe that the Total Propagated Delay (TPGD) is, on average, better for delay averse solutions compared to the nominal approach. The difference between the nominal and delay averse cases becomes larger, particularly in the higher percentiles.

In Tables 5.5 to 5.9, the performance of delay averse solutions are presented with values of $\gamma = 5$ and $\gamma = 10$, respectively. When comparing these two solutions, we do not observe significant differences between them. This indicates that increasing the delay aversion parameter does not lead to improvements in the overall performance. This suggests that the model is relatively stable with respect to changes in the delay aversion level, and the choice of γ value has a diminishing effect.

In general, the delay averse solution gives lower total propagated delays (TPGD) by introducing more buffer time between flights within duties. This additional buffer time helps prevent delays from propagating further through the schedule. Consequently, we observe a reduction in duty connection times as the total available time for each duty is limited and fixed. This allocation of time ensures that the crew has more flexibility to handle minor delays and reduce the risk of significant disruptions in the flight schedule. The gap between the nominal solution and the delay-averse solution is more significant in the right tail of the TPGD distribution. This difference becomes more noticeable when considering more severe delay scenarios.

Table 5.5: Total Propagated Delays for 75R,25IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD

pth	75R, 25IR-70			75R, 25IR-80			75R, 25IR-90		
	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$
90	263.48	118.65	116.48	380.63	165.01	156.24	724.99	404.54	373.98
91	275.78	132.89	117.6	386.18	183.51	184.81	755.64	434.5	376.88
92	302.19	138.72	123.15	404.62	190.61	199.21	857.5	482.15	408.55
93	369.42	141.95	134.17	411.38	202.33	202.08	870.95	493.26	451.33
94	377.26	159.06	159.06	434.7	222.58	218.85	897.36	541.15	476.32
95	429.27	167.72	167.23	449.34	282.86	277.89	937.41	573.47	505.53
96	441.26	187.37	177.83	645	285.16	284.71	956.02	575.87	572.81
97	446.08	187.93	187.36	660.99	292.27	339.27	1196.69	612.76	575.77
98	521.6	191.99	187.92	776.13	341.82	592.22	1226.45	657.25	620.16
99	671.54	318.31	190.36	1097.23	659.48	623.11	1308.1	847.57	826.24
100	749.73	461.48	281.8	1322.8	771.1	653.36	1937.85	1294.83	1084.51

Table 5.6: Total Propagated Delays for 50R,50IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD

pth	50R, 50IR-70			50R, 50IR-80			50R, 50IR-90		
	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$
90	363.42	211.93	145.89	541.85	416.06	356.75	1103.93	603.72	489.45
91	427.16	231.03	171.7	637.32	450.93	367.7	1129.66	612.83	544.18
92	476.5	287.84	172.8	645.46	456.96	399.58	1152.8	622.15	577.85
93	545.07	295.2	174.54	656.95	483.7	424.41	1167.25	662.69	580.55
94	560.64	304.02	187.73	694.53	498.77	450.7	1274.96	684.47	598.52
95	598.01	307.93	209.56	730.8	521.09	453.1	1302.94	785.47	630.27
96	715.32	308.17	233.92	753.25	524.06	453.53	1385.07	844.87	639.39
97	729.68	314.1	292.01	777.27	529.03	458.55	1437.86	899.03	698.37
98	880.8	340.34	340.34	924.68	531.06	496.37	1464.34	960.57	865.98
99	919.47	355.21	354.97	985.34	566	613.17	1851.01	1220.42	995.69
100	1081.58	409.44	385.67	990.55	599.9	614.17	2002.43	1221.59	1225.56

Table 5.7: Total Propagated Delays for 25R,75IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD

pth	25R, 75IR-70			25R, 75IR-80			25R, 75IR-90		
	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$
90	676.36	302.61	210.18	778.23	396.14	375.17	1076.94	540.55	524.95
91	729.14	314.64	242.41	808.92	426.22	381.41	1114.03	550.36	538.67
92	752.21	322.37	259.81	847.32	432.88	409.26	1216.8	604.12	565.31
93	859.7	336.61	275.72	857.96	453.42	421.61	1245.4	714.18	623.57
94	880.2	407.77	304.55	888.24	473.78	459.6	1292.33	727.6	655.6
95	938.43	429.91	355.69	920.15	512.92	507.03	1340.7	742.85	728.22
96	938.98	432.63	376.29	928.52	554.84	529.28	1395.78	759.86	762.5
97	944.66	489.41	426.41	991.17	565.79	557.49	1564.22	812.21	851.98
98	1080.12	493.99	432.02	1115.75	573.43	565.74	1636	890.38	995.14
99	1124.6	534.12	517.88	1297.78	602.47	577.24	1698.6	986.47	1070.61
100	1442.88	534.97	534.11	1619.25	1023.98	882.81	1948.81	1115.78	1135.22

Table 5.8: Total Propagated Delays for 100IR Scenarios with (70,80,90);pth represents the pth percentile of the TPGD

pth	100IR-70			100IR-80			100IR-90		
	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$
90	441.61	254.21	281.99	703.52	350.13	297.16	1235.76	568.16	476.53
91	456.46	260.49	294.19	707.01	359.5	336.39	1263.02	575.33	482.02
92	463.53	282.73	304.24	718.62	442.2	340.4	1288.56	577.05	535.88
93	476.34	307.09	306.11	773.88	516.02	386.8	1303.04	584.98	589.49
94	486.32	327.8	313.64	792.89	526.73	573.79	1358.28	652.37	631.35
95	496.08	339.22	321.74	817.02	545.43	598.32	1382.32	663.05	693.38
96	558.24	382.63	338.18	876.92	584.76	606.07	1451.38	670.59	703.84
97	573.04	387.93	340.09	1048.99	632.69	628.58	1473.58	707.3	710.58
98	720.04	395.7	387.95	1264.63	861.38	798.68	1480.62	999.21	884.88
99	801.45	480.85	399.67	1301.57	881.45	849.21	1558.76	1106.24	944.57
100	1095.01	557.93	557.93	1400.98	911.89	881.14	1804.28	1182.22	1014.61

Table 5.9: Total Propagated Delays for 100R Scenario;pth represents the pth percentile of the TPGD

pth	100R-0		
	$\gamma = 0$	$\gamma = 5$	$\gamma = 10$
90	172.74	85.37	101.04
91	193.04	115.24	119.61
92	198.84	122.37	122.37
93	237.08	146.41	146.88
94	268.99	160.02	166.43
95	295.99	172.58	172.58
96	484.28	215.26	215.63
97	536.02	295.53	306.78
98	846.45	306.45	374.93
99	933.51	393.51	393.51
100	1004.48	464.48	464.48

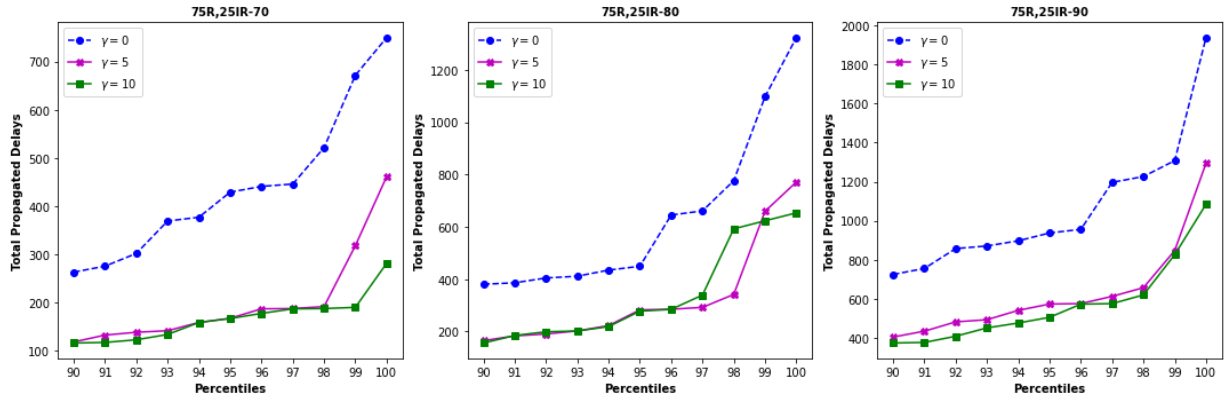


Figure 5.2: Total Propagated Delays for 75R,25IR Scenarios (70,80,90)

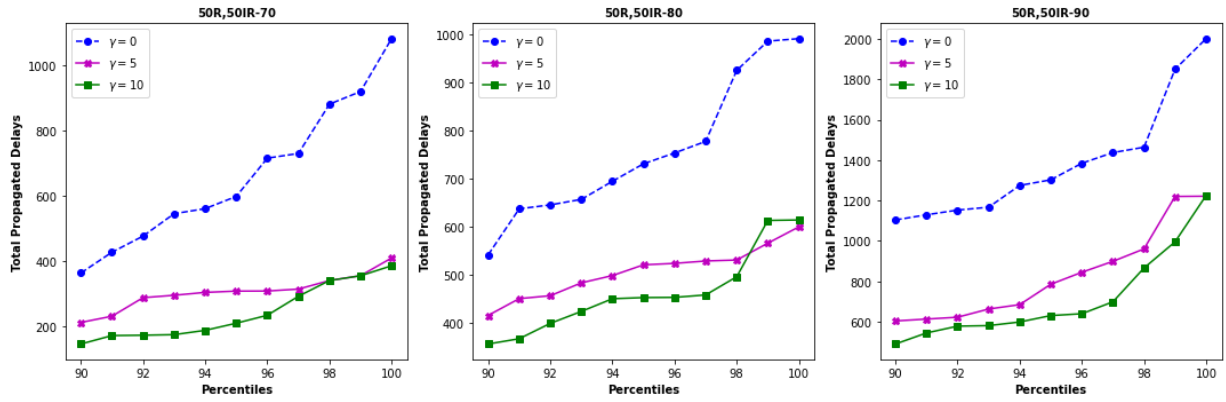


Figure 5.3: Total Propagated Delays for 50R,50IR Scenarios (70,80,90)

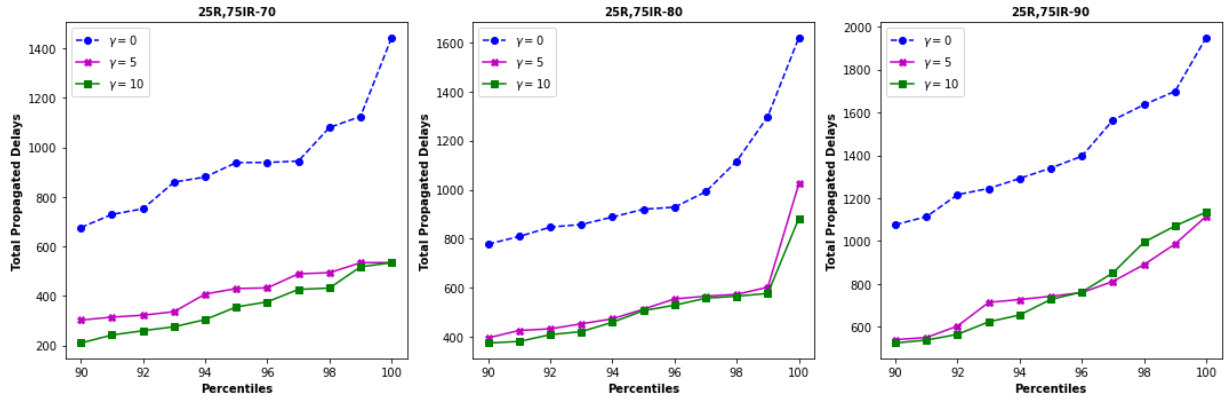


Figure 5.4: Total Propagated Delays for 25R,75IR Scenarios (70,80,90)

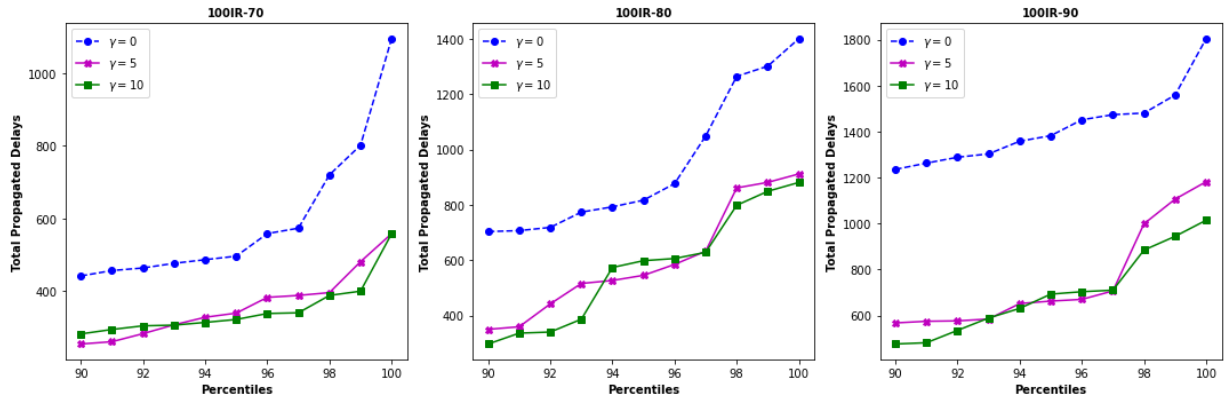


Figure 5.5: Total Propagated Delays for 100IR Scenarios (70,80,90)

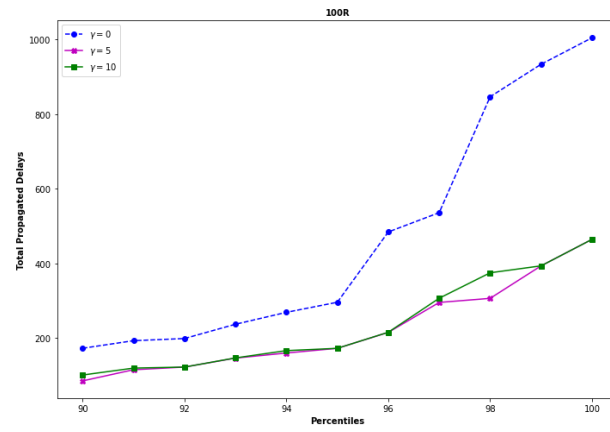


Figure 5.6: Total Propagated Delays for 100R

Chapter 6

Conclusions

This study focuses on reducing the gap between scheduled and actual crew costs. To achieve this, we develop a mixed-integer programming model that builds reliable crew pairings by accounting for flight delay uncertainty. We incorporate a pay-and-credit-based crew payment structure built on a duty network. This structure accounts for factors such as duty length, elapsed time, and guaranteed minimum pay. Traditional deterministic models typically neglect the impact of flight delays. To incorporate flight delays, we calculate delay penalties for each flight using buffer times, which are determined based on flight delay estimates and minimum connection time requirements. These penalties are then summed to calculate the total delay cost for each duty. This cost penalizes duties with insufficient buffer times between flight connections.

The crew pairing problem becomes complex due to the number of possible pairings. To solve this, we apply Lagrangian relaxation by relaxing the flight covering constraint, and we decompose the problem into $|K|$ subproblems. To solve the Lagrangian master problem, we use a cutting-plane algorithm to obtain bounds on the Lagrangian dual problem. The

paths generated during this algorithm are used to solve the path-based formulation and obtain integer solutions.

We test the nominal and delay-averse model on a 60-crew instance from BTS. In the numerical experiments, we use three values for the delay penalty weight γ : 0, 5, and 10, where $\gamma = 0$ represents the nominal case. For each duty, delay penalty costs are calculated using delay estimates from the 90th to 99th percentiles of each flight’s empirical delay distribution. Our findings demonstrate that incorporating delay risk costs significantly reduces overall delay risk. The model shifts its preferences from shorter duties to those with longer elapsed times, increasing the buffer between flights and reducing the risk of delay propagation.

We use propagated delays as a recovery method in the simulation to assess solution quality. This allows us to evaluate how well the schedules perform under realistic delay scenarios. We compare the total propagated delay of the delay-averse model against the nominal model. Simulation results show that the delay-averse model enhances operational stability by producing more resilient schedules.

This work contributes to the optimization of airline operations by integrating operational constraints, payment structures, and delay aversion into the crew pairing process. This work builds on [Biswal \[2024\]](#), who used flight networks and survival analysis to predict delays and adjust nominal pairing costs. While we adopt their simulation framework, our model differs by operating on duty networks, treating duties as full sequences of flights rather than isolated connections. The model can be extended to account for more complex airline regulations and additional factors, such as crew rest requirements. Future research could further test the model’s scalability with larger datasets and explore advanced statistical methods for predicting delays.

References

- Bts airline on-time performance data. <https://www.bts.gov>, 2024. Accessed: 2024-05-09.
- Yana Ageeva. *Approaches to incorporating robustness into airline scheduling*. PhD thesis, Massachusetts Institute of Technology, 2000.
- Divyam Aggarwal, D. Saxena, M. Emmerich, and S. Paulose. On large-scale airline crew pairing generation. *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, null:593–600, 2018. doi: 10.1109/SSCI.2018.8628699. URL <https://www.semanticscholar.org/paper/4c123be0c10aea637599f448988f880869e3aaa8>.
- Shervin AhmadBeygi, Amy Cohn, Yihan Guan, and Peter Belobaba. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management*, 14(5):221–236, 2008.
- Shervin Ahmadbeygi, Amy Cohn, and Marcial Lapp. Decreasing airline delay propagation by re-allocating scheduled slack. *IIE transactions*, 42(7):478–489, 2010.
- Mohamed Ben Ahmed, Maryia Hryhoryeva, Lars Magnus Hvattum, and Mohamed Haouari. A matheuristic for the robust integrated airline fleet assignment, aircraft routing, and crew pairing problem. *Computers & Operations Research*, 137:105551, 2022.

- R. Anbil, John J. H. Forrest, and W. Pulleyblank. Column generation and the airline crew pairing problem. *Documenta Mathematica*, pages 677–686, 1998. URL <https://www.semanticscholar.org/paper/dd1208ac3c685b5df5d56fd99e74141b4a2803ee>.
- E. Andersson, E. Housos, Niklas Kohl, and Dag Wedelin. Crew pairing optimization. 1998. doi: 10.1007/978-1-4615-5501-8_8. URL <https://www.semanticscholar.org/paper/54a5113694dba2161b45a2eb7095592929c3b6e1>.
- David Antunes, Vikrant Vaze, and António Pais Antunes. A robust pairing model for airline crew scheduling. *Transportation Science*, 53(6):1751–1771, 2019.
- C. Barnhart, A. Cohn, Ellis L. Johnson, D. Klabjan, G. Nemhauser, and P. H. Vance. 14 airline crew scheduling. 2002. URL <https://www.semanticscholar.org/paper/c9f96614b465324cfb9079d0bf502a48424c8fd6>.
- C. Barnhart, A. Cohn, Ellis L. Johnson, D. Klabjan, G. Nemhauser, and P. H. Vance. Airline crew scheduling. 2003. doi: 10.1007/0-306-48058-1_14. URL <https://www.semanticscholar.org/paper/dbd4733a1adb95520a4c4bf406d5fc0acabf6c0a>.
- D. Bayer. Pairing generation for airline crew scheduling. 2012. URL <https://www.semanticscholar.org/paper/5bd860bf7efeca33812c4f0a0e05e785b3447675>.
- Aniket Biswal. Learning and optimization for reliable airline operations. Master’s thesis, University of Waterloo, Waterloo, Canada, 2024.
- L. Clarke, Ellis L. Johnson, G. Nemhauser, and Z. Zhu. The aircraft rotation problem. *Annals of Operations Research*, 69:33–46, 1997. doi: 10.1023/A:1018945415148. URL <https://www.semanticscholar.org/paper/c832b0a3c161c141b91fee0a853df2e8a8fe5870>.

- Michael Dudley Delano Clarke. Irregular airline operations: a review of the state-of-the-practice in airline operations control centers. *Journal of Air Transport Management*, 4(2):67–76, 1998.
- Jens Clausen, Allan Larsen, Jesper Larsen, and Natalia J Rezanova. Disruption management in the airline industry—concepts, models and methods. *Computers & Operations Research*, 37(5):809–821, 2010.
- A. Cohn and C. Barnhart. Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, 51:387–396, 2003. doi: 10.1287/opre.51.3.387.14759. URL <https://www.semanticscholar.org/paper/105c6f2006d245ee2ddff2d12f59d010eb45756e>.
- J. Cordeau, G. Stojković, F. Soumis, and J. Desrosiers. Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science*, 35:375–388, 2000. doi: 10.1287/trsc.35.4.375.10432. URL <https://www.semanticscholar.org/paper/ce867bfcd8316156582fcceb3bb0a8d2458acc75>.
- G. Desaulniers, J. Desrosiers, M. Solomon, F. Soumis, Y. Dumas, S. Marc, and B. Rioux. Crew pairing at air france. 1993. doi: 10.1016/S0377-2217(96)00195-6. URL <https://www.semanticscholar.org/paper/4e1d2df8c616bc962c2e33e6ce909be3563713b9>.
- Viktor Dück, Lucian Ionescu, Natalia Kliewer, and Leena Suhl. Increasing stability of crew and aircraft schedules. *Transportation Research Part C: Emerging Technologies*, 20(1): 47–61, 2012.
- Michelle Dunbar, Gary Froyland, and Cheng-Lung Wu. Robust airline schedule planning: Minimizing propagated delay in an integrated routing and crewing framework. *Transportation Science*, 46(2):204–216, 2012.

- M. Ehrgott and D. Ryan. Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-criteria Decision Analysis*, 11:139–150, 2002a. doi: 10.1002/MCDA.321. URL <https://www.semanticscholar.org/paper/e912359940b967c90e265988b751590a1da07df4>.
- Matthias Ehrgott and David M Ryan. Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-Criteria Decision Analysis*, 11(3):139–150, 2002b.
- Chunhua Gao, Ellis Johnson, and Barry Smith. Integrated airline fleet and crew robust planning. *Transportation Science*, 43(1):2–16, 2009.
- B. Gopalakrishnan and Ellis L. Johnson. Airline crew scheduling: State-of-the-art. *Annals of Operations Research*, 140:305–337, 2005. doi: 10.1007/s10479-005-3975-3. URL <https://www.semanticscholar.org/paper/0ce5f65ee228398dc3afc2ec35136b7bebd99582>.
- Yufeng Guo. A decision support framework for the airline crew schedule disruption management with strategy mapping. In *Operations Research Proceedings 2004: Selected Papers of the Annual International Conference of the German Operations Research Society (GOR). Jointly Organized with the Netherlands Society for Operations Research (NGB) Tilburg, September 1–3, 2004*, pages 158–165. Springer, 2005.
- Randolph Hall. *Handbook of transportation science*, volume 23. Springer Science & Business Media, 2012.
- Atoosa Kasirzadeh, Mohammed Saddoune, and F. Soumis. Airline crew scheduling: models, algorithms, and data sets. *EURO Journal on Transportation and Logistics*, 6:111–137, 2014. doi: 10.1007/s13676-015-0080-x. URL <https://www.semanticscholar.org/paper/b49c45d605d7b111bf14ea4f39d02b0a66ea749a>.

- D. Klabjan. Large-scale models in the airline industry. 2005. doi: 10.1007/0-387-25486-2_6. URL <https://www.semanticscholar.org/paper/469a8d7cf121b91b03af68d511b309bb388a4b3b>.
- D. Klabjan, Ellis L. Johnson, G. Nemhauser, E. Gelman, and S. Ramaswamy. Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications*, 20:73–91, 2001a. doi: 10.1023/A:1011223523191. URL <https://www.semanticscholar.org/paper/ac97f7398e68e22c4364a2fb758ff1db6d72cf24>.
- D. Klabjan, Ellis L. Johnson, G. Nemhauser, E. Gelman, and S. Ramaswamy. Airline crew scheduling with regularity. *Transportation Science*, 35:359–374, 2001b. doi: 10.1287/trsc.35.4.359.10437. URL <https://www.semanticscholar.org/paper/d836f34b0dd17c7da37741002f78ea69b95a829f>.
- D. Klabjan, Ellis L. Johnson, G. Nemhauser, E. Gelman, and S. Ramaswamy. Airline crew scheduling with time windows and plane-count constraints. *Transportation Science*, 36: 337–348, 2002. doi: 10.1287/trsc.36.3.337.7831. URL <https://www.semanticscholar.org/paper/a557865a6c7407cd3004df6ac4740c4ed0b476e6>.
- Diego Klabjan, Ellis L Johnson, George L Nemhauser, Eric Gelman, and Srini Ramaswamy. Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications*, 20:73–91, 2001c.
- Niklas Kohl, A. Larsen, J. Larsen, A. Ross, and S. Tiourine. Airline disruption management - perspectives, experiences and outlook. *Journal of Air Transport Management*, 13:149–162, 2007a. doi: 10.1016/J.JAIRTRAMAN.2007.01.001. URL <https://www.semanticscholar.org/paper/bb1dd53a3ee04a7ad3f42b92e87547c81379984f>.

- Niklas Kohl, Allan Larsen, Jesper Larsen, Alex Ross, and Sergey Tiourine. Airline disruption management—perspectives, experiences and outlook. *Journal of Air Transport Management*, 13(3):149–162, 2007b.
- Shan Lan, J. Clarke, and C. Barnhart. Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. *Transportation Science*, 40:15–28, 2006. doi: 10.1287/trsc.1050.0134. URL <https://www.semanticscholar.org/paper/b6c4f153413dd458cf85405bdc1344751c6cd27a>.
- Ladislav Lettovsky, Ellis L. Johnson, and G. Nemhauser. Airline crew recovery. *Transportation Science*, 34:337–348, 2000. doi: 10.1287/trsc.34.4.337.12316. URL <https://www.semanticscholar.org/paper/0d28b1712f3201f43c2eee84bd3881285088f901>.
- Da Lu. *Robust optimization for airline scheduling and vehicle routing*. Ph.d. thesis, University of Waterloo, Waterloo, ON, Canada, 2014. Available at: <https://uwspace.uwaterloo.ca/items/8eb6ef31-c8a0-4262-8323-62fae135cf79>.
- Da Lu and Fatma Gzara. The robust crew pairing problem: model and solution methodology. *Journal of Global Optimization*, 62:29–54, 2015.
- A. Makri and D. Klabjan. A new pricing scheme for airline crew scheduling. *INFORMS Journal on Computing*, 16:56–67, 2004. doi: 10.1287/ijoc.1020.0026. URL <https://www.semanticscholar.org/paper/3cbc1f08820e8f029358818fe6e0228899d7868a>.
- Claude P Medard and Nidhi Sawhney. Airline crew scheduling from planning to operations. *European Journal of Operational Research*, 183(3):1013–1027, 2007.
- A. Mercier and F. Soumis. An integrated aircraft routing, crew scheduling and flight retiming model. *Computers & Operations Research*, 34:2251–2265, 2005.

- doi: 10.1016/j.cor.2005.09.001. URL <https://www.semanticscholar.org/paper/0836aee761e083477934b2fdce1c482eacbec695>.
- A. Mercier, J. Cordeau, and F. Soumis. A computational study of benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research*, 32:1451–1476, 2003. doi: 10.1016/j.cor.2003.11.013. URL <https://www.semanticscholar.org/paper/a57dda843630b7f0b046f0f6591239742039ef60>.
- Rüdiger Nissen and Knut Haase. Duty-period-based network model for crew rescheduling in european airlines. *Journal of Scheduling*, 9:255–278, 2006.
- Axel Parmentier. Aircraft routing: complexity and algorithms. 2013. URL <https://www.semanticscholar.org/paper/d9cd0d9e84fedf995addc18afa3414a13427ef1a>.
- Reuters. Air canada pilots ratify labor deal, avoiding strike. *Reuters*, 10 2024. URL <https://www.reuters.com/business/aerospace-defense/air-canada-pilots-ratify-labor-deal-2024-10-10/>. Accessed: 2025-05-09.
- J. Rosenberger, A. Schaefer, D. Goldsman, Ellis L. Johnson, A. Kleywegt, and G. Nemhauser. A stochastic model of airline operations. *Transportation Science*, 36:357–377, 2002. doi: 10.1287/trsc.36.4.357.551. URL <https://www.semanticscholar.org/paper/86d26e3b5bad4ac7477db12025fba01b960fef15>.
- J. Rosenberger, Ellis L. Johnson, and G. Nemhauser. A robust fleet-assignment model with hub isolation and short cycles. *Transportation Science*, 38:357–368, 2004. doi: 10.1287/trsc.1030.0038. URL <https://www.semanticscholar.org/paper/84e2459f97351722511500c8e1db091e29b6dc06>.
- Jay M Rosenberger, Andrew J Schaefer, David Goldsman, Ellis L Johnson, Anton J Kleywegt, and George L Nemhauser. Simair: A stochastic model of airline operations. In

- 2000 Winter Simulation Conference Proceedings (Cat. No. 00CH37165), volume 2, pages 1118–1122. IEEE, 2000.
- R. Sandhu and D. Klabjan. Integrated airline fleet and crew-pairing decisions. *Oper. Res.*, 55:439–456, 2007. doi: 10.1287/opre.1070.0395. URL <https://www.semanticscholar.org/paper/3010606a74052cfa3262d22d0d37059a80010937>.
- A. Schaefer, Ellis L. Johnson, A. Kleywegt, and G. Nemhauser. Airline crew scheduling under uncertainty. *Transportation Science*, 39:340–348, 2005a. doi: 10.1287/trsc.1040.0091. URL <https://www.semanticscholar.org/paper/bdbb6fd97e77d3a492109d3bd5e844f30edeae0f>.
- Andrew J Schaefer, Ellis L Johnson, Anton J Kleywegt, and George L Nemhauser. Airline crew scheduling under uncertainty. *Transportation Science*, 39(3):340–348, 2005b.
- Sergey Shebalov and D. Klabjan. Robust airline crew pairing: Move-up crews. *Transportation Science*, 40:300–312, 2006a. doi: 10.1287/trsc.1050.0131. URL <https://www.semanticscholar.org/paper/35fe1bd9279f49cbebcd090c02db90e73014a003>.
- Sergey Shebalov and Diego Klabjan. Robust airline crew pairing: Move-up crews. *Transportation Science*, 40(3):300–312, 2006b.
- Hanif D Sherali, Ebru K Bish, and Xiaomei Zhu. Airline fleet assignment concepts, models, and algorithms. *European Journal of Operational Research*, 172(1):1–30, 2006.
- Barry C. Smith and Ellis L. Johnson. Robust airline fleet assignment: Imposing station purity using station decomposition. 2005. doi: 10.1287/trsc.1060.0153. URL <https://www.semanticscholar.org/paper/d5b1cfa0806f402f81fdc0a7d4f1716670a1552d>.

- Chellappan Sriram and A. Haghani. An optimization model for aircraft maintenance scheduling and re-assignment. *Transportation Research Part A-policy and Practice*, 37:29–48, 2003. doi: 10.1016/S0965-8564(02)00004-6. URL <https://www.semanticscholar.org/paper/1d5078406d8859db964c09d73fca2a5b2e914dac>.
- M. Stojkovic, F. Soumis, and J. Desrosiers. The operational airline crew scheduling problem. *Transportation Science*, 32:232–245, 1997. doi: 10.1287/trsc.32.3.232. URL <https://www.semanticscholar.org/paper/2003c4ffdd654837f8fd5b0530eaec4f3431f839>.
- Mirela Stojković, Francois Soumis, and Jacques Desrosiers. The operational airline crew scheduling problem. *Transportation Science*, 32(3):232–245, 1998.
- Xuting Sun, Sai-Ho Chung, and Hoi-Lam Ma. Operational risk in airline crew scheduling: do features of flight delays matter? *Decision Sciences*, 51(6):1455–1489, 2020.
- Bassy Tam, Matthias Ehrgott, David Ryan, and Golbon Zakeri. A comparison of stochastic programming and bi-objective optimisation approaches to robust airline crew scheduling. *OR Spectrum*, 33:49–75, 2011.
- Hatice Tekiner, S. Birbil, and Kerem Bülbül. Robust crew pairing for managing extra flights. *Computers & Operations Research*, 36:2031–2048, 2009a. doi: 10.1016/j.cor.2008.07.005. URL <https://www.semanticscholar.org/paper/8370f3f9c727b4d0ca8ba18ecf06568ef8683398>.
- Hatice Tekiner, Ş İlker Birbil, and Kerem Bülbül. Robust crew pairing for managing extra flights. *Computers & Operations Research*, 36(6):2031–2048, 2009b.
- P. H. Vance, C. Barnhart, Ellis L. Johnson, and G. Nemhauser. Airline crew scheduling: A new formulation and decomposition algorithm. *Oper. Res.*, 45:188–200,

1997. doi: 10.1287/opre.45.2.188. URL <https://www.semanticscholar.org/paper/fb781230cd8cad44c91970d17e35cc90e5cd5d0a>.
- O. Weide, D. Ryan, and M. Ehrgott. An iterative approach to robust and integrated aircraft routing and crew scheduling. *Computers & Operations Research*, 37:833–844, 2010a. doi: 10.1016/j.cor.2009.03.024. URL <https://www.semanticscholar.org/paper/e2090a075ca8ba693623376882e08aa2dda4ce05>.
- Oliver Weide, David Ryan, and Matthias Ehrgott. An iterative approach to robust and integrated aircraft routing and crew scheduling. *Computers & Operations Research*, 37(5):833–844, 2010b.
- Xin Wen, Xuting Sun, Yige Sun, and Xiaohang Yue. Airline crew scheduling: Models and algorithms. *Transportation research part E: logistics and transportation review*, 149: 102304, 2021.
- Shuai Wu, Enze Liu, Rui Cao, and Qiang Bai. Airline recovery problem under disruptions: A review. *arXiv preprint arXiv:2401.04866*, 2024.
- S. Yan and Yu-Ping Tu. A network model for airline cabin crew scheduling. *Eur. J. Oper. Res.*, 140:531–540, 2002. doi: 10.1016/S0377-2217(01)00215-6. URL <https://www.semanticscholar.org/paper/2bc4a5fcd78ca8b4106ae1c946b1482370cdfea8>.
- Yufeng Yao. Topics in fractional airlines. 2007. URL <https://www.semanticscholar.org/paper/8321f4541e9cf74e0a5d448b00fa452137006e0f>.
- Xugang Ye. Airlines’ crew pairing optimization: A brief review. 2007. URL <https://www.semanticscholar.org/paper/3919d9221641d9a3eeaa975d8798d8a4b365c135>.

Joyce W. Yen and J. Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40:3–14, 2006a. doi: 10.1287/trsc.1050.0138. URL <https://www.semanticscholar.org/paper/169998d64a210b20d083113ffa2f0427b87a94f9>.

Joyce W Yen and John R Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3–14, 2006b.

Gang Yu, Michael F. Argüello, Gao Song, Sandra M. McCowan, and A. White. Special issue: 2002 franz edelman award for achievement in operations research and the management sciences: A new era for crew recovery at continental airlines. *Interfaces*, 33: 5–22, 2003. doi: 10.1287/inte.33.1.5.12720. URL <https://www.semanticscholar.org/paper/5909fc82cbece474d36506dbca6dabad59e4e953>.