AN APPLICATION OF STOCHASTIC PROGRAMMING ON ROBUST AIRLINE SCHEDULING

A THESIS

SUBMITTED TO THE DEPARTMENT OF INDUSTRIAL ENGINEERING

AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

By

Nil Karacaoğlu

July, 2014

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. M. Selim AKTÜRK(Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Prof. Hande YAMAN PATERNOTTE(Co-Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Özlem ÇAVUŞ

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Seçil SAVAŞANERİL TÜFEKÇİ

Approved for the Graduate School of Engineering and Science:

Prof. Dr. Levent Onural
Director of the Graduate School

ABSTRACT

AN APPLICATION OF STOCHASTIC PROGRAMMING ON ROBUST AIRLINE SCHEDULING

Nil Karacaoğlu
M.S. in Industrial Engineering
Supervisor: Prof. M. Selim AKTÜRK
Co-Supervisor: Prof. Hande YAMAN PATERNOTTE

July, 2014

The aim of this study is to create flight schedules which are less susceptible to unexpected flight delays. To this end, we examine the block time of the flight in two parts, cruise time and non-cruise time. The cruise time is accepted as controllable within some limit and it is taken as a decision variable in our model. The non-cruise time is open to variations. In order to consider the variability of non-cruise times in the planning stage, we propose a nonlinear mixed integer two stage stochastic programming model which takes the non-cruise time scenarios as input. The published departure times of flights are determined in the first stage and the actual schedule is decided on the second stage depending on the non-cruise times. The objective is to minimize the airline's operating and passenger dissatisfaction cost. Fuel and CO_2 emission costs are nonlinear and this nonlinearity is handled by second order conic inequalities. Two heuristics are proposed to solve the problem when the size of networks and number of scenarios increase. A computational study is conducted using the data of a major U.S. carrier. We compare the solutions of our stochastic model with the ones found by using expected values of non-cruise times and the company's published schedule.

Keywords: Airline Scheduling, Stochastic Programming, Robust Optimization, Nonlinear Programming.

ÖZET

RASSAL PROGRAMLAMANIN DAYANIKLI HAVAYOLU ÇİZELGELEME ÜZERİNDE UYGULANMASI

Nil Karacaoğlu
Endüstri Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Prof. Dr. M. Selim AKTÜRK
Eş-Tez Yöneticisi: Prof. Dr. Hande YAMAN PATERNOTTE
Temmuz, 2014

Bu çalışmanın amacı beklenmeyen uçuş gecikmelerinden daha az etkilenen çizelgeler yaratmaktır. Bu amaç doğrultusunda, uçuşun blok süresini seyir süresi ve seyir dışı süre olmak üzere iki kısımda inceledik. Seyir süresi belli limitler dahilinde kontrol edilebilir kabul edildi ve modelimizde karar değişkeni olarak alındı. Seyir dışı süre ise değişikenliğe açıktır. Seyir dışı sürenin değişkenliğini planlama aşamasında göz önünde bulundurmak adına, seyir dışı süre senaryolarını girdi olarak alan karma tamsayılı doğrusal olmayan iki aşamalı rassal model önerdik. Uçuşların yayınlanmış kalkış zamanlarına ilk aşamada karar verildi ve gerçekleşen çizelge ise seyir dışı süresi senaryolarına göre ikinci aşamada belirlendi. Amaç havayolu şirketinin işletme ve yolcu memnuniyetsizliği maliyetini enazlamaktır. Ağın boyutu ve senaryo sayısı arttıkça problemi çözebilmek adına iki sezgisel algoritma geliştirildi. ABD'li büyük bir havayolu şirketinin verileri kullanılarak sayısal bir çalışma gerçekleşti ve bizim rassal modelimizin sonuçları seyir dışı sürenin beklenen değerleri kullanılarak bulunan sonuçla ve şirketin yayınlanmış çizelgesiyle kullanıldığında bulunan sonuçla karşılaştırıldı.

Anahtar sözcükler: Havayolu Çizelgeleme, Rassal Programlama, Gürbüz Optimizasyon, Doğrusal Olmayan Programlama.

to my mother...

Acknowledgement

Foremost, I want to express my deepest gratitude to my advisor Prof. Selim Aktürk for his excellent guidance, patience, support and understanding. He is a great advisor and I have learned many valuable lessons from him. I would like to thank my co-advisor Prof. Hande Yaman. This thesis would not take its current form without her valuable ideas. I could not imagine better advisors for my MS study.

I also would like to acknowledge the financial support of The Scientific and Technological Research Council of Turkey (TUBITAK) for the Graduate Study Scholarship Program they awarded.

I would like to thank my officemates Nihal Berktaş and Özüm Korkmaz. We were together during the sleepless nights to finish our theses. I would like to Merve Meraklı, Burcu Tekin, Hüseyin Gürkan, Oğuz Çetin being in the same office with them was always a pleasure for me. I will also remind my homemate Özge Şafak, her presence made the time I spent at home enjoyable.

I would like to thank my mother, father and brother. I would not be able to finish this thesis without their continuous support. My mother believed in me even when I did not believe in myself. Her presence and motivation is priceless for me. Last but not least, I want to express my gratitude to my grandmother since she constitutes a great example for me with her courage and ambition. I love them all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main aim of Two Stage Stochastic Robust Airline Scheduling is developing a flight schedule that is less susceptible to unexpected flight delays and that minimizes airline operating, passenger delay and disruption costs at the same time. Creating such schedule is a challenging problem with many parameters like demand of flights, passenger connections, cost parameters and decision variables such as departure times and cruise times of flights. In this study, to solve this problem a mathematical model is developed and implemented in Java with a connection to CPLEX, a commercial optimization software.

1.1 Motivation

Even though the number of passengers who prefer air travel increased considerably in the past decade, the entrance of new players to the airline industry and government regulations increased competition. In order to thrive in this competitive industry, airlines should adopt operational research's methodologies to utilize their expensive resources efficiently.

Airline industry is one of the sectors which have to consider and control high number of factors. The large networks, the number of passenger and aircraft connections, working requirements of crews are some of these variables. The companies need to take several long term decisions in the planning process. According to Belobaba [1], the most important decisions that are faced by airlines during the planning process are fleet composition, route planning and schedule development. In the fleet planning phase, airlines decide the types and the number of aircraft they will purchase. Route planning is determining which routes they will serve. The schedule development phase is composed of four different tasks; planning the frequency, the departure times, hence roughly the arrival times of flights, determining fleet assignment and aircraft rotations. These different decisions should be considered simultaneously for effective airline scheduling. However, even solving these problems individually is a difficult job that involves millions of variables. In this thesis, we focused on determining departure and arrival times of flights by assuming that frequency of flights, fleet assignments and rotations are given and fixed.

Each flight is assigned scheduled block time at airline scheduling phase that is equal to the duration from its scheduled departure time to scheduled arrival time. While creating this schedule, the airlines should consider minimum aircraft turnaround time, which is necessary in order to prepare the aircraft for the next flight, as well as the minimum passenger turnaround time that is the time required for passengers to connect from their current flight to the next flight in their itinerary. However, longer block times might lead to under utilization of aircraft by keeping such expensive equipments idle on the ground. In this work, idle time is referred as the time aircraft spend on the ground from after their arrival and preparation process is handled to their departure for the next flight. On the other hand, shorter block times may result in aircraft, passenger delays and disruptions. According to Desphande and Arıkan [2] airlines have tendency to assign shorter scheduled block times to reduce operating cost. However, airlines work with tight profit margins, usually less than 2%, and flight delays decrease this profit. U.S. Department of Transportation (DOT) considers a flight late if it arrives its destination 15 minute or more later than its scheduled arrival time. Flight delays are affected by various uncontrollable factors. Weather delays, security delays, and national aviation delays account for approximately 4%, 40% and 3% of airline delays in the last ten years, respectively. Moreover, flight delays constitute a major component of airline's cost. According to the report of Joint Economic Committee of the U.S. Congress estimated cost of airline delays to U.S. economy is $41 billion dollar in 2007 [3]. Moreover, these delays might also lead to negative consequences for the passengers. The passengers are considered as disrupted if they miss the next flight in their itinerary because of the late arrival of their current flight. Robust airline scheduling is a pro-actively taking possible flight delays into account in the airline scheduling phase and considering their effect on passengers and consequent flights of the same aircraft. In this way, the profitability of airlines can be improved while passenger disruptions and delays are reduced. Since this complex system involves many different components, optimization methods should be adopted to obtain solutions in a reasonable amount of time.

Even though robustness is a way to increase profit of airlines, quantifying and defining it is a challenging task. Robustness comes with a cost, hence companies should decide how much they are willing to pay for a robust schedule. Robustness can be obtained by inserting more idle time in the system, but this may lead to under utilization of aircraft. Another way is considering the effect of delays on passengers and subsequent flights in terms of cost at the planning stage.

1.2 Contributions

The airlines determine and publish a flight schedule in the planning phase. However, the actual departure and arrival times might deviate from the published departure times because of unexpected delays and not properly planned schedules.

Airlines separate actual block time of flights into five different components: departure delay, taxi-out, cruise time, taxi-in and arrival delay. The total duration of departure delay, taxi-out, taxi-in and arrival delay is also called as non-cruise time. Cruise time is less susceptible to variations, hence it can be considered as deterministic. However, non-cruise times are affected by weather conditions, airport congestion and air traffic.

Idle times, delays and passenger disruptions depend on the published flight schedule no matter what the actualized non-cruise time of the flight is. If non-cruise times are realized shorter than planned, it would lead to idle time and if it was longer, it may cause delays. Thus, the published schedule is the major determinant factor on airline's operating cost.

In our study, we aim to develop a flight schedule that is less susceptible to unexpected flight delays by developing and solving nonlinear mixed-integer two-stage stochastic programming model. In the first stage, we decide on the published schedule. In the second stage, according to the realized non-cruise times adjusting the speed of the aircraft is considered as recourse action. The published schedule is determined by taking into account the operating and passenger costs of different non-cruise time scenarios. Idle time insertion and adjusting the speed of the aircraft are considered options to obtain a robust schedule. Moreover, delay cost is included in the objective function and it depends on the number of passengers as well as the duration of delay.

In order to consider the variability of non-cruise times in the planning stage, we integrated several non-cruise time scenarios into our model. In this way, instead of using a single value for non-cruise time of each flight we utilize more information to capture the variability. The non-cruise time scenarios are specific to origin and destination airports of the flights, hence the information about the congestion of the airports are included for a more realistic approach. The scenarios consist of departure delay, taxi-out, taxi-in and arrival delay information for each airport. Moreover, critical airports are determined and more information about these airports is introduced. The data used in this process is obtained from the database of Bureau of Transportation Statistic (BTS).

Assigning longer block times is an irreversible decision and it may lead to keeping such expensive resources idle, hence increases cost. In our study, we consider trade-off between inserting idle time, speeding up the aircraft or experiencing delay in the system for different non-cruise time scenarios. The published block time in the first stage is determined by considering its effect on different realizations. Insertion of idle time would decrease the utilization of aircraft and crew. On the other hand, speeding up the aircraft increases the fuel consumption and CO_2 emission which might be more costly than inserting idle time in some cases. Another option is allowing passengers and aircraft to experience delays. The cost of delay is handled by introducing cost of passengers disruption and delay in the objective function. Passenger disruption is represented by a binary variable in our model. Speeding up the aircraft and experiencing delays are scenario specific, however longer block times, which is equivalent to preferring to inserting idle time, affects every scenario. Hence, the consequences of adopting these options are examined on each scenario and over-all system.

Moreover, the departure times also affect the market share of airlines. They prefer to schedule the flights where the demand is high, especially at the airports where the competition for the same route is high. Hence, in order to protect the current market share departure times of flights are allowed to change within some limits from the published schedule of the airline generated. In addition, existing passenger and aircraft connections are kept feasible in the newly generated schedule.

One of the contributions of our study is generation of a valid inequality to speed up the solution process. Another important contribution is our way of handling non-linear cost terms in the objective function. This non-linearity is handled by introducing second order conic inequalities into formulation. The proposed nonlinear mixed-integer two-stage stochastic programming model is solved with a commercial solver IBM ILOG CPLEX. The decrease in cost, obtained by introducing non-cruise time scenarios instead of using a single value, is presented in computational study section. Furthermore, aircraft utilization is increased and the number of disrupted passengers is decreased.

In order to solve the problem with large number of scenarios, two heuristics are developed. The heuristic takes the published schedule as given and evaluates its impact on each scenario.

1.3 Overview

In the next Chapter, brief information about stochastic programming and extensive review about scenario generation in airline operations are provided. Moreover, a short review on robust optimization in airline flight scheduling and cruise time controllability is given.

In Chapter 3, the dynamics of the problem, the parameters, and the model are explained. In order to strengthen the formulation, a valid inequality is proposed and its validity is proved. Moreover, the conic representation of the nonlinear cost function and conic reformulation of the model are demonstrated.

Generation of non-cruise time scenarios is explained in detail in Chapter 4. The information about the congestion of airports is given and the scenario generation mechanism is demonstrated on Chicago O'Hare Airport and the calculation of non-cruise time of a flight is shown.

A numerical example on a small network, which involves two aircraft and eight flight legs, is given in the Chapter 5. A new schedule is generated by solving the model for four different non-cruise time scenarios. The performance of stochastic programming solution is compared with the performance of using optimistic, pessimistic and expected times for non-cruise time values.

A heuristic algorithm is given in Chapter 6 in order to solve the problems with large number of scenarios on large networks.

In the computational study section, two networks are considered. Network 1 contains 31 flight legs and 9 aircraft, whereas network 2 is composed of 114 flight legs and 31 aircraft. The performance improvements obtained by solving the stochastic model on network 1 for 18 scenarios, instead of using the published schedule of airline or solving the model by using single deterministic value for non-cruise times are provided. Furthermore, the performance of heuristic in terms of cost and CPU time are demonstrated on network 1 for 18 scenarios. The performance of heuristic algorithm in terms of cost and CPU time instead of using expected values of non-cruise times is demonstrated on network 1 for 228 scenarios and on network 2 for 104 scenarios. Moreover, factor analysis is conducted in order to analyze the effect of cost parameters on the quality of the solution as well as devoted block time of each flight. Finally, in Chapter 8 we considered the extensions of the problem for the future studies.

Chapter 2

Literature Review

In the first chapter of this section, review about airline flight scheduling is given. Brief summary of stochastic programming and detailed literature about scenario generation in airline industry constitute the following section. Finally, introductory information to second order cone programming is provided.

2.1 Airline Scheduling

Optimization methods have been adopted by airlines since late 1970's as a result of increasing competition in the industry. At the beginning, operations research practices were restricted to revenue management. However, in the recent decades its application is extended to other areas [4].

Schedule Design, Fleet Assignment, Aircraft Maintenance Routing and Crew Scheduling are the four core steps in airline scheduling. Since the combination of these steps causes computational complexity, the problems are generally considered separately in the current literature. In the schedule design phase, airlines decide on which markets they would serve, with what frequency in order to match the forecasted demand, and departure times of flights are determined to generate an initial schedule. The assignment of specific fleet types to flights to match the seat capacity of aircraft with the demand for the flight is decided on fleet assignment phase. Aircraft need to go under regular maintenance in order to continue their operation. In aircraft maintenance routing phase, feasible sets of flight legs of aircraft are determined such that maintenance requirements of aircraft are satisfied. The given fleet assignment is an input in this stage. In crew scheduling, assignment of crews to flights is handled by considering regulations. Detailed review about airline operations and usage of optimization methods in the industry are given in Barnhart et al. [4].

High volume of air traffic, congestion, weather or security issues cause deviations from schedules. Bureau of Transportation Statistics reported that approximately 21% of U.S. domestic flights are delayed whose 5% is air-carrier delay, 5% is National Aviation System delay, 7% is late arriving aircraft, 1% is canceled flights and the left is weather delay, diverted flights and security delay [5]. The deviation from the schedule not only affects airlines, which work with tight profit margins, but also passengers. Companies face with incremental cost and decrease in revenue due to delays. On the other hand, passengers see increases in the time required for travel, experience inconvenience and stress. In 2007, delays caused 8.3 billions dollar cost industry wide [6]. Robust optimization is one of the approaches which is applied in order to create schedules which are less susceptible to unexpected flight delays.

2.1.1 Robust Airline Scheduling

In the schedule design phase, airlines usually assume that flights depart and land according to the published schedule. Even though this approach increases aircraft utilization, its effect on operational cost is significant when deviations from the plans are experienced. In reality the weather conditions, security issues or crew sickness cause deviations from the plans. In robust flight scheduling, these deviations are considered in airline scheduling phase and preventive actions are taken.

Lan et al. [7] differentiate between the propagated and non-propagated delay and in the first part of their work they focused to minimize expected propagated delay. They formulated a mixed-integer program that allows changing the assignments of aircraft to flights and finds an aircraft rotation. In the first part, they kept the departure times of flights fixed. They also considered re-timing the departure times of flights within a small time window, when re-assigning fleet is not an option, to minimize the expected number of passenger disruptions.

Marla and Barnhart [8] focused on aircraft routing in order to create a robust schedule. They considered three different models for generating routing: extreme-value based, probabilistic approach and tailored approach which is proposed by Lan, Clarke and Barnhart. Marla and Barnhart measured the quality of routings created by different models using simulation and different performance metrics like total aircraft delay, on-time performance and passenger disruption metrics.

Ahmedbeygi et al. [9] focused on re-distributing the existing slack in the system by re-timing flights within given time window while aircraft and crew assignments are fixed. In this way, they aimed to reduce the downstream effect of delay. They defined a surrogate objective function which is an approximation to delay propagation and formulated an mixed-integer programming model. The constraint matrix of the model is totally unimodular, hence it can be solved as a linear programming problem.

Chiraphadnakul and Barnhart [10] proposed a model that re-allocates the existing slack by re-timing departure times of flights and adjusting the flight block times. They defined several different measures like passenger delay and delay propagation to measure the robustness of their schedule in different terms. Moreover, they considered different delay scenarios while re-timing the flight departure times. In the scenario generation part, they took real demand values of sixty days in January and February. The matrix of their model is totally unimodular and they can solve it as a linear programming problem. They claimed that even little adjustments in departure times lead to substantial improvement in performance metrics.

Aktürk et al. [11] focused airline recovery. For maintaining disrupted schedules, they considered two options speeding up aircraft and aircraft swaps. The trade-off between flight delays and cost of recovery are taken into account. Moreover, the nonlinear fuel cost function is handled by cone programming. This is the first study which incorporates the cruise speed control in airline recovery model. In addition, nonlinear delay cost function, delay function in the step form and match-up model are the extensions considered in this study.

Duran et al. [12] studied re-scheduling flights within a given time window while ensuring passenger service levels with chance constraints. They assumed that cruise times of flights are controllable and compressing them to some extend is allowable. They considered flight duration as a decision variable. Moreover, their model considers the trade-off between the speeding up the aircraft and putting idle time between flights. Şafak et al. worked on an extension of this problem. They integrated the model proposed in [12] with fleet assignment decision. To achieve robustness in fleet assignment, they considered the fuel efficiency, idle time cost and capacity of aircraft while making an assignment decision.

Sohoni et al. [13] developed stochastic binary integer programming model for incorporating block-time uncertainty. They included block time uncertainty through chance constraints in the model. They considered two different objectives. In the first one, they defined two different target service levels, while satisfying these target levels, their objective is profit maximization. Another variant they considered is maximizing the service level while desired profitability level is reached. The model is solved by cut generation after the chance constraints are linearized.

Dunbar et al. [14] considered aircraft and crew routing problem simultaneously to minimize the cost of the propagated delay. Since both problems are individually NP-hard, they developed algorithms to bring solution to this combined problem. Ageeva et al. [15] worked on airline recovery model, they aimed to increase flight swap opportunities in order to minimize the propagated delay after delay is observed.

Arıkan and Desphande [2] showed that how on time performance of flights are affected from the scheduled block time. Structural estimation technique is used. They showed that the block time devoted to flight is closely related to the definition of delay and the block time that airlines assigned to flights is usually less than the expected flight duration. Moreover, they emphasized that increase in number of passengers and passenger connections do not improve the on-time arrival probability. It was concluded that new definitions for flight delays should be adopted in order to increase on-time arrival probabilities of airlines.

2.2 Stochastic Programming and Scenario Generation

Mathematical modeling of the systems has been widely studied topic for many years. The classical approach considers parameters of the models as deterministic. However, in the real world some parameters are not completely known when some decisions are need to be taken. The traditional method is using the expected value of random parameters. Although it might provide an approximation, this may lead to inferior solutions. One way to handle the uncertainty is using stochastic programming.

Stochastic programming, as a widely used approach for modeling optimization problems that involve uncertainty, tries to take advantage of the fact that probability distributions of governing data are known or can be estimated. The goal of stochastic programming is finding a policy that is feasible for almost all of the possible parameter realizations and optimizes the expectation of objective function. The most widely used stochastic programming formulation is two-stage model. In that model, a number of decisions are taken before the realization of random parameters when the decision maker does not have full information on the random event. These decisions are called as first-stage decisions. After the realization of the random parameters, corrective actions are taken; these actions are referred as second stage decision.The second-stage decisions are recourse decisions which are taken in order to mitigate the possible bad effects that might occur as a result of first stage decisions. In multi-stage stochastic programming, decisions are taken in a sequential order and it can be viewed as an extension of two-stage stochastic programming problem [16].

2.2.1 Two Stage Stochastic Programming Model

In this section, general formulation of two-stage stochastic programming is demonstrated. First stage decisions are represented by vector x and the realized random vector is denoted by ξ. After the realization of random parameters, second stage decisions, or by their other name corrective actions, y are taken. In mathematical programming, the two-stage stochastic programming is generally represented in the following form

$$\begin{aligned} \text{Minimize } & c^T x + E_\xi Q(x,\xi) \\ \text{subject to } & Ax = b \\ & x \geq 0 \end{aligned}$$

where $Q(x,\xi)$ represents the recourse function of the second stage for given first stage decision vector x and realization ξ.

$$Q(x,\xi) = min\{q^T y | Wy = h - Tx, y \geq 0\}$$

when W does not change according to realization of ξ the model is called as fixed recourse model. $Q(x,\xi)$ is called as recourse function. More detailed information about stochastic programming can be found in Shapiro [17] and Birge et al. [18].

2.2.2 Scenario Generation and Scenarios Generation Methods in Airlines

Stochastic programming can only handle discrete samples of limited size, hence the discrete approximations of continuous distributions should be used. For computational tractability number of scenarios should be limited, however theoretically reasonable accuracy is desired. The main problem with the scenarios is the exponential growth of number of scenarios. Hence, increasing accuracy of approximation and computational tractability of problem are two conflicting objectives [16].

There are several sources of scenarios. Historical data, experts' opinion, simulation based on a mathematical model or a combination of methods can be used for scenario generation purposes. Actually, scenarios are not natural part of the problem; they are a result of the methodology that is adopted to solve problems. A good scenario generation method should influence the solution only as little as possible and the scenario-based solution should converge to the true optima, with increasing number of scenarios. However, a good scenario generation method is problem-dependent and bad methods might spoil the result of the whole optimization [19].

Scenarios that consider demand of flights:
One of the major applications of revenue management is improving profits by controlling the prices and availabilities of various products that are produced with scarce resources. Airline industry exemplifies one of the best practices of this area. In airline industry, tickets can be considered as products and seats on flights refer to scarce resources. As in every industry, demand distribution plays a key factor on revenue management problem. In general, separate demands for individual itinerary-class pairs are taken into account since each itinerary and class produce different revenue. Also, each class has specific behaviors and different price sensitivity. Moreover, predicting demand for a flight is a critical step for determining fleet assignment and fleet composition. Seat allocation problems are modeled as linear programming models and expectations of demand distributions are used. Even though this approach eases the computation process, it does not allow user to utilize more information from demand distribution that might reveal as time passes. A proposed approach to utilize demand distribution more is re-solving the deterministic linear programming model repeatedly when new information is revealed [20]. However, Cooper [19] showed that re-solving deterministic problem repeatedly might result in lower expected revenue.

Chen and Homem-de-Mello [20] focused solving origin-destination model in airline revenue management. As stated above, each itinerary-class fare has different demand distribution. The seat allocation process might be considered as sequential decision process that involves rejection or acceptance of each demand request and can be modeled by using multistage stochastic programming (MSSP).

However, problem tractability becomes an issue as the number of stages, demand classes, flights and scenarios increase. Instead of solving a single MSSP, they proposed solving a sequence of two stage problems with a simple recourse model. Even though this procedure might deteriorate the solution quality, it can be considered as a good approximation. They do not specify the method they use for scenario generation.

Another paper related with revenue management in airline operations is written by Möller et al. [21] In this paper, authors worked on determining protection levels for origin-destination revenue management problem. The stochastic values in their model are demand and cancellation values in each stage. They divided time horizon into data collection points (DCP). They modeled the booking problem as a linear programming model, hence they utilized computational efficiency. Their decision variables are protection levels for each fare class, itinerary, and at each DCP and their objective is maximizing total revenue by considering booking request and cancellations. During scenario generation steps, they utilized from the method proposed by Gröwe and Kuska [22] which is based on computation of Kantrovich distances. This scenario generation algorithm has several advantages. First, it does not require any assumption on underlying demand distribution. As a result of this benefit, the authors utilized from the historical discrete data and developed a fan shaped tree which later turned into a scenario tree with the algorithm.

In the later paper written by the same authors [23], a model for airline revenue network management was presented. They modeled this problem as a mixed integer programming model. Their scenario generation algorithm relies on the algorithm proposed by Heitsch and Römisch [24] which is a stability-based recursive reduction and bundling technique which allows to handle multi-dimensional and multivariate stochastic processes.

Lardeux et al. [25] also focused revenue management in airlines. They proposed a method for solving the availability calculation for itineraries in real time while considering uncertainty. The availability calculation means determining whether a ticket for an itinerary is available or not in a given period. The decision variables are continuous and they represent the number of seats in the whole network allocated to each product. Their objective is minimizing maximum regret. In scenario generation step, they generated scenarios by considering the remaining demand for each itinerary and fare class. Even though they assumed that demand follows a normal distribution, they did not specify a method that is adopted when selection of discrete points is handled.

Listes and Dekker [26] studied on creating an approach to the airline fleet composition problem that accounts explicitly the stochastic demand fluctuations. The authors proposed a mixed integer multi-commodity flow model in order to decide robust fleet composition under stochastic passenger demand. In their stochastic programming model (SP), they considered the number of aircraft of each type as their first stage decision variable and assignment of aircraft to flights and their positioning on ground arcs are second stage decision variables. Even the deterministic version of this problem is NP-hard for more than three aircraft types. Hence, they developed an approximation algorithm to find a robust solution for fleet composition problem. They generated scenarios using descriptive sampling and all scenarios has equal probability.

Scenarios that consider the delay of aircrafts or capacity of airports: Delay and capacity scenarios generally consider the operation level at the planning stage and aim to minimize delays. The estimated cost of airport congestion to the industry is $31.2 billion in 2007. Moreover, additional time cost of airlines and passengers is approximately $6 billion [6]. In addition, the cost of delay can be examined in several component and additional crew cost is one of them. Crew cost is the second major cost components of airlines after fuel cost [27]. Hence, developing an effective mechanism that focuses on minimizing the crew cost by considering delays is a research topic. Yen and Birge [28] focused on this problem. They devised a model that incorporates effect of random disruptions in the operational level into the crew assignment decision. They proposed a standard two-stage model where the first stage decision variable is crew assignment and second stage decision variables are actual arrival and departure time of flights under different scenarios. However, this assumption is not realistic because one pairing decision might affect other flights' delay. Hence, they consider this interaction in the second stage recourse problem. In recourse problem, they separated delays such as the delays that are caused by aircraft connections and delays caused by crew connections. Objective function of this problem considers cost of delay that is result of the crew assignment decisions. They developed a branch and bound algorithm called "flight-pair branching algorithm" which uses a variation of constraint branching. In their scenario generation part, they used data of Air New Zeland, they generated 100 scenarios from truncated gamma or log normal distribution by matching mean, second moment and range of disruption data.

Yan et al. [29] studied gate reassignment models by considering random departure times. They differentiated flights as deterministic flights and stochastic flights. The deterministic flights have certain departure and arrival times, which are the flights within one hour interval according to the model. The assignment of the deterministic flights can be regarded as the first stage decision variable and the assignment of the stochastic flights for each possible scenario realization is second stage decisions. The second stage variables are temporary and do not have permanent effect on the final solution. Their final assignment is determined when the flights become deterministic. At this stage, they are beneficial for considering effect of assignment of deterministic flights on the stochastic flights; hence, downstream effect is considered. They solved the reassignment problem based on the recent updates of the flight data once in every 30 minutes. The assignment problem with perfect information about the arrival and departure times of flights is solved in order to find a lower bound to compare the stochastic solutions. After testing different number of scenarios, they realized that when less than 40 scenarios are considered, there are deviations among objective function values. However, when the number of scenarios is larger than 40, incorporating additional ones do not have substantial impact on the result. The objective function value varies less than 3%. For scenario generation purposes, they determined departure/arrival distribution for each flight and they selected random values from this distribution for each flight and combined them to obtain a scenario.

Ball [30] defines ground delay program (GDP) as a mechanism used to decrease the rate of in-coming flights into an airport when it is projected that arrival demand will exceed capacity. Ground delay is the action of delaying take-off beyond a flights schedule departure time. However, this procedure might result in unnecessary ground delays if the capacity forecasts prove to be pessimistic. Mukherjee and Hansen [31] worked on developing a dynamic stochastic integer programming model for single airport ground holding program. They developed a dynamic stochastic optimization model that assign ground delay to individual flights and allows revision according to the most recent updates. In their scenario tree, they considered the airport arrival capacity as time passes and each branch represents a capacity scenario as the day progresses.

2.3 Second Order Cone Programming

In this study, nonlinear cost function is handled by transforming it to second order conic inequalities. This method enables us to obtain exact solutions to the problem instead of approximations. Detailed information about second order cone programming can be found in Ben-Tal and Nemirovski [32] and Günlük and Linderoth [33].

The applications of second order cone programming are presented in many studies. Aktürk et al. [34] worked on conic quadratic reformulations to solve machine job assignment problem with separable cost function. Moreover, Duran et al. [12] and Şafak et al. [35], Aktürk et al. [11] worked on conic reformulations of chance constraints and nonlinear cost functions.

2.4 Summary

Robust optimization is one of the methods that is adopted to build schedules that are less susceptible to unexpected flight delays. Even though the topic is widely studied, there is not an exact definition of robustness. The passenger service levels, total delay, total propagated delay in the system or total operating cost are some of the criteria to measure the robustness of given schedule. Moreover, in the literature in order to obtain more robust schedule fleet re-assignment, schedule re-timing, slack re-allocation and crew assignment are considered.

Moreover, one of the other methods to handle uncertainties in the system is using stochastic programming. In real life not all problem parameters are deterministic and values of random parameters reveal as time passes. However, some decisions need to be taken before the realization of random parameters. Stochastic programming assumes that the possible realizations, which are called as scenarios, can be estimated in advance and the decisions can be taken by considering these scenarios. The solution quality of stochastic programming depends on scenarios, hence the scenario generation method is critical.

In the airline planning problems, usage of stochastic programming techniques is relatively new. The most widely used application of stochastic programming in airline industry is on revenue management. Even though there are several papers in other areas of airline planning process focus on stochastic programming techniques, scenario generation step is usually overlooked. In addition, application of stochastic programming on delay disruption remains limited. In our work scenarios are generated for non-cruise time of flights by examining each part of non-cruise time separately for each airport.

Chapter 3

Problem Definition and Stochastic Model Formulation

The proposed model is a nonlinear two-stage stochastic programming model which is referred as stochastic model in the rest of this thesis. Stochastic model takes non-cruise time scenarios, departure and arrival times of the flights determined by the airline, passenger and aircraft connections as input and generates a robust schedule that is less susceptible to unexpected flight delays by re-timing the departure time of flights. The objective is to minimize expected cost of fuel consumption, CO_2 emission, idle times, passenger disruption and delays experienced by passengers. The stochastic model determines the new published schedule in the first stage and in the second stage the actual departure times of the flights are selected by considering the non-cruise time information and adjusting cruise time under each scenario.

In a schedule, the time between departure and arrival of an aircraft is called as block time. The block time is separated into two parts, cruise time and non-cruise time. Cruise time is considered as controllable in the model since changing the speed of an aircraft within some limits is an option. Controllable cruise time, idle time insertion and experiencing delay option are considered in each scenario for each flight to adjust the actual departure times. Moreover, considering these options and their results in terms of cost have effect on the first stage decision which is published departure times of flights. Non-cruise times of the flights involve departure delay, taxi-out, arrival delay taxi-out stages which can be shorter or longer depending on the congestion of the origin and destination airports, weather conditions, and security issues.

An aircraft connection is possible between flights A1 and A2, if sum of the arrival time of A1 and required turnaround time for the aircraft A1 is less than the departure time of A2 and the origin airport of A2 is the same as the destination airport of A1. Moreover, passenger connection from flight A3 to A4 is possible if the sum of the published arrival time of the flight A3 and passenger turnaround time is less than the departure time of the flight A4 and destination of the flight A3 is same with the origin of A4. While developing a robust schedule, the existing passengers connections are satisfied by adding constraint to the model.

The parameters and decision variables that are used in the proposed model is given below. In the model, J represents the set of flights and J_o represents the first flight of an aircraft in a given day. Set of all non-cruise time scenarios is denoted by Ω.

A is the set of flights connected with the same aircraft. For each $(i, j) \in A$, ta_{ij} is the turnaround time needed to prepare the aircraft after flight i to its next flight j. It depends on the congestion of the destination airport of flight $i \in J$.

P_i is the set of flights which are next flights in the itinerary of the passengers of i. tp_{ij} represents the turn-time needed for the passenger connection between flights i and j. A passenger is considered as disrupted if the connection time is insufficient between his/her current and next flight. Cost of disruption per passenger is denoted by c_m.

For each flight idle time cost is denoted by c_i^s where $s \in J$. The cost of fuel consumption is calculated by multiplying the amount of fuel consumed (in kg) with the fuel price (\$\kg) represented by c_f. The cost of emission is also calculated by multiplying the amount of CO_2 emission (kg) with the unit cost of emission (\$), c_c. Late landing of an aircraft to its destination causes loss of goodwill of passengers and the cost of late arrival is c_d per passenger for each minute.

nd_i^ω represents the non-cruise time of flight $i \in J$ under scenario $\omega \in \Omega$ and p_ω is the probability that scenario ω would be realized. The sum of the p_ω over all $\omega \in \Omega$ is equal to one.

dur_i is the ideal duration of flight $i \in J$. $[f_i^l, f_i^u]$ is the time window for the cruise time of flight $i \in J$, where f_i^l is determined by the maximum compression of the initial flight time. $[x_i^l, x_i^u]$ is the time interval of the possible departure time of flights due to the marketing requirements.

The first stage decision variable of stochastic model is the published departure time of each flight $i \in J$ which is denoted by x_i. The remaining ones are the second stage decision variables, hence they are scenario specific. Actual departure time of the flight $i \in J$ under scenario ω is represented by y_i^ω. Under each scenario ω, s_i^ω indicates idle time after flight i, d_i^ω represents the delay and f_i^ω is the cruise time of flight i. For each passenger connection between two flights, a binary variable z_{ij}^ω is included in the model. When the passengers who are connecting from flight i miss their next flight j this variable is equal to one, otherwise it is zero.

The notation is given below:

Parameters

J: set of all flight legs

J_o : set of first flight leg of each aircraft

Ω: set of possible delay scenarios

P_i: set of flights that have passenger connection with flight $i \in J$

A: set of consecutive flights of the same aircraft

ta_{ij}: turntime of an aircraft between flights i, $j \in J$

tp_{ij}: turntime of passengers between flights $i \in J$, , $j \in P_i$

c_f: cost of fuel ($/ kg)

c_c: cost of CO_2 emission ($/ kg)

c_i^s: unit idle time cost of flight $s \in J$($)

c_d: unit delay cost of a passenger ($)

c_m: cost of passenger disruption ($)

k: CO_2 emission constant

nd_i^ω: Noncruise time delay for flight $i \in J$, in scenario $\omega \in \Omega$

p_ω: Probability of scenario $\omega \in \Omega$

dur_i: published duration of $i \in J$

$[f_i^l, f_i^u]$: time window for cruise time of flight $i \in J$

$[x_i^l, x_i^u]$: time window for departure time of flight $i \in J$

Decision Variables

x_i : departure time of flight $i \in J$

d_i^ω : delay of flight $i \in J$ for a given scenario $\omega \in \Omega$

y_i^ω : actual departure time of flight $i \in J$ for a given scenario $\omega \in \Omega$

s_i^ω : idle time after flight $i \in J$ for a given scenario $\omega \in \Omega$

f_i^ω : cruise time of flight $i \in J$ for a given scenario $\omega \in \Omega$

z_{ij}^ω: 1 if passengers in flight $i \in J$ miss flight $j \in P_i$ for a given scenario $\omega \in \Omega$, 0 o.w.

The calculation of costs which are used in the objective function is shown below. For each flight $i \in J$ and scenario $\omega \in \Omega$, fuel and CO_2 emission cost function is defined by the functions below. These costs components, $c_1^i, c_2^i, c_3^i, c_4^i$ are aircraft specific and detailed information about these cost can be found in

Aktürk et al. [11].

$$C^{i,\omega}_{fuel}(f^\omega_i) = c_f \cdot \left(c^i_1 \frac{1}{f^\omega_i} + c^i_2 \frac{1}{(f^\omega_i)^2} + c^i_3 (f^\omega_i)^3 + c^i_4 (f^\omega_i)^2 \right)$$

$$C^{i,\omega}_{CO_2}(f^\omega_i) = c_c \cdot k \cdot \left(c^i_1 \frac{1}{f^\omega_i} + c^i_2 \frac{1}{(f^\omega_i)^2} + c^i_3 (f^\omega_i)^3 + c^i_4 (f^\omega_i)^2 \right).$$

The idle time cost, delay cost and disruption cost is calculated as follows:

$$C^{i,\omega}_{idle}(s^\omega_i) = c_i \cdot s^\omega_i$$

$$C^{i,\omega}_{del}(d^\omega_i) = c_d \cdot numpas_i \cdot d^\omega_i$$

$$C^{ik,\omega}_{disrupt}(z^\omega_{ik}) = c_m \cdot PAS_{ik} \cdot z^\omega_{ik}$$

where

$numpas_i$: number of passengers in flight $i \in J$

PAS_{ij}: number of passengers connecting from flight $i \in J$ to flight $j \in P_i$.

3.1 Mathematical Model

As indicated in the previous sections, assigned block time of the flights have great impact on the operating cost of the airlines. Increasing the block time of the flights might decrease the delays when some scenarios realize, on the other hand it may cause unnecessary idle time in other cases. When idle time cost is relatively greater than delay and passenger disruption cost, the block times of the flights tend to decrease. The published schedule is affected by the variability of scenarios as well as their probabilities.

Moreover, compensating insufficient idle time and preventing the delays by increasing the speed of the flight is another option. Controlling the speed of the aircraft is preferable to the idle time insertion, since inserting idle time between the flights is an irreversible decision. However, speeding up the aircraft in case of delay and congestion, or allowing delay in some scenarios can be less costly than increasing the block times of the flights in the published schedule which affects all the scenarios and may cause long idle times under some scenarios. Experiencing delay and passenger disruption might be a better option if the fuel cost and idle time cost are high. Therefore, considering the trade-off between these three options under each scenario and the relative weights of scenarios is a complex problem with too many variables and parameters. Solving this problem even on a small network with a few scenarios requires a global optimization tool.

Therefore, we developed a mathematical model that tries to minimize the expected cost and determines the published schedule at the first stage and actual schedule, which are recourse decisions, for each scenario at the second stage.

The mathematical model is given below. The first part corresponds to the first stage of the model where the vector x denotes the published departure times of the flights.

$$\text{minimize } E(Q(x)) \tag{3.1}$$

$$\text{subject to } x_k - (x_i + dur_i + tp_{ik}) \geq 0 \qquad i \in J, \quad k \in P_i \tag{3.2}$$

$$x_i^l \leq x_i \leq x_i^u \qquad i \in J \tag{3.3}$$

where the recourse function $E(Q(x)) = \sum_{\omega \in \Omega} Q_\omega(x) p_\omega$ is defined as

$$Q_\omega(x) = \min \sum_{i \in J} (C_{idle}^{i,\omega}(s_i^\omega) + C_{fuel}^{i,\omega}(f_i^\omega) + C_{CO_2}^{i,\omega}(f_i^\omega) + C_{del}^{i,\omega}(d_i^\omega) + \sum_{k \in P_i} C_{disrupt}^{ik,\omega}(z_{ik}^\omega))$$

subject to

$$
\begin{aligned}
y_j^w - y_i^\omega - ta_{ij} - f_i^\omega - nd_i^\omega - s_i^\omega \quad &= 0 \quad && i \in J, (i,j) \in A, \omega \in \Omega && (3.4) \\
y_i^\omega \quad &\geq x_i \quad && i \in J, \omega \in \Omega && (3.5) \\
y_i^\omega \quad &= x_i \quad && i \in J_o, \omega \in \Omega && (3.6) \\
y_i^\omega + tp_{ik} + f_i^\omega + nd_i^\omega - y_k^w \quad &\leq M \times z_{ik}^\omega \quad && i \in J, k \in P_i, \omega \in \Omega && (3.7) \\
f_i^l \leq f_i^\omega \quad &\leq f_i^u \quad && i \in J, \omega \in \Omega && (3.8) \\
(y_i^\omega + f_i^\omega + nd_i^\omega) - (x_i + dur_i) \quad &\leq d_i^\omega \quad && i \in J, \omega \in \Omega && (3.9) \\
0 \quad &\leq d_i^\omega \quad && i \in J, \omega \in \Omega && (3.10) \\
0 \quad &\leq s_i^\omega \quad && i \in J, \omega \in \Omega && (3.11) \\
z_{ik}^\omega \quad &\in \{0,1\} \quad && i \in J, k \in P_i, \omega \in \Omega && (3.12)
\end{aligned}
$$

The most common objective function is minimizing the operating cost of the airlines and it is represented by fuel cost and idle time cost in our model. Moreover, passenger perspective is also taken into account. Dissatisfaction of the customers would increase the objective function both by cost of goodwill, which is represented by cost of delay, and by cost of disruption which is a result of either finding a new itinerary to the passenger or reimbursing passengers. These costs are scenario specific and their expectation is taken in the objective function.

In (3.3) time frame is put on the published departure time of flights in the new published schedule, hence the departure times of flights in the new published schedule is allowed to deviate from the ones in the published schedule within some limits. This constraint is inserted to protect the current market share of the airline. Constraint (3.2) ensures that if there exists a passenger connection between two flights in the published schedule of airline, this connection should still be satisfied in the new published schedule generated by the model. Hence, existing passenger connections are taken into account while generating the new schedule. Constraint (3.4) guarantees that if two flight legs are assigned to the same aircraft and flight j follows flight i then flight j cannot depart before flight i arrives and the aircraft is prepared for the next flight this time can be called as the ready time of flight j. If the flight j departs later than its ready time then the time between its departure and ready time denotes the idle time of aircraft after flight i. Constraint (3.5) ensures that the actual departure time of a flight cannot be earlier than its scheduled departure time. Constraint (3.6) requires that the first flight of every aircraft should depart on time. When there are connecting passengers from flight i to k and there is not enough time for passenger connection between departure of flight k and arrival of flight i then the passengers miss flight k and in this case constraint (3.7) ensures that z_{ik}^{w} is equal to 1. In (3.8) we put time frame on the cruise time of flights. It can be shorter than or equal to its ideal cruise duration which depends on by max-range cruise speed and the lower bound is determined by the maximum allowable compression amount of ideal cruise time duration. Increase in delay leads to increase in objective function value, hence the model tries to assign it to its lower bound. Constraints (3.10) and (3.9) determine the lower bound of delay of a flight. When a flight arrives late then its delay is set equal to the difference between its actual and published arrival time by constraint (3.9). If it arrives earlier than its published arrival time or on time then delay is set to zero by (3.10).

In order to strengthen the formulation and speed up the solution process a valid inequality is developed. The valid inequality is presented and its validity is proven in the following propositions. l_j^ω denotes the lower bound of the actual departure time of flight j under scenario ω (y_j^ω). This value is set to possible earliest departure time of the flight j which is x_j^l for each $j \in J$.

Proposition 3.1.1. *Let $i \in J$, $j \in P_i$, $\omega \in \Omega$ and l_j^ω,l_i^ω be a lower bound for y_j^ω,l_i^ω respectively. The inequality*

$$l_i^\omega + f_i^l + tp_{ij} + nd_i^\omega \leq (l_i^\omega + f_i^l + tp_{ij} + nd_i^\omega - l_j^\omega)z_{ij}^\omega + y_j^\omega \tag{3.13}$$

is a valid inequality for the feasible set of stochastic model.

Proof. If $z_{ij}^\omega = 0$ then for feasibility we need $y_i^\omega + f_i^\omega + tp_{ij} + nd_i^\omega \leq y_j^\omega$. Since $y_i^\omega \geq l_i^\omega$, inequality (3.13) is satisfied.

If $z_{ij}^\omega = 1$, inequality (3.13) becomes $l_j^\omega \leq y_j^\omega$. Since l_j^ω is a lower bound for y_j^ω, the inequality is again satisfied. □

3.2 Conic Reformulation of the Stochastic Model

The objective function of the model involves non-linearity due to controllable cruise time. Solving nonlinear mixed integer models require excessive computation time and it might not give exact solutions. This non-linear cost function could be handled with second order conic inequalities as demonstrated in Aktürk et al. [11] and Günlük et al. [33]. Providing the solution in this way is computationally tractable and results in exact solutions. In order to simplify the representation, flight and scenario indices of cruise time variable are dropped.

Since the formulations of fuel cost and carbon emission cost functions are similar except the cost multiplier, they are combined into a function as demonstrated below.

$$C^f_{total} = C_{fuel}(f) + C_{CO_2}(f) = (c_f + c_c \cdot k) \cdot \left(c_1 \frac{1}{f} + c_2 \frac{1}{(f)^2} + c_3 (f)^3 + c_4 (f)^2 \right).$$

This nonlinear cost function in the objective is expressed with the constraints in the following form:

$$t \geq (c_{fuel} + k.c_{CO_2})(c_1 \cdot q + c_2 \cdot \delta + c_3 \cdot \varphi + c_4 \cdot \vartheta) \tag{3.14}$$

$$1^2 \leq q \times f \tag{3.15}$$

$$1^4 \leq f^2 \times \delta \times 1 \tag{3.16}$$

$$f^4 \leq 1^2 \times \varphi \times f \tag{3.17}$$

$$f^2 \leq \vartheta \times 1 \tag{3.18}$$

The constraints (3.15), (3.16), (3.17) and (3.18) can also be shown as below:

$$\frac{1^2}{f} \leq q$$
$$\frac{1^4}{f^2} \leq \delta$$
$$\frac{f^3}{1^2} \leq \varphi$$
$$\frac{f^2}{1} \leq \vartheta$$

(3.15) and (3.18) are hyperbolic inequalities whereas (3.16) and (3.17) can be presented as a combination of two hyperbolic inequalities. (3.16) can be represented as

$$1^2 \leq wf \text{ and } w^2 \leq \delta.1$$

and (3.17) can be restated as

$$f^2 \leq w1 \text{ and } w^2 \leq \varphi.f$$

Safak [35] proved that these hyperbolic inequalities can be expressed as second order conic inequalities. Conic reformulation of the cost function is presented below.

For constraint (3.15)**:** Two auxilary variables W_1 and $W_2 \geq 0$ are introduced and denoted as below,

$$W_1 = q_i^\omega - f_i^\omega \qquad \text{i} \in J, \omega \in \Omega (3.19)$$
$$W_2 = q_i^\omega + f_i^\omega \qquad \text{i} \in J, \omega \in \Omega (3.20)$$
$$4 \times 1^2 \leq (W_2)^2 - (W_1)^2 \qquad \text{i} \in J, \omega \in \Omega (3.21)$$

For constraint (3.16)**:**

$$(Q_i^\omega)^2 \leq \delta_i^\omega \times 1 \qquad \text{i} \in J, \omega \in \Omega (3.22)$$
$$1^2 \leq Q_i^\omega \times f_i^\omega \qquad \text{i} \in J, \omega \in \Omega (3.23)$$

Two auxilary variables W_3 and $W_4 \geq 0$ are presented and defined as follows,

$$\begin{aligned} W_3 &= \delta_i^\omega - 1 && \text{i} \in J, \omega \in \Omega (3.24) \\ W_4 &= \delta_i^\omega + 1 && \text{i} \in J, \omega \in \Omega (3.25) \end{aligned}$$

Then, the constraint (3.22), can be rewritten so as,

$$4(Q_i^\omega)^2 \leq (W_4)^2 - (W_3)^2 \qquad i \in J, \omega \in \Omega \qquad (3.26)$$

Two auxilary variables W_5 and $W_6 \geq 0$ are introduced and denoted as below,

$$\begin{aligned} W_5 &= Q_i^\omega - f_i^\omega && \text{i} \in J, \omega \in \Omega (3.27) \\ W_6 &= Q_i^\omega + f_i^\omega && \text{i} \in J, \omega \in \Omega (3.28) \end{aligned}$$

Then, let's rewrite the constraint (3.23) as below,

$$4(1)^2 \leq (W_6)^2 - (W_5)^2 \qquad \text{i} \in J, \omega \in \Omega (3.29)$$

For constraint (3.17)**:** It can be redefined as follows,

$$\begin{aligned} (Q_i^\omega)^2 &\leq \varphi_i^\omega \times f_i^\omega && \text{i} \in J, \omega \in \Omega (3.30) \\ f_i^{\omega 2} &\leq Q_i^\omega \times 1 && \text{i} \in J, \omega \in \Omega (3.31) \end{aligned}$$

Two auxilary variables W_7 and $W_8 \geq 0$ are introduced and denoted as follows,

$$\begin{aligned} W_7 &= \varphi_i^\omega - f_i^\omega && \text{i} \in J, \omega \in \Omega (3.32) \\ W_8 &= \varphi_i^\omega + f_i^\omega && \text{i} \in J, \omega \in \Omega (3.33) \end{aligned}$$

Then, the constraint (3.30), can be rewritten so as,

$$4(Q_i^{\omega})^2 \leq (W_8)^2 - (W_7)^2 \qquad \text{i} \in J, \omega \in \Omega \tag{3.34}$$

Let, introduce two auxilary variables W_9 and $W_{10} \geq 0$ and define them as below,

$$\begin{aligned} W_9 &= Q_i^{\omega} - 1 & \text{i} \in J, \omega \in \Omega & \quad (3.35) \\ W_{10} &= Q_i^{\omega} + 1 & \text{i} \in J, \omega \in \Omega & \quad (3.36) \end{aligned}$$

Then, the constraint (3.31), can be rewritten so as,

$$4(f_i^{\omega})^2 \leq (W_{10})^2 - (W_9)^2 \qquad \text{i} \in J, \omega \in \Omega \tag{3.37}$$

For constraint (3.18)**:** Two auxilary variables W_{11} and $W_{12} \geq 0$ are introduced and denoted as follows,

$$\begin{aligned} W_{11} &= \vartheta_i^{\omega} - 1 & \text{i} \in J, \omega \in \Omega & \quad (3.38) \\ W_{12} &= \vartheta_i^{\omega} + 1 & \text{i} \in J, \omega \in \Omega & \quad (3.39) \\ 4(f_i^{\omega})^2 &\leq (W_{12})^2 - (W_{11})^2 & \text{i} \in J, \omega \in \Omega & \quad (3.40) \end{aligned}$$

3.2.1 Reformulated Stochastic Model

When the non-linear cost function is expressed with second order conic inequalities and a valid inequality are introduced into model, the model becomes:

$$\min \sum_{\omega \in \Omega} p_{\omega} \sum_{i \in J} \quad (C_{idle}^{i,\omega}(s_i^{\omega}) + (c_f + c_c \times k) \times (c_1^i \times q_i^w + c_2^i \times \delta_i^w + c_3^i \times \varphi_i^w + c_4^i \times \vartheta_i^w)$$
$$+ C_{del}^{i,\omega}(d_i^{\omega}) + \sum_{k \in P_i} C_{disrupt}^{ik,\omega}(z_{ik}^{\omega}))$$

subject to

$$
\begin{aligned}
q_i^\omega \times f_i^\omega \geq & \quad 1 & i \in J, \omega \in \Omega \quad (3.41)\\
\delta_i^\omega \times (f_i^\omega)^2 \times 1 \geq & \quad 1 & i \in J, \omega \in \Omega \quad (3.42)\\
\vartheta_i^\omega \times 1 \geq & \quad (f_i^\omega)^2 & i \in J, \omega \in \Omega \quad (3.43)\\
\varphi_i^\omega \times f_i^w \times 1^2 \geq & \quad (f_i^w)^4 & i \in J, w \in \Omega \quad (3.44)\\
l_i^\omega + f_i^l + tp_{ij} + nd_i^\omega \leq & \quad (l_i^\omega + f_i^l + tp_{ij} + nd_i^\omega - l_j^\omega) z_{ij}^\omega + y_j^\omega & i \in J, j \in P_i, \omega \in \Omega \quad (3.45)
\end{aligned}
$$

(3.4)-(3.12)

The cruise and CO_2 emission cost components are changed since the non-linear cost function is represented with conic constraints (3.41)-(3.44). The constraints (3.45) represent the proposed valid inequality. Remaining constraints are same with the proposed model in the previous section.

3.3 Summary

In this section definition of the problem, which is explained briefly in the first section, is extended. The parameters and decision variables used in the model are explained in detail. The stochastic model formulation to solve this nonlinear two stage stochastic programming problem is provided. A valid inequality is introduced to decrease the CPU time of the stochastic model. Moreover, conic reformulation of the nonlinear cost function is demonstrated. Finally, the extended model is presented.

Chapter 4

Scenario Generation

In stochastic programming problems, the underlying distributions of random variables are assumed to be known but these are usually continuous distributions. However, due to the limited computing power they should be reduced to limited number of discrete points. Hence, the problems cannot be solved for exact values. Since the approximations of problems are solved, the solution is substantially affected from the discrete approximation of stochastic variables which are called as scenarios. Therefore, adopted scenario generation methodology is one of the most important factors that determine the quality of the solution of stochastic programming.

4.1 Airport Classification and Discrete Point Selection

Data used in this study are obtained from the website of Bureau of Transportation Statistics (BTS). The governing non-cruise time data are known for each flight and each airport. Although non-cruise time values are flight specific, the flights that depart from and land to the same airport are subject to same weather and congestion conditions. Hence, determining taxi-out, departure delay, arrival delay and taxi-in times of airports instead of flights during the scenario generation process would not deteriorate solution quality. Moreover, number of scenarios increases exponentially. In our computational experiments, we work on two different networks. In the first case, which is referred as network 1, the number of airports is 10 and the number of flights is 31. If three data points are generated for each airport and all possible combinations are considered then the number of scenarios is equal to 3^{10}. On the other hand, if the flights are used then the number of scenarios would be 3^{31}. Thus, using airports instead of flights would substantially reduce number of scenarios while the solution quality is not affected substantially, because the flights that depart from or land to same airport are subject to similar conditions which are the effective on non-cruise time of flights. For our problem, instead of picking the values from the distribution, they are selected from historical data by using conditional sampling. This method is preferable, since number of data points that are used in our study for each airport is at most three and size of the data is very large. Hence, instead of fitting a distribution to data, one point is selected to represent the average values of parameters, the other one indicates values of non-cruise times for delayed flights, the last one is selected by examining the historical data and picking the point that represents the remaining probability when probabilities of the first two data points are extracted. The method that is used for determining the value of each parameter and classifying airports is explained in detail in the following paragraphs.

Ω represents the set of non-cruise time scenarios in our model. Each non-cruise time scenario is composed of departure delay, taxi-out time, arrival delay and taxi-in time of each airport. The non-cruise time of a flight is equal to sum of departure delay, taxi-out time of its origin airport and arrival delay and taxi-in time of destination airport. This sum is represented by nd_i^w where $i \in J$, $w \in \Omega$. Each scenario has a probability which is denoted by p_w for $w \in \Omega$. The sum of p_w over $w \in \Omega$ is equal to one.

The calculation of non-cruise time of a flight is demonstrated on an example. The flight with tail number N535AA and flight number 1446 departs from ORD and lands to EWR. While its taxi-out time and departure delay depend on the congestion of ORD, taxi-in time and arrival delay are determined by the congestion of EWR. The departure delay, taxi-out, arrival delay, taxi-in times of ORD are equal to 15, 11, 9 and 3 and of EWR are 22, 10, 7 and 9, respectively. These values are belong to scenario which is referred as ω_1. Then the non-cruise time of flight 1446 is equal to 42 as indicated below.

$$
\begin{aligned}
nd^{\omega_1}_{1446} &= \text{taxi-out}_{ORD} + \text{dep.del.}_{ORD} + \text{taxi-in}_{EWR} + \text{arr. del.}_{EWR} \\
&= 15 + 11 + 7 + 9 = 42 \text{ min}
\end{aligned}
$$

BTS publishes monthly reports that present on-time performance of airlines and airports in that specific month. They classify some airports as major airports and they are ranked according to their on-time arrival and departure percentages. While generating scenarios the major airports are considered as priority since more passengers and flights are affected from the congestion at those airports. The major airports are classified according to their on-time and departure probability.

According to number of scenarios that will be considered in the stochastic model, number of data points for each airport is determined. For example, when 18 scenarios would be included for network 1, three data points are selected for the two most crowded airports which are ORD and EWR. For the remaining airports two instances are created. In the first instance, the data point which has the smallest non-cruise time value among the three data points is selected for each of the remaining airports and this case is referred as optimistic case. In the second instance, the data point which has the largest non-cruise time value among the three data points is selected for each of the remaining airports and this case is referred as pessimistic case. In order to capture the variability of the remaining airports, these optimistic and pessimistic cases are also included in the scenario generation process. Hence, in total two different values are used for the remaining airports. All possible combinations of data points, three points for ORD and EWR and two instances for the remaining airports, are considered hence, the number of scenarios that would be used for solving stochastic model on network 1 is $3 \times 3 \times 2 = 18$. When the number of scenarios that would be considered is equal to 288, three data points for ORD and EWR, two data points for DCA, LAS, MIA, MCO are selected. For the remaining airports two data points are included, as explained for the 18 scenario case, one of them is all optimistic case and the other is all pessimistic case. When all possible combinations are considered 288 scenarios are generated.

As indicated above, the data for the taxi-out, departure delay, taxi-in and arrival delay information are obtained from the website of BTS. Three data points are generated for each airport. These points represent optimistic, most likely and pessimistic non-cruise time realizations. Since solving the stochastic model becomes computationally untractable as number of scenarios increases, not all of the data points are used in scenario generation. However, to have more accurate view about the non-cruise time variability of airports and assign the probabilities of data points more precisely by taking the other non-cruise time realizations into account three data points are generated for each airport.

In the BTS' website there are options to examine the each component of non-cruise time in detail or to obtain summary tables that show the average values taken over a specified time span of the desired components. The first data point indicates the average non-cruise times of flights that depart from or land to the specified airport between 2013-2014. Departure delay and taxi-out times are related with origin airport of the flight while the taxi-in time and arrival delay are dependent on the destination airport. The snapshots in figures 4.1 (http://apps.bts.gov/xml/ontimesummarystatistics/src/ddisp /OntimeSummarySelect.xml?tname=OntimeSummaryOrigData) and 4.2 (http://apps.bts.gov/xml/ontimesummarystatistics/src/ddisp /OntimeSummarySelect.xml?tname=OntimeSummaryDestData), that demonstrate these values for Chicago O'Hare airport, are taken from BTS' website. Origin Airport indicates the name of the airport, airline shows that which airline's flight data are used to generate these averages. Time period specifies the time span during which the data is obtained. The first row of the table, included in the snapshots, shows the average values for all airlines and second row presents average values for the airline company, which is referred as AA in the rest of the thesis, whose flight data is used for generation on network 1 and 2. We use the airline specific information as shown in the second row of the figures 4.1 and 4.2 since airlines are assigned specific gates and their taxi-out, departure delay performances might be affected from the position of the gate that they use. The first column indicates total number of flights of AA that depart from Chicago O'Hare during the specified time span. Second and third columns demonstrate average departure delay and average taxi-out minutes, respectively. These are fractional numbers but for convenience they are rounded to the closest integer. Hence, for the first data point of Chicago O'Hare departure delay value is equal to 13 and taxi-out time is equal to 15. In figure 4.2, Destination Airport represents the name of the airport, Airline and Time Period are same with figure 4.1. The first column indicates the total number of flights of AA that land to Chicago O'Hare during the time period. Second, third and fourth columns demonstrate average arrival delay, airborne delay and average taxi-out minutes, respectively. Airborne delay is not included in our study, since airborne delays are extreme cases also they are not a part of the statistics that airlines report to DOT. As it is done for taxi-out time and departure delay, the taxi-in and arrival delay are also rounded to the closest integer. Hence, for the first data point of Chicago O'Hare arrival delay value is equal to 6 and taxi-in time is equal to 9.

Summary Statistics
Origin Airport

Origin Airport: Chicago-Naperville-Joliet, IL-IN-WI - O'Hare International (ORD)
Airline: American Airlines (AA)
Time Period: January 1, 2013 to January 1, 2014

Note: A complete listing of airline and airport abbreviations is available. Times are reported in local time using a 24 hour clock.

Excel | CSV

Carriers	All Flights				Late Flights				Total Number Cancelled	Percent Flights Cancelled	Total Number Diverted	Percent Flights Diverted	Percent Flights Late
	Total Number	Average Departure Delay (minutes)	Average Taxi-Out (minutes)	Average Scheduled Departure to Take-off (minutes)	Total Number	Average Departure Delay (minutes)	Average Taxi-Out (minutes)	Average Scheduled Departure to Take-off (minutes)					
ALL*	307,107	16.09	17.38	33.47	82,272	60.89	19.46	80.35	9,366	3.05	709	0.23	26.79
AA	51,671	13.00	15.26	28.26	12,089	55.56	17.71	73.27	854	1.65	151	0.29	23.40

Figure 4.1: ORD-Origin Airport

Summary Statistics
Destination Airport

Destination Airport: Chicago-Naperville-Joliet, IL-IN-WI - O'Hare International (ORD)
Airline: American Airlines (AA)
Time Period: January 1, 2013 to January 1, 2014

Note: A complete listing of airline and airport abbreviations is available. Times are reported in local time using a 24 hour clock.

Excel | CSV

Carriers	All Flights				Late Flights				Total Number Cancelled	Percent Flights Cancelled	Total Number Diverted	Percent Flights Diverted	Percent Flights Late
	Total Number	Average Arrival Delay (minutes)	Average Airborne Time (minutes)	Average Taxi-In (minutes)	Total Number	Average Arrival Delay (minutes)	Average Airborne Time (minutes)	Average Taxi-In (minutes)					
ALL*	307,205	9.83	103.08	9.91	71,445	71.03	102.15	12.85	10,188	3.32	711	0.23	23.26
AA	51,676	5.70	144.00	9.03	10,475	70.55	144.45	10.32	887	1.72	173	0.33	20.27

Figure 4.2: ORD-Destination Airport

Second data point represents the average values for delayed flights. For this purpose the late flights information in Figure 4.1 and 4.2 are used. In figure 4.1, the fifth, sixth and seventh columns denote total number of flights that depart late, average departure delay and average taxi-out times of these flights respectively. The last column of the table denotes the percentage of the flights that depart late from O'Hare. In figure 4.2, the fifth, sixth and seventh column denotes total number of flights that arrive late, average arrival delay and average taxi-in times of these flights respectively. The last column of the table denotes the percentage of the flights that arrive late to O'Hare.

Third data point is selected by descriptive sampling by considering the data of randomly selected thirty days of 2013. First, the data are sorted in ascending order. Then, a cumulative discrete function is fit. The value that coincides with the probability of third data point is selected. This process is exemplified on taxi-out time value selection of the third data point of ORD. The first data point represents the average of all flights, hence the probability of this point is taken as 0.5. Then the probability of second is taken as 0.23. Thus, the probability of the third data point is equal to 0.27. First data point has the highest probability, therefore it corresponds to most likely situation. Second data point indicates the non-cruise time values of delayed flights, hence it is the pessimistic case. Third data point is selected to represent the optimistic case. The data are ordered in ascending order and the value that is selected from upper part of the data.

Finally, probabilities are assigned to each data point. The probability of the first data point is taken as 0.5. The probability of second data point is computed by computing total number of late departures and late arrivals and dividing this number to the sum of arrivals and departures. For instance, for O'Hare the probability of second data point is 0.23. As mentioned above the third data point is selected according to its probability, since the probability of this point is computed by subtracting sum of probabilities of first and second data points from 1. However, the probabilities of data points are normalized so that their sum adds up to 1. For the airports that include only one data point, the probability of this data point is equal to one.

The chart obtained by from arrival delay data, which is used in the generation of third data point, is given in 4.3. The arrival delay value for third data point is 0. Moreover, when the chart is examined it can be observed that the values selected for the first and the second data points, which are 13 and 56, are also appropriate to represent the most likely and pessimistic cases when their probabilities are taken into account.

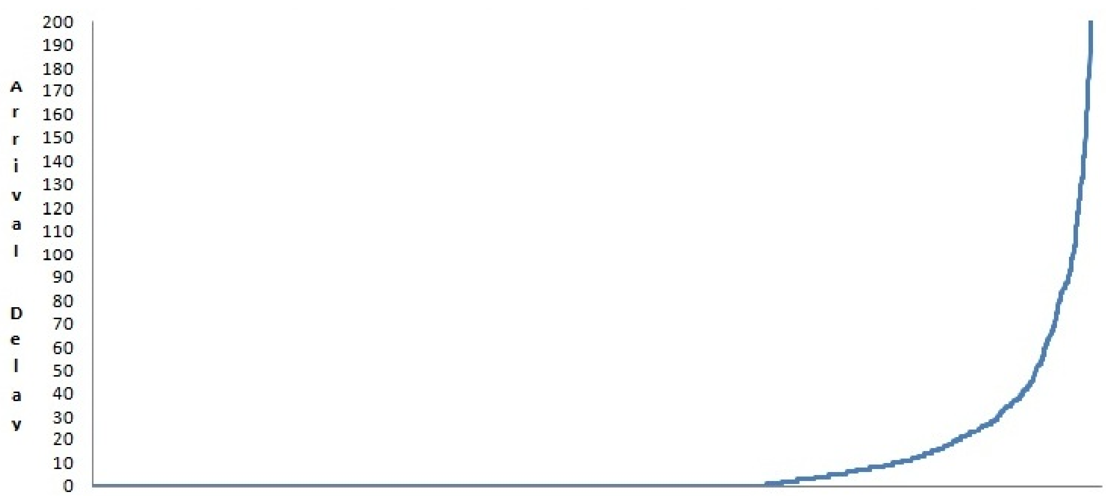

Figure 4.3: Arrival Delay Chart-ORD

A scenario is a vector which includes departure delay, taxi-out, arrival delay, taxi-in time for each airport which is a destination or an origin airport of a flight in the given schedule. It is assumed that non-cruise time values of airports are independent of each other. Hence, the probability of a scenario can be computed by multiplying the probability of selected data point of each airport. When all possible combinations are included in the scenario tree, the final probability of each scenario is equal to this multiplication. However, if this is not the case, the probability of scenarios should be normalized such that sum of probabilities of scenarios is equal to 1.

4.2 Summary

In this study, non-cruise time scenarios are generated using the historical data. The components of non-cruise time, which are taxi-out time, departure delay, taxi-in time and arrival delay, are examined separately. While generating the scenarios, non-cruise time values of airports are used instead of focusing on individual flights. Since the number of airports are significantly less than number of flights, this method leads to substantial decline in number of scenarios while sacrifice from solution quality is not significant when compared to the gains in solution time and computational tractability.

Three data points are generated for each airport to represent the most likely, optimistic and pessimistic cases. The average values are used for most likely case. For pessimistic instance, the average non-cruise time values of delayed flights are considered. For the optimistic case, conditional sampling methods are adopted. Data of randomly selected thirty days of 2013 are examined and value that corresponds to the probability of the selected point is picked as the value of the third data point.

A scenario consists departure delay, taxi-out, arrival delay and taxi-in times for each airport. The non-cruise times of airports are assumed to be independent of each other. Hence, the probability of each scenario is computed by multiplying individual probabilities of data points included in the scenario. If all of the combinations are not considered, the probabilities of scenarios should be normalized.

Chapter 5

Numerical Example

In this chapter, we demonstrate the mechanics of the model on a small network by providing a numerical example. This example would provide a good understanding about how stochastic model works by taking non-cruise time scenarios into account, re-timing the published departure time of the flights, inserting idle times and controlling the speed of the aircraft. The paths that comprise the network considered in this chapter which is used in this numerical example is given in table 5.1 that presents the departure times of flights published by the airline, which is called as initial departure times in the rest of the chapter, and idle time between the flights. In our model, a schedule that is less susceptible to delays is generated by considering different realizations of non-cruise time. In this way the departure times and block times are adjusted so that the expected cost is minimized.

Since our problem is a two-stage stochastic programming problem, there are first stage and second stage decision variables as explained in Chapter 2. The published departure times of the flights, allowed to change within some limits of its initial value, are our first stage decision variables. In the second stage, for each scenario the actual departure and arrival times of the flights, delays, idle time and whether the passenger connection is satisfied or not are decided depending on the published departure times and realization of non-cruise times. The objective is minimizing the expected operating cost of airline and cost of customer dissatisfaction. Hence, in our approach the published flight schedule is generated by considering several realizations of non-cruise time. This is achieved by considering different scenarios as well as their probabilities in the stochastic model. In addition, both in generating the published schedule and determining actual schedule the trade-off between adjusting the speed of the aircraft, idle time insertion and allowing delay options are taken into account.

The small schedule, which will be used in the numerical example, is given in Table 5.1. It includes 2 paths operated by 2 different aircraft. The tail numbers of the aircraft are given in the first column. In the second column, the flight numbers are given. The information about the origin and destination airport of the flights are provided in the following two columns. The last three columns represent the departure times, block times and arrival times to its destination under the published schedule. On the path of the aircraft with tail number N3ETAA there exists two flights with the same flight number. They share the same flight number because there is one or more intermediate airports between the origin and destination airports, it is called a through flight.

Tail #	Flight #	From	To	Dep.Time	Duration	Arr.Time
N535AA	2460	ORD	RSW	06:45	02:45	09:30
	564	RSW	ORD	10:20	03:05	13:25
	1446	ORD	EWR	14:55	02:45	17:40
	1411	EWR	ORD	18:45	02:45	21:30
N3ETAA	1704	ORD	EWR	06:35	02:05	08:40
	1883	EWR	ORD	09:30	02:40	12:10
	810	ORD	DCA	13:10	01:45	14:55
	2013	DCA	ORD	15:45	02:15	18:00
	2013	ORD	LAS	19:00	04:10	23:10

Table 5.1: Published Schedule of Numerical Example

For this small example, considering even four different non-cruise time scenarios would be enough to show the performance of the schedule generated by stochastic model over the initial schedule. The non-cruise time scenarios are demonstrated in table 5.2. In the first column all airports experience the lowest congestion level, hence non-cruise time values are the smallest among all possibilities for each flight. This scenario corresponds to selecting optimistic non-cruise time value for each airport. The scenario in the second column demonstrates the pessimistic case, since all airports experience highest congestion level. In the third and the fourth columns the most likely values of non-cruise times of Chicago O'Hare (ORD) and New Jersey Newark (EWR) airports realize. In the third column the remaining airports experience lowest congestion level, on the other hand in the fourth one highest congestion option realizes for the remaining airports. The last case represents the expected non-cruise time value of each airport taken over the scenarios that are used in stochastic model. The probability of each scenario is denoted in the last row. They are calculated by multiplying the probability of each data point of each airport. For example, for the first scenario the probability of data points that correspond to ORD, EWR, RSW, DCA and LAS are 0.27, 0.29, and 0.5 for the remaining airports respectively. If all of the possible combinations are considered then the probability of the scenario 1 is equal to $0.27 \times 0.29 \times 0.5 \times 0.5 \times 0.5 = 9.78 \times 10^{-3}$. However, only four scenarios are included. Hence, after normalizing the probabilities of scenarios included in numerical example its probability for this problem becomes 0.23.

How airline separates the flight duration into cruise and non-cruise time is not known. Airline companies have tendency to under-estimate the value of non-cruise time in order to obtain higher utilization from aircraft [2]. Hence, we assume 40 minutes of flight duration are devoted to non-cruise time. For example, flight block time of flight 1883 is 2 hours 40 minutes, 40 minutes represent the non-cruise time and the remaining 2 hours are the cruise time.

The values of second stage decision variables when the initial schedule is used as the published schedule and most likely non-cruise time scenario realizes are given in table 5.3. The actual schedules when other scenarios are realized are given in tables A.1, A.2, and A.3 and can be found in appendix. As it is observed, using initial schedule causes long idle times if optimistic scenario realizes. On the other hand if pessimistic scenario actualizes long delays are observed and the delays have cascading effect. The main reason that creates this difference is the variability of scenarios. The initial schedule does not take the variability of the scenarios into account and this causes major differences in delay and idle time values for different realizations.

Even though no passenger connection is demonstrated in this small example, when there is a passenger connection between two flights this connection remains feasible in the schedule generated by stochastic model. In this study, we assume that passenger connection between flights i and j is possible when the initial departure time of the flight j is within 45 and 180 minutes of initial arrival time of flight i and the destination airport of flight i is the same with the origin airport of flight j. The passenger connection for a turn back flight is not allowed.

While arriving late might cause inconvenience to passengers and disruption, keeping such expensive equipments and crew idle would also increase the operating cost. The trade-off between different options should be recognized. A decision maker might create the published schedule by focusing on one scenario depending on his evaluation. Although focusing only on one scenario and generating the published schedule accordingly would produce optimal results in that specific instance, this schedule might perform catastrophically in others when compared with the solution of the stochastic model. Hence, considering all scenarios and their probabilities simultaneously would give the best overall result, even though it might not perform optimally in any one of the scenarios. The performance of the schedule generated by the stochastic model is compared with initial schedule and schedules generated by focusing solely on the individual scenarios which are included in the stochastic model and using the expected values of non-cruise time. Solving the problem by focusing one scenario is same with the reducing the problem into deterministic case. The first stage decision variables are generated such that the cost is minimized when the considered scenario realized. Since the first stage decision variables are determined, the problem is decomposed into solving the second stage of the stochastic model for every scenario that is considered in the stochastic model. The second stage problems are solved to optimality by taking these fixed first stage solution as published schedule. The optimistic scenario,pessimistic scenario, most likely scenario, most likely for the congested airport and pessimistic for the other ones are referred as scenario 1, scenario 2, scenario 3 and scenario 4 respectively from that point on until the end of the chapter.

The time-space network of the published schedule when most likely scenario

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Expected
ORD	Taxi-out	11	18	15	15	14
	Dep.Del.	0	56	13	13	10
	Taxi-in	6	10	9	9	8
	Arr.Del	0	71	6	6	5
EWR	Taxi-out	20	25	22	22	22
	Dep.Del.	0	74	10	10	8
	Taxi-in	6	8	7	7	7
	Arr.Del	0	65	9	9	7
RSW	Taxi-out	13	15	13	15	13
	Dep.Del.	11	80	11	80	15
	Taxi-in	4	5	4	5	4
	Arr.Del	7	52	7	52	10
DCA	Taxi-out	16	20	16	20	16
	Dep.Del.	9	68	9	68	12
	Taxi-in	6	8	6	8	6
	Arr.Del	5	52	5	52	8
LAS	Taxi-out	11	18	11	18	11
	Dep.Del.	16	70	16	70	19
	Taxi-in	9	10	9	10	9
	Arr.Del	9	54	9	54	12
	Probability	0.23	0.01	0.72	0.04	

Table 5.2: Scenarios used in stochastic model

is realized is given in Figure 5.1. The continuous lines represents the actual departure times of the flights and the dashed lines represents the planned departure times. The red and blue paths represent aircraft N535AA and N3ETAA respectively. Turnaround times of the aircraft are indicated by the continuous ground lines and idle times are represented by the dashed ground lines. As it can be observed in this figure, flight 1446 is delayed for 4 minutes and also there is 35 minutes of idle time after that flight. However, when the published schedule generated by the stochastic model is used no idle time and delay is observed for flight 1446 when the most likely scenario is realized as it can be observed in Figure 5.2. The departure times and block times of the flights are re-adjusted to decrease the overall cost. The total delay in the system is 5 minutes and no idle time exists. On the other hand, the initial published schedule results in 15 minutes of delay and 138 minutes of idle time.

In table 5.5, the improvements obtained by using stochastic model, instead of generating the schedule by focusing on only one scenario is demonstrated. The TotalCost denotes the expected cost of the schedule, TotalCost is equal to the value of objective function and calculating the objective function value is

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	6:45	6:45	9:30	9:29	0	0
	564	RSW	ORD	10:20	10:20	13:25	13:24	0	13
	1446	ORD	EWR	14:55	14:55	17:40	17:44	4	35
	1411	EWR	ORD	18:45	18:45	21:30	21:30	0	6
N3ETAA	1704	ORD	EWR	6:35	6:35	8:40	8:44	4	0
	1883	EWR	ORD	9:30	9:30	12:10	12:17	7	8
	810	ORD	DCA	13:10	13:10	14:55	14:54	0	14
	2013	DCA	ORD	15:45	15:45	18:00	18:00	0	22
	2013	ORD	LAS	19:00	19:00	23:10	23:10	0	40

Table 5.3: Initial schedule is considered and most likely scenario is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	7:30	7:30	10:15	10:06	0	0
	564	RSW	ORD	10:44	10:44	13:49	13:37	0	0
	1446	ORD	EWR	14:33	14:33	17:18	17:10	0	0
	1411	EWR	ORD	18:00	18:05	20:45	20:45	0	0
N3ETAA	1704	ORD	EWR	7:20	7:20	9:25	9:25	0	0
	1883	EWR	ORD	10:03	10:03	12:43	12:49	5	0
	810	ORD	DCA	13:27	13:28	15:12	15:12	0	0
	2013	DCA	ORD	15:40	15:40	17:55	17:55	0	0
	2013	ORD	LAS	18:15	18:15	22:25	22:25	0	0

Table 5.4: Stochastic model's schedule is considered and most likely scenario is realized

	Stochastic Model	Expected	Initial	Scenario1	Scenario2	Scenario3	Scenario4
TotalCost	48837	49388	67778	49541	92307	49268	69809
TotalCost*	5882	6433	24823	6586	49352	6313	26854
TotalDelay	3031	2894	2409	3050	1839	2887	2332
TotalIdleTime	87	117	425	90	903	117	434
WeightedDelay	51	46	42	59	38	47	37
WeightedIdle	19	27	160	23	336	26	175

Table 5.5: Performance of schedules generated by solving the model according to first row

	Stochastic Model	Expected	Initial	Scenario1	Scenario2	Scenario3	Scenario4
TotalCost	0	1.12	27.95	1.42	47.09	0.87	30.04
TotalCost*	0	8.57	76.3	10.69	88.08	6.83	78.1
TotalDelay	0	-4.73	-25.82	0.62	-64.82	-4.99	-29.97
TotalIdleTime	0	25.64	79.53	3.33	90.37	25.64	79.95
WeightedDelay	0	-10.87	-21.43	13.56	-34.21	-8.51	-37.84
WeightedIdle	0	29.63	88.13	17.39	94.35	26.92	89.14

Table 5.6: Improvements gained by using the schedule generated by solving the stochastic model

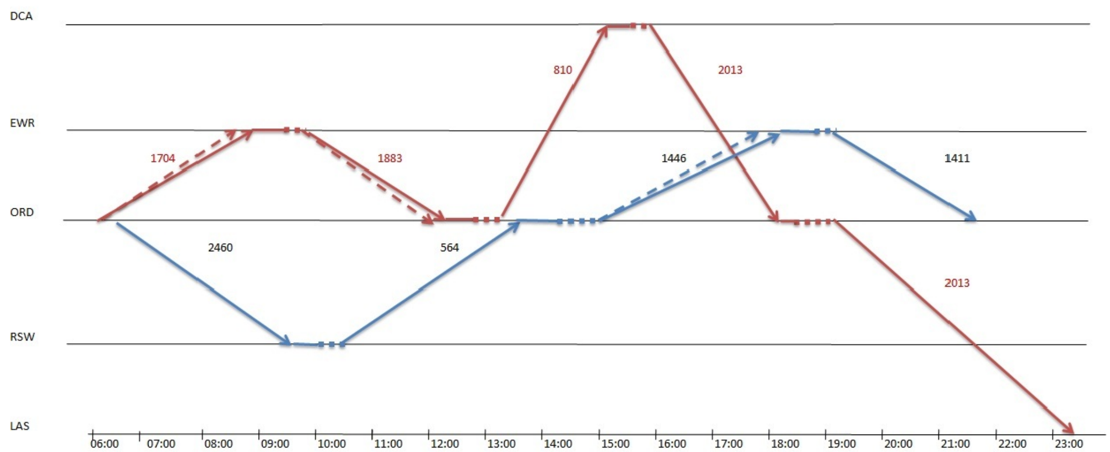

Figure 5.1: Time Space Network for the Published Schedule when most likely scenario is realized

explained in detail in Chapter 3. Since in the model we assume that the upper bound of the cruise time is equal to the max cruise speed, there exists a fixed cost of the fuel and CO_2 emission which is independent of scenarios, TotalCost* denotes the remaining cost when this fixed part is subtracted from the TotalCost. TotalDelay and TotalIdle are the total delay time and idle time in the whole system respectively. They are calculated by summing the delay/idle time of each flight over all scenarios. Their weighted versions are calculated by the sum of multiplication of delay/idle minutes in each scenario with the probability of scenario. In the 5.6 the relative increase/decrease observed in the values specified in the first column by using the published schedule generated by the stochastic model instead of using the schedule generated by individual scenarios, expected values of non-cruise times or taking the initial schedule is demonstrated. This values are calculated by the formula below:

$$\text{Change} = 100 \times \frac{\text{Other Schedule} - \text{Stochastic model's Schedule}}{\text{Other Schedule}}$$

The improvement in the cost varies between 0.87% and 47.09% even for this small example and when fix part of fuel cost is subtracted the improvements goes up to 88%. Stochastic model outperforms the others by decreasing idle time hence, block times. The idle time improvement is higher than 3% for all

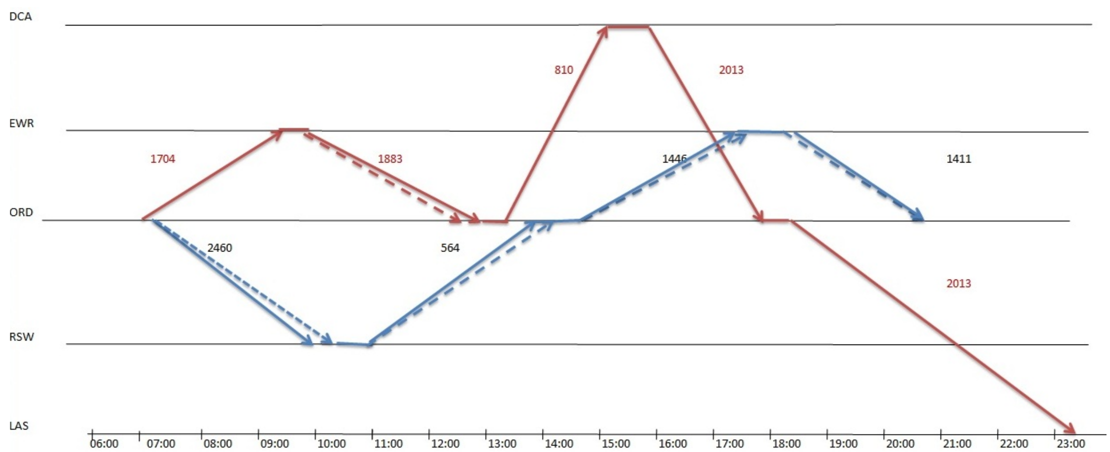

Figure 5.2: Time Space Network for the Stochastic Model's Schedule when most likely scenario is realized

cases and the stochastic model outperforms the pessimistic case by 90% in idle time. Even though, the delay is higher in stochastic model, the gains from the idle time compensate the increase in delay cost. The main reason that lead to difference between delay/idle times is the difference of block times of the flights in different schedules. As expected, the tightest flight block times are observed in the optimistic case and the longest ones are obtained in the pessimistic case. The increase in flight block times decreases delays, on the other hand for the cases where the non-cruise time scenarios are not as pessimistic it might lead to unnecessary idle time. The total flight block time of the initial schedule is 180 minutes higher than the stochastic model.

	Scenario1	Scenario2	Scenario3	Scenario4
Stochastic Model	218	52697	472	18324
Expected	4380	46697	257	14552
Initial	25470	30287	19910	7367
Scenario1	0	50954	1616	17181
Scenario2	50939	0	45380	23551
Scenario3	4674	46211	0	14603
Scenario4	27170	31039	22675	0

Table 5.7: Regret

The relative regret of a first stage solution at a given scenario is the difference between the cost of this solution and the optimal cost in that scenario. The regret table given in 5.7 shows the increase in the cost when the published schedule generated by scenario in the first column is used as the given first stage solution

	Scenario1	Scenario2	Scenario3	Scenario4
Stochastic Model	0.004	0.415	0.011	0.341
Expected	0.08	0.368	0.006	0.271
Initial	0.465	0.239	0.46	0.137
Scenario1	0	0.402	0.037	0.32
Scenario2	0.93	0	1.049	0.438
Scenario3	0.085	0.364	0	0.272
Scenario4	0.496	0.245	0.524	0

Table 5.8: Regret percentage

and the table 5.8 denotes the percentage increase in cost. As it can be observed the largest regret in each scenario, except the pessimistic cases, is experienced when the schedule generated by pessimistic case is used as first stage solution. The optimistic case gives the smallest regret in most likely case, since the most likely values of ORD and EWR; optimistic values of the remaining airports are used. It performs badly in pessimistic cases. The expected model performs better than stochastic model in most likely case since the impact of the most likely case is higher in expected case than the stochastic model. However, in the optimistic case the best results are obtained by using stochastic model. The initial schedule gives the smallest regret in pessimistic case, since the block times of initial schedule is longer than other cases except the pessimistic ones.

The stochastic model minimizes the cost and it leads to significant improvements in idle time. It performs better than using any one of the scenarios. The improvements are observable even in the small example.

The stochastic model given in chapter 3 works with these mechanics. The published schedule is determined by using stochastic model which considers non-cruise time scenarios simultaneously and the actual schedules are determined with respect to the given published schedule for each scenario in the second stage.

Chapter 6

Heuristic Algorithm

We solved mixed integer nonlinear two stage stochastic programming model for robust airline scheduling with nonlinear cost function by using second order conic programming. Instead of using other methods which gives an approximation, using conic programming enabled us to get the exact solution for the model. As demonstrated in the following chapter, we solved the model to optimality on network 1 and 18 scenarios. However, as the network size and the number of scenarios increase, the problems becomes more complex since the number of binary variables and conic inequalities increase. In some cases, the model does not reach to optimality in ten hours. Therefore, we propose two heuristic algorithms, which are referred as relaxation heuristic and binary assignment heuristic, in order to solve the large scale problems in reasonable time.

6.1 Relaxation Algorithm

The binary variables z_{ij}^{w} connect the path of flight i to flight j during the solution process and they are incorporated in the model by big-M constraints. These variables are only integer variables in stochastic model and relaxation of integrality restriction of binary variables reduces to problem into linear programming model which is referred as relaxed model in the rest of the thesis. The linear relaxation of the model can be solved within reasonable amount of time.

In this heuristic, the relaxed model is solved and the departure times of flights, which are first stage decision variables, are taken as input for the second stage. When the first stage decision variable is set, the problem is decomposed into scenarios and each scenario subproblem can be solved independently from other scenarios.

Heuristic 1: Relaxation Heuristic

```
Input: Published Schedule, Delay Scenarios
1 Solve the LP relaxation of the stochastic model and take the published departure times of the flights
  refer it as x̄;
2 for each ω in Ω do
3     Solve Q_ω(x̄);
4     Return the actual schedule S_ω ;
5 end
Output: S_ω ∀ω ∈ Ω, x̄
```

The optimal cost for given first stage decision variable x is computed by $E(Q(x)) = \sum_{\omega \in \Omega} R_{\omega}(x) p_{\omega}$. The algorithm terminates when the second stage models are solved for each scenario.

6.2 Binary Assignment Algorithm

Passenger connections and aircraft connections are the constraints which create dependencies between flights when solving the model. When passenger connections are disregarded or it is assumed that passengers miss their connecting flight, the problem is decomposed into paths. Moreover, the passenger connections introduce binary variables and big-M constraints which makes the problem more complex.

One possible approach for solving the problem is decreasing the number of binary variables by fixing some of them to either 0 or 1 and simplifying the problem. In order to ease the solution process and obtain an initial schedule for all scenarios in a fast manner, integrality constraints of binary variables are relaxed. The fractional z_{ij}^{w} values are considered after the solution of stochastic model with relaxed constraints. These fractional variables are divided into two sets, set Z and set N. It is computed that whether assigning to these fractional values to zero or one is less costly. When assigning a fractional variable to zero is less costly, this variable is put on set Z else set N. The variables in set Z

are forced to take value zero by adding constraints to the stochastic model.The variables in set N are ordered in the descending order of time difference values. If no integral solution is obtained after thirty minutes, the variables which are in the first half of the set N are forced to take value one in the stochastic model by a constraint. This algorithm is solved iteratively and it continues until an integral solution is obtained.

$$\text{timedifference}_{\omega jk} = y_j^\omega + nd_j^\omega + tp_{jk} - y_k^\omega \qquad (6.1)$$

Heuristic 2: Binary Assignment Algorithm

```
Input: Published Schedule, Delay Scenarios
Solve the linear relaxation of the stochastic model and name the current schedule as S and put the binary variables in F;
Set Z and N to empty ;
for each variable in F do
    Compute timedifference of variable in schedule S ;
    if timedifference ≤ 0 then
        Add this variable to Z and delete it from F;
    endif
end
while F not empty do
    Take the variable in F with the smallest timedifference, say it selected;
    Compute approximate cost of assigning selected to zero, say zerocost ;
    Compute approximate cost of assigning selected to one, say onecost ;
    if zerocost ≤ onecost then
        Add selected to Z;
        Adjust the actual schedule S such that passenger connection selected is feasible and call it as S_new ;
        S ← S_new ;
        Delete selected from F;
    endif
    else
        Add selected to N and delete it from F
    endif
end
while each item in N do
    Compute timedifference ;
    if timedifference ≤ 0 then
        Add this variable to Z and delete it from N;
    endif
end
initial is equal to the size of set Z ;
C ← F ;
temp ← 0;
while temp ≠ initial do
    temp = initial ;
    Take the variable in F with the smallest timedifference, say it selectedone;
    Compute approximate cost of assigning selectedone to zero, say zerocost ;
    Compute approximate cost of assigning selectedone to one, say onecost ;
    if zerocost ≤ onecost then
        Add selectedone to Z;
        Adjust the actual schedule S such that passenger connection selectedone is feasible, call this schedule
        S_new and delete selectedone from N ;
        S ← S_new ;
        temp = initial + 1
    endif
end
nothavesol ← true ;
T ← {} ;
while nothavesol do
    Solve the stochastic model by assigning the binary variables in Z to zero and in T to one ;
    if An integer solution is generated in thirty minutes then
        Record this schedule as S_best ;
        Set nothavesol to false ;
        Return S_best ;
    endif
    else
        Order in decreasing order the variables in N. ;
        Put the first half of the variables to T ;
        Remove this items from N ;
    endif
end
Output: S_best
```

6.3 Summary

The stochastic model is solved less than 15 minutes for each instance on network 1 and 18 scenarios. However, when the size of the network and number of scenarios increase, the model does not give optimal solutions in ten hours. In order to overcome this obstacle and produce solutions by considering the scenarios two heuristics are proposed. Both of these heuristics utilize relaxation of integrality constraints.

In the relaxation heuristic, the first stage decision variables are determined by solving stochastic model with relaxed constraints and published departure time of flights generated by this model are taken as fixed. Second stage of the stochastic model is solved by taking these fixed first stage decision variables as input.

In binary assignment heuristic, to ease the solution process and obtain an initial schedule for all scenarios in a fast manner, integrality constraints of binary variables are relaxed. Then, the trade-off between assigning each fractional binary variable to 0 or 1 is considered and the assignment with the minimum cost is done. The stochastic model is solved by fixing the binary variables which are assigned to zero.

Finally, detailed performance evaluations of these heuristics are presented in the next chapter. The trade-off between improvements CPU time and degradation of solution quality are demonstrated by comparing heuristics with the optimal values of stochastic model. Moreover, the comparison of these heuristics with using initial schedule for the first stage decision variables and expected values of non-cruise times to generate schedules is presented.

Chapter 7

Computational Study

In this study, we proposed a mixed integer second order conic two-stage stochastic programming formulation, which is referred as stochastic model, that generates a published schedule by taking non-cruise time scenarios into account. In order to find solutions when the network size and the number of scenarios increase, relaxation and binary assignment heuristics are developed. Binary assignment heuristic relaxes the integrality constraint of binary variables and solves the stochastic model with relaxed constraints to obtain a published schedule then determines that which binary variables should be set to zero. This heuristic runs until an integer solution is obtained. Moreover, the relaxation heuristic takes the published schedule which is generated by solving the stochastic model with relaxed integrality constraints. This published schedule corresponds to the first stage decision variables and the second stage is solved to optimality according to the given first stage solution. In this section, the performance of stochastic model is compared with using the expected values of non-cruise time to generate published schedule and initial schedule in terms of cost, idle and delay time and passenger disruption. Table 7.2 summarizes the size of the network and number of scenarios that are taken into account in computational study. The model is solved to optimality on network 1, which contains 31 flight legs and 9 paths, for 18 scenarios and the heuristic is used to solve the problem on network 1 by considering 228 scenarios. On network 2, which contains 114 flight legs and 31 aircraft, the heuristics are used to solve the problem when 18 and 108 scenarios are considered.

To examine the effects of selected parameters on results, 2^k full-factorial experimental design is conducted. The three experimental factors and their corresponding levels are given in table 7.1.

Factor	Description	Low(0)	High(1)
A	Idle cost	Base Level	3*Base Level
B	Delay cost	0.4	1.2
C	Disruption cost	200	500

Table 7.1: Factor Values

	# of Flights	# of Aircraft	# of Scenarios	Stochastic Model	Relaxation Heuristic	Binary Assignment Heuristic
Network 1	31	9	18	✓	✓	✓
	31	9	228		✓	✓
Network 2	114	31	18		✓	✓
	114	31	108		✓	

Table 7.2: Networks and Scenarios

Aircraft Type	B727 228	B737 500	MD 83	A320 111	A320 212	B767 300
C_1	158100	535994	132111	341985	267834	516220
C_2	1999459	2301504	299628	215437	75819	11208873
C_3	0.001	0.0022	0.0082	0.0037	0.0047	0.0078
C_4	0.0137	0.0096	0.0019	0.0023	0.0013	0.009
Idle Cost	150	140	142	136	144	147

Table 7.3: Aircraft Parameters

There are six types of aircraft and their cost parameters are given in table 7.3. First four parameters are used to calculate the fuel and CO_2 emission costs. The idle cost represents unit idle time cost of the aircraft and the values in table 7.3 is taken as base level. Idle time cost can be considered from two different perspectives, since it is caused by keeping the aircraft and crew idle. Moreover, this cost parameter affects the published schedule generated by stochastic model substantially. To examine its impact on stochastic model's solution, two different levels are selected.

Delay cost represents cost of goodwill per passenger due to the flight delays per minute. This cost is different for the economy class passengers and business class passengers, since their value to company differs in monetary terms. Thus, two different levels are set for this cost parameter.

Disruption cost denotes the cost per passenger that is incurred when the passenger misses next flight in his/her itinerary. In the literature, this cost depends on waiting time of the passenger until a new itinerary is arranged to compensate missed connection. If passenger has to stay until the next day, this cost increases substantially. However, there is no information about current itinerary of passengers, so re-organizing the itineraries is not an option in our problem. To incorporate different waiting times, two different levels are used for passenger disruption cost.

In this study, the schedule used by Aktürk et. al. [11] is used. Information about the flights are taken from the BTS' website. Each column in the table represents the tail number, origin airport, destination airport, departure time and block time of the aircraft respectively. The flights that share the same tail number are flown by the same aircraft and these subsequent flights represent the paths in the schedule.

In order to find the cruise time of the flight, 40 minutes of non-cruise time is subtracted from the given block time. The maximum compression amount is taken as the 15% of the current cruise time and lower bound for cruise time is set to the 85% of the current cruise time.

Number of passengers is selected randomly and the aircraft are assumed serve above 50% fullness. The passenger connection level is taken below fifty percent and passenger connection is possible between flights i and j if the departure time of j is within 45 to 180 minutes of arrival of i and the origin of j is same with the destination of i. The turn back flights are not considered for passenger connection. The passenger turnaround time is selected between 25 minutes and 40 minutes uniformly.

When flight j is operated after i by the same aircraft, turnaround time is necessary to prepare the aircraft to flight j after flight i. The aircraft turnaround times are selected according to the congestion of the airports. The airport congestion levels are determined by the on-time arrival and departure probabilities of airports. The most congested airport has the 1.4 congestion level and the least congested one's is 0.8. The turnaround time is longer in congested airports.

Tail #	Flight #	From	To	Dep. Time	Dur.
N530AA	398	ORD	LGA	06:15	134
	319	LGA	ORD	09:25	170
	2329	ORD	DFW	13:35	155
	2364	DFW	ORD	17:00	150
N459AA	394	ORD	LGA	06:50	135
	321	LGA	ORD	10:00	170
	366	ORD	LGA	13:55	140
	347	LGA	ORD	17:15	170
N531AA	2303	ORD	DFW	06:45	155
	2336	DFW	ORD	10:10	140
	1053	ORD	AUS	13:25	170
	336	AUS	ORD	17:00	165
	336	ORD	LGA	20:40	125
N4XGAA	2079	ORD	SAN	08:45	270
	1438	SAN	ORD	14:00	250
	346	ORD	LGA	19:50	135
N598AA	1341	ORD	SFO	07:50	295
	348	SFO	ORD	13:30	265
	1521	ORD	TUS	19:15	235
N439AA	2455	ORD	PHX	07:10	240
	358	PHX	ORD	11:55	210
	358	ORD	LGA	16:25	145
	371	LGA	ORD	20:00	155
N475AA	407	ORD	STL	06:20	70
	755	STL	ORD	08:35	75
	755	ORD	SAT	10:45	180
	408	SAT	ORD	14:30	160
	408	ORD	PHL	18:05	125
N3EEAA	876	ORD	BOS	06:35	130
	413	BOS	ORD	09:35	185
	413	ORD	SNA	13:45	275
	1262	SNA	ORD	19:10	230
N4YDAA	451	ORD	SFO	09:45	295
	554	SFO	ORD	15:45	265
N3ERAA	496	ORD	DCA	06:45	100
	1715	DCA	ORD	09:15	130
	1715	ORD	LAS	12:25	255
	1708	LAS	ORD	17:20	220
N5CLAA	1425	ORD	SNA	08:25	280
	556	SNA	ORD	14:00	240
	1940	ORD	MIA	19:25	180
N535AA	2460	ORD	RSW	06:45	165
	564	RSW	ORD	10:20	185
	1446	ORD	EWR	14:55	165
	1411	EWR	ORD	18:45	165
N3DMAA	568	ORD	FLL	07:25	175
	711	FLL	ORD	11:10	195
	2021	ORD	SJU	15:25	275
N544AA	2463	ORD	MCI	06:25	90
	754	MCI	ORD	08:40	90
	2321	ORD	DFW	11:15	155
	2356	DFW	ORD	14:40	140
	2487	ORD	DEN	17:50	165
N3EBAA	1565	ORD	MSP	06:40	90
	779	MSP	ORD	09:00	85
	779	ORD	SAN	11:35	260
	1358	SAN	ORD	16:45	235
	1358	ORD	BOS	21:50	125
N3ETAA	1704	ORD	EWR	06:35	125
	1883	EWR	ORD	09:30	160
	810	ORD	DCA	13:10	105
	2013	DCA	ORD	15:45	135
	2013	ORD	LAS	19:00	250
N3DYAA	1063	ORD	LAX	08:50	275
	874	LAX	ORD	14:30	255
	874	ORD	BOS	19:45	135
N3DRAA	1021	ORD	LAS	08:30	245
	1544	LAS	ORD	13:25	215
	1544	ORD	DCA	18:00	105
N5DXAA	1048	ORD	MIA	07:35	190
	1763	MIA	ORD	11:55	200
	1899	ORD	MIA	16:20	185
N454AA	2441	ORD	ATL	06:30	120
	1986	ATL	ORD	09:15	135
	1872	ORD	MCO	12:25	160
	1131	MCO	ORD	15:50	185
N4YMAA	1137	ORD	MSY	08:20	145
	1768	MSY	ORD	11:30	150
	1768	ORD	PHL	15:05	125
	1697	PHL	ORD	18:00	155
N467AA	1823	ORD	PBI	09:20	175
	2067	PBI	ORD	13:00	200
	2067	ORD	STL	17:15	70
	1186	STL	ORD	19:10	80
N536AA	2305	ORD	DFW	07:45	160
	2344	DFW	ORD	11:35	140
	1201	ORD	STL	14:50	65
	1815	STL	ORD	17:00	80
	1815	ORD	SLC	19:15	270
N420AA	1686	ORD	RDU	06:50	110
	2435	RDU	ORD	09:25	135
	2435	ORD	PHX	12:35	235
	1206	PHX	ORD	17:15	205
N546AA	1462	ORD	EWR	08:00	140
	1387	EWR	ORD	11:25	160
	1397	ORD	MCO	15:00	160
	1221	MCO	ORD	18:25	175
N4WPAA	2311	ORD	DFW	09:05	155
	2348	DFW	ORD	12:35	140
	1797	ORD	STL	15:50	70
	1982	STL	ORD	18:00	80
	1339	ORD	SAN	20:15	270
N5EBAA	2375	ORD	EGE	08:10	175
	2378	EGE	ORD	12:25	165
	1677	ORD	SNA	18:40	270
N3DUAA	2099	ORD	LAX	07:00	270
	1972	LAX	ORD	12:40	245
	1972	ORD	RDU	17:45	115
N3ELAA	2057	ORD	SJU	08:30	290
	2078	SJU	ORD	14:25	335
N3DTAA	2363	ORD	HDN	09:50	170
	2318	HDN	ORD	13:40	170
N412AA	2345	ORD	DFW	17:15	155
	2374	DFW	ORD	20:40	130

Table 7.4: Published schedule for 114 flight network

7.1 Analysis of Results on Network 1 with 18 Non-Cruise Time Scenarios

In this section, the stochastic model is applied on network 1, which contains 31 flight legs and 9 aircraft. The number of non-cruise time scenarios included is 18. Three data points are used for ORD and EWR, for the remaining 8 airports 2 instances are created. In the first instance optimistic case is realized and in the other one pessimistic case is actualized for remaining airports.

The schedule generated by stochastic model considers the effect of published departure times of flights on the objective function when different non-cruise time scenarios are realized. According to the published departure times and realized non-cruise time scenario, the actual departure, arrival, delay and idle times are determined in the second stage such that the objective function is minimized. The advantage of using the stochastic model instead of expected values is demonstrated. Moreover, in order to indicate the performance of relaxation heuristic and binary assignment heuristic, a comparison between the optimal values and results of heuristics in terms of solution time and cost are presented.

		Expected						Initial					
		TotalCost			TotalCost*			TotalCost			TotalCost*		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	3.5	4.2	5.4	8.2	10.3	12.1	9.1	13	16.5	17.4	28.9	39.8
	1	6.4	9.2	12	10.8	18	25.3	27	32.8	38.4	40	50.8	61.5
B	0	3.5	7.7	12	10.7	18	25.3	16	27.2	38.4	38.5	50	61.5
	1	4.3	6	7.1	8.6	10.8	12	9.1	18.6	27.6	17.4	29.6	40.9
C	0	3.5	6.6	11.8	8.2	14	25	9.1	23	38.4	17.4	40	61.5
	1	3.7	6.8	12	8.6	14.3	25.3	9.2	22.8	38	17.5	39.6	60.8

Table 7.5: Effect of Factors on Cost

7.1.1 Computational Analysis of Stochastic Model

Stochastic model provides improvement in cost, on the other hand it causes longer CPU times when compared with using one deterministic value. The schedule generated by focusing one deterministic value might perform good in some cases, but the stochastic model considers all scenarios and generates a schedule which

		Expected						Initial					
		TotalDelayMinutes			TotalIdleTime			TotalDelayMinutes			TotalIdleTime		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	-8.1	3.9	16.2	-57.4	-9.8	33.6	-31.8	-16.3	-1.8	33.4	53.7	73.3
	1	-13.4	-10.2	-7.4	30	35.6	42.4	-39.5	-35.4	-31.3	72.8	75.8	78.2
B	0	-13.4	-10	-7.5	23.4	33.4	42.4	-39.5	-33.7	-27.8	68.8	74.1	78.2
	1	-9.3	-0.4	15.9	-57.4	4.8	34	-35.2	-17.9	-1.8	33.4	55.4	75.4
C	0	-12.6	-3.2	16.2	-48.5	14.9	42.4	-39.5	-26.7	-2.8	36.5	65.6	78.2
	1	-13.4	-3.2	15.9	-57.4	10.9	39.2	-38.3	-25	-1.8	33.4	63.9	77.1

Table 7.6: Effect of Factors on Delay and Idle Time

		Expected						Initial					
		WeightedDelay			WeightedIdle			WeightedDelay			WeightedIdle		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	-9.8	6.7	22.8	-52.1	-6.4	37.7	-38.9	-17.3	2.1	35.4	55.4	74.7
	1	-16.5	-12.1	-8.1	35.2	39.7	44	-47.5	-42.1	-35	74.6	77	78.8
B	0	-16.5	-11.7	-5.4	34.3	38.8	44	-47.5	-40	-32.9	71.4	75.7	78.8
	1	-9.8	1.1	22.7	-52.1	8.3	38.4	-42.3	-19.4	2.1	35.4	56.7	76.6
C	0	-15	-2.8	22.8	-51.4	17.3	44	-47.5	-30.7	0.6	37.8	66.8	78.8
	1	-16.5	-2.7	22.7	-52.1	16	42.2	-46.4	-28.8	2.1	35.4	65.6	78.6

Table 7.7: Effect of Factors on Weighted Delay and Idle Time

		Expected						Initial					
		NumofMissedConnect			NumofPassengerMiss			NumofMissedConnect			NumofPassengerMiss		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	-12.9	16.6	52.6	-18.2	26.2	52.1	-60	-3	60	-21.8	25.4	77
	1	-14	-5.2	9.5	-78	-20.6	12.2	-80	-23.3	37	-65.6	5.5	45.2
B	0	-14	1.2	15.8	-78	-1.6	52.1	-80	-30.4	15	-65.6	3.2	48.2
	1	-13.6	8.1	52.6	-52.1	2.4	51.2	-52	4.1	60	-27.6	27.7	77
C	0	-14	1	37	-78	-11.6	46.9	-60	-5.1	37.5	-21.8	21.5	48.2
	1	-13.6	10.4	52.6	-29.1	17.3	52.1	-80	-21.3	60	-65.6	9.4	77

Table 7.8: Effect of Factors on Passenger Disruption

	Replication1			Replication2			Replication 3		
	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min
TotalCost	12.4	6.9	3.6	12.8	7.3	4.0	12.9	7.1	4
TotalCost*	26.5	14.7	8.9	26.9	15.5	10.4	27.5	15.2	8.6
TotalDelayMinutes	15.8	-4.1	-14.7	16.1	-3.2	-13.4	15.9	-4	-14.5
TotalIdleTime	46.4	14.8	-57.4	44.8	18.3	-39.0	45.9	16.6	-53.7
WeightedDelay	22.5	-3.9	-17.9	21.4	-3.5	-17.7	22.1	-3.9	-17.3
WeightedIdle	47.5	18.0	-53.1	47.1	19.3	-47.6	48.5	19.6	-47.8
NumofMissedConnect	80.0	31.9	0.0	47.1	13.5	-12.9	57.5	21.1	-28.6
NumofPassengerMiss	87.3	24.0	-6.2	25.9	11.6	-17.1	78.6	42.8	9.7

Table 7.9: Effects of Replications

minimizes the overall expected cost even though it might not provide optimal results in any of the scenarios. The performance of stochastic model with respect to idle time, delay time and passenger disruption as well as the optimal cost is compared in detail with initial published schedule and using expected values of non-cruise times for determining the published schedule. Definitions of TotalCost, TotalCost*, TotalDelayMinutes, TotalIdleTime, WeightedDelay, WeightedIdle, NumofMissedConnect, NumofPassengerMiss are provided in Chapter 5.

Three different combinations of random variables are considered in order to examine the effect of selection of random variable on the results. These different combinations are referred as replications. For each level of parameters, minimum, average and maximum values of percentage improvement/decline in the considered performance indicators are calculated. For example, the improvement obtained by using the stochastic model instead of generating published schedule by considering expected values of non-cruise times is calculated using the formula below. The tables 7.5, 7.6, 7.7, and 7.8 summarize the results obtained by generating the schedule by stochastic model rather than using initial schedule or considering expected values of non-cruise time.

$$\text{Improvement in cost} = 100 \times \frac{\text{Cost of Expected Schedule-Cost of Stochastic Model's Schedule}}{\text{Cost of Expected Schedule}} \tag{7.1}$$

The stochastic model improves the cost of the system up to 12% and when the fixed cruise cost is subtracted this improvement goes up to 25% when it is compared with using expected values. The improvements are even more significant the stochastic model is compared with initial schedule. The cost decrease goes up to 38% percent and the improvements exceed 61% when the fixed part of the cruise cost is subtracted. To sum up, the improvement is above 3.5% for all factor combinations and when the fixed part is subtracted it leads more than 9% decrease in the controllable part of the cost. On the average, the stochastic model decreases the cost by 6.8% and 22.9% when it is compared with using expected values and initial published schedule, respectively. The results about decline in cost is given in table 7.5. The largest decline in the cost is observed when Factor A and C are in their high levels, B is set to its low level. This is coherent with the insight gained by examining the published schedules of stochastic model and schedule generated by using expected values of non-cruise times, since using the expected values of non-cruise time decreases the delay times and causes larger idle times for our instances. Hence, using stochastic model is more advantageous in that case.

The effect of factor A, the idle time cost, can be observed in the total cost improvement. When the idle time cost is set to its high level, the decrease in the total cost goes up to 12%. On the average stochastic model outperforms using expected value more than 9% and initial schedule by 32.8%. The stochastic model decreases the block times when the idle time costs are higher and because of this reason idle time improvement climbs up to 36% and 76% when it is compared with the schedule generated by using expected solutions and initial published schedule as indicated in table 7.6. However, the block times get longer and using the expected values or initial published schedule decrease the delay in the system. Even in this case, stochastic model outperforms initial schedule solution in terms of number of disrupted passengers. However, as it might be expected the improvement in number of disrupted passengers decreases.

When the Factor B, delay cost, is set to its high level, the improvements in cost decrease. It might be expected, since the schedule generated by considering the expected values puts more emphasis on decreasing delay values even factor B is in its low level. On the other hand, the stochastic model performs approximately same with expected solution in decreasing the total delay and outperforms it in weighted delay time if delay cost takes its high level. Moreover, decrease in delays leads to increase in the number of passengers who can catch their connecting flights.

Importance of factor C depends on the level of passenger connection in this study and it is assumed that less 50%. Effect of factor C, passenger disruption cost, is not very significant in this case, because it comprises at most two percent of the total cost. Hence, changing its level causes small alteration on the average total cost. However, fixing C to its high level substantially decreases the number of missed flight connections and disrupted passenger. Even though stochastic model has more delay than using initial published schedule it outperforms initial schedule in decreasing the number of passengers who miss their connection for each factor level as it can be observed in table 7.8. This result might indicate that the stochastic model is flexible enough to consider number of connecting passengers, even if their effect on the total cost is less than 3%.

The reason that might lead to the results mentioned above is that while using the stochastic model we are more knowledgeable about the scenarios since the variability among them and their individual probabilities are taken into account. However, when expected values are used, the system does not have any information about the neither probability nor the variability. The non-cruise times take only a single value. Moreover, the expected values might be affected from extreme cases substantially even if the associated probability of this extreme case is small. However, in stochastic model since the probabilities of the extreme case is known, the system can evaluate it by taking the other scenarios and their probability into account. Thus, using expected values or extreme cases might lead to very tight or long block times. Hence, focusing on only one scenario might cause poor performance when the overall system is considered.

Factors A and B, idle cost and delay cost, lead the model in opposite directions, since increasing the block times would decrease the delays when the pessimistic scenarios realizes, on the other hand it might cause excessive idle times when the non-cruise times are smaller than the duration devoted to non-cruise time in the block time of flight. As it can be observed in table 7.6, when the factor B is set to its low level the improvements in the average idle time are about 33%, but when it takes its high value the average improvement falls to 4.8%. The same pattern is also observed for delay times when factor A is considered. When factor A takes low level, the stochastic model improves the total delay.

These results also indicate the flexibility of stochastic model. The published departure times adapt to changing levels of factors. For instance, when the delay cost is higher, which means the significance of delay increases for the decision maker, the stochastic model outperforms using expected values of non-cruise times by decreasing the weighted delay in the system. It works in the same way when the idle time cost is higher. The improvements in weighted delay and idle time is higher than total delay and idle time. This is also an indicator of the emphasis stochastic model put on the probability of scenarios.

The number of passengers in each flight, passengers connecting from flight i to j, turnaround times of aircraft and passengers depend on the replication. In order to observe whether the results are affected by the value of random components of the model or not three replications are considered. For each replication minimum, average and maximum values of the change observed for each replication is calculated using equation 7.1. The results are summarized in table 7.9. Considerable difference is not observed between the replications.

7.1.2 CPU Time Analysis of Stochastic Model

The two-stage stochastic mixed integer second order conic programming formulation, which is referred as stochastic model, is implemented in JAVA programming language with a connection to the commercial solver IBM ILOG CPLEX Optimization Studio 12.5. All these experiments are performed on 64-bit Windows 7 Ultimate with 4 GB memory and Intel(R) Core(TM)2 Duo 2.26 GHz CPU. The impact of valid inequality in decreasing the solution time is inspected for each factor level and three replications. The CPU times provided below is in seconds.

		No Valid			Valid		
		Min	Avg	Max	Min	Avg	Max
A	0	38.4	298.7	665.2	5.9	259.3	806.3
	1	49.1	121	292.5	29.9	73.2	282.4
B	0	38.4	115.7	295	29.9	205.5	806.3
	1	65.1	304	665.2	5.9	127	620.6
C	0	49.1	260	642.4	5.9	206	806.3
	1	38.4	152.2	282	29.9	126.5	362.5
		38.4	**208.6**	665.2	5.9	**166.2**	806.3

Table 7.10: CPU Time Analysis of Stochastic Model

Including the valid inequality decreases average solution time more than 20%. However, in some cases including valid inequality might cause longer solution times in some cases. However, no significant pattern is observed.

The problem is harder to solve when factor A and C, idle cost and miss-connection cost, are set to their low level and factor B, delay cost, takes its high level. The reason behind this result might be the trade-off between inserting idle time and experiencing delay. The stochastic model outperforms using expected values of non-cruise times by decreasing the idle time and experiencing more delay. Hence, stochastic model might put more emphasis on decreasing idle times. However, when the delay cost is higher, considering the trade-off between emphasis put on decreasing idle time and high cost of delay might lead to increase in solution time.

In addition, solving the problem when factor A is set to its low level is more time consuming than it takes its high level. Reverse trend is observed for factor B. These trends also support the idea that the stochastic model put more emphasis on decreasing idle time and when it takes its high level solving the problem becomes even more faster. However, when factor B is set to its high level the model might consume more time to consider trade-off between idle time and delay since they work in opposite ways. Solution time is shorter when Factor C takes its high level rather than its low level. However, as indicated in the previous section the contribution of the passenger disruption cost to objective function value is small when it is compared with delay and idle time costs. Hence, the results about different levels of factor C might not be representative of all cases.

The stochastic model is solved to optimality in approximately 209 seconds on the average when no valid inequalities are included in the stochastic model over the network which includes 31 flight legs and 9 paths when number of scenarios in the system is equal to 18. Including the valid inequality decreases the solution time to 166.2 seconds. However, when the size of the network or the number of scenarios increases the problem is not solved optimality in three hours. In order to solve the problems on large networks or when the number of scenarios increases, relaxation heuristic algorithm is used.

7.1.3 Computational Analysis of Heuristics

The exact solution is obtained for stochastic model when nonlinear cost function is transformed into second order conic inequalities. As indicated above, the optimal solution of stochastic model is obtained for NW1 18. Significant savings in cost and idle time are observed when stochastic model is used instead of the expected values of non-cruise times. In this section, we compare the performance of the relaxation heuristic and binary assignment heuristic with the optimal results and using expected values of non-cruise times. To demonstrate the impact of using the heuristics instead of solving the model to optimality, total cost of the schedule generated by the stochastic model and two heuristics are juxtaposed. For each experimental factor combination the minimum, average and maximum optimality gap are calculated over three replications as seen is table 7.1.3, the results are shown under the first column which is referred as gap. The gap between the objectives of relaxed heuristic and optimal solution of stochastic model is calculated as follows:

$$\text{Gap} = 100 \times \frac{\text{Cost of Heuristic - Cost of Stochastic Model}}{\text{Cost of Stochastic Schedule}} \tag{7.2}$$

Moreover, the performance of heuristics is compared with using expected values in terms of total cost. The results are demonstrated in the second and fifth columns of table 7.1.3. The improvements obtained by heuristics instead of using the expected values are computed by the function 7.3. The third and last column of 7.1.3 denote the average increase in the number of disrupted passengers when the heuristics are used instead of the optimal schedule generated by the stochastic model.

$$\text{Improvement} = 100 \times \frac{\text{Cost of Expected Schedule - Cost of Heuristic}}{\text{Cost of Expected Schedule}} \tag{7.3}$$

2^3 full-factorial experimental design with three replications for each combination is conducted in to examine the impact of different factor levels on the quality

A	B	C	Relaxation Heuristic Gap	Improvement	Increment in # of Disrupted Passengers	Binary Assignment Heuristic Gap	Improvement	Increment in # of Disrupted Passengers
0	0	0	0.956	2.656	127	0.277	3.311	34
0	0	1	0.437	3.594	15	0.306	3.719	50
0	1	0	0.544	3.916	196	0.097	4.343	10
0	1	1	0.556	4.4	28	0.354	4.593	404
1	0	0	1.015	11.488	485	1.921	10.694	-94
1	0	1	1.361	11.124	88	1.517	10.987	42
1	1	0	0.63	6.14	299	0.476	6.283	4
1	1	1	0.941	5.776	54	0.511	6.177	126
			0.805	6.137	162	0.682	6.263	72

Table 7.11: Comparison of heuristics

of the solution generated by the heuristics. Setting factor A to its high level increases the gap between the optimal cost of stochastic model and heuristics for each factor combination. The largest gap is obtained when the factor A is set to its high level and others take to their low level. Moreover, the increment in the number of disrupted passengers increases in relaxation heuristic but opposite trend is observed for binary assignment heuristic. When more emphasis is put on decreasing idle time values, the schedule generated by the relaxation have tighter block times. Hence, delay values and passenger disruption increase. On the other hand, when factor B takes its high level, the gap between the optimal value and objective function value of stochastic model decreases even though the increment in the number of disrupted passengers increases. As indicated in the previous chapters, the performance improvement of the stochastic model decreases when it is compared with using expected values of non-cruise time when the factor B is set to its high level. Both the stochastic model and heuristics perform more closely when factor B takes its high level. Setting factor C to its high level, decreases increment in the number of disrupted passenger. This result is understandable, since decreasing passenger disruption becomes more important when cost of disruption is higher. However, the gap between the optimal value and solution of relaxation heuristic increases when focus is shifted to decreasing disruption.

However, the heuristics perform better than using expected values of non-cruise times for all factor combinations. Especially, the improvement is higher when factor A, idle time, takes its high level which is the case where the largest gaps between the solution of stochastic model's and heuristics' objective function value are obtained. These results are complementary with the comparison between the stochastic model and using expected values presented in the previous sections. Moreover, there is no difference in CPU times between using the expected values and heuristics. Therefore, it can be concluded that using the heuristics proposes a better schedule in the same duration when it is compared with using expected values of non-cruise times.

7.2 Analysis of Heuristics

In this section performance of heuristics is compared with the initial published schedule and using expected values of non-cruise times. NW1 228, NW2 18, NW2 108 are considered in this section. Only the relaxation heuristic is applied on network 2 for the case that includes 108 scenarios since preliminary analyses indicate that the binary assignment heuristic does not perform better than the relaxation heuristic even though it doubles CPU time of relaxation heuristic.

To compare the heuristics with the stochastic model, the stochastic model is solved with 1800 seconds time limit and the best LP solution is taken as the lower bound of the stochastic model. The heuristics are compared with this solution by the formula 7.4 and the result is reported as Gap From LB. Moreover, the improvements obtained over initial schedule and using expected values of non-cruise times are calculated by the formula 7.3 as indicated in the previous section. Finally, the performances of heuristics are compared with initial schedule and using expected values of non-cruise times in terms of number of disrupted passengers.

The maximum LB gap 7.5% is obtained on network 1 for 228 scenarios for relaxation heuristic as indicated in 7.12 and minimum LB gap 0.3% is obtained on network 2 for 18 scenarios for relaxation heuristic as presented in 7.14. The heuristics lead more than 5.4% improvement over initial schedule for all factor levels, networks and scenarios and on the average 15.5% improvement is observed. The largest improvement is obtained on network 2 for 108 scenarios. The increase over using expected values of non-cruise time is more than 0.9% for each factor level and on the average it is 6.4%. Even though the largest improvements are observed on network 1, the best average improvements are obtained on network 2 for 18 scenarios as shown in table 7.17.

When the effect of different factor level is considered a trend similar to the one obtained on NW1 18 is observed. The high level of idle cost leads to more improvement and the opposite inclination is observed for high level of Factor B, delay cost. However, no significant effect of factor C, misconnection cost, is observed. Since the misconnection variables are relaxed at the first stage of both heuristics, the different levels of factor C might not effect the result.

Including more scenarios increases the gap from lower bound as it can be observed on network 2 in table 7.17. Moreover, this trend is also observable on network 1. Increasing number of scenarios lead to decrease in improvements over using expected values of non-cruise time scenarios. On the other hand, the cost improvements are more substantial when the heuristics are compared with initial published schedule. Including more scenarios has impact on both stochastic model, hence on heuristics, and expected values. However, using the initial published schedule is similar to decomposing the problem into second stage problems for given first stage solution. Hence, incorporating more scenarios enable the user to consider more option at the planning stage and the published schedule obtained at the end of the heuristics are more robust in terms of decreasing cost.

Using binary assignment heuristic leads to more improvement when it is compared with relaxation heuristic even though it consumes more CPU time. Even though the improvements in idle time is less in binary assignment heuristic, the decrease in delays is larger than the relaxation heuristic. For the network 1 with 228 scenarios, the binary assignment heuristic is solved in two iterations, which means after fixing the variables assigned to zero the other half of the binary variables are fixed to one. This leads to increase in passenger misconnections. The changes in the percentage of misconnections are given in Figure 7.1. Relaxation Heur. Exp., Bin. Assgn. Heur. Exp. denotes the percentage of change observed in number of misconnecting passengers when relaxation heuristic, for the second instance binary assignment heuristic is used instead of using expected values of non-cruise times. Relaxation Heur. Init., Bin. Assgn. Heur. Init. represents the percentage of change observed in number of misconnecting passengers when relaxation heuristic, for the second instance binary assignment heuristic is used instead of initial schedule. In that case, as indicated in Figure 7.2, binary assignment heuristic always performs worse than initial schedule in terms of decreasing number of misconnected passengers and this degradation is more significant when the idle cost is set to its high level. On the other hand, the binary assignment heuristic is solved at one iteration on network 2 for 18 scenarios and in that case binary assignment heuristic performs better than both initial schedule and using expected values of non-cruise times in decreasing the misconnections as indicated in Figure 7.1. However, relaxation heuristic performs worse than initial schedule when idle time cost is set to its high level. This result is also valid on network 2 for 108 scenarios and network 1 for 228 scenarios as indicated in Figure 7.3 and 7.2 respectively.

$$\text{Gap} = 100 \times \frac{\text{Cost of Heuristic - LB of Stochastic Model}}{\text{LB of Stochastic Schedule}} \tag{7.4}$$

		Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	0.5	1.8	3.1	0.9	5.5	21.3	6.2	6.9	8.1
	1	1	2.9	4.9	1.3	6.6	11.2	12.5	19.9	27.5
B	0	1.6	2.8	4.8	0.9	6.1	11.2	6.2	16.9	27.5
	1	0.5	1.9	4.9	1.3	5.9	21.3	6.2	9.9	14.3
C	0	0.5	1.9	4.9	1.3	6.5	21.3	6.3	13.5	27.5
	1	1.2	2.8	4.8	0.9	5.6	10.7	6.2	13.2	26.2
		0.5	2.4	4.9	0.9	6	21.3	6.2	13.4	27.5

Table 7.12: Performance of Relaxation Heuristic on Network 1 for 228 scenarios

7.2.1 CPU Time Analysis of Heuristics

Computation times are less than one hour for each heuristic on Network 2 when 18 scenarios are considered. However, when 108 scenarios are considered a solution that performs better than the result of relaxation heuristic is not found in two hours in the preliminary analysis. Binary assignment heuristic does not give a feasible solution in the first iteration on Network 1. Hence, each run takes longer

		Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	0.4	1.9	4	1	5.6	21.4	5.4	7	8.5
	1	1.1	3.6	7.5	3.5	6.9	11.1	14.1	20.2	27.2
B	0	1.5	3.7	7.5	1	5.9	11.1	6.3	16.7	27.2
	1	0.4	1.8	3.9	3.5	6.6	21.4	5.4	10.5	15
C	0	0.4	2.1	7.5	2.4	6.9	21.4	5.7	13.9	27.2
	1	1.1	3.4	5.7	1	5.6	10.7	5.4	13.3	26.3
		0.4	2.8	7.5	1	6.3	21.4	5.4	13.6	27.2

Table 7.13: Performance of Binary Assignment Heuristic on Network 1 for 228 scenarios

		Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	0.5	1.3	2.4	2.3	5.6	9	7.3	7.9	8.6
	1	0.3	1.2	2.3	3.7	7.6	12	12.2	19.1	26.4
B	0	0.3	1.4	2.4	2.3	6.9	12	7.5	16.7	26.4
	1	0.4	1.1	2.3	3.7	6.4	9	7.3	10.4	13.4
C	0	0.3	0.7	1.5	2.3	6.7	12	7.3	13.7	26.4
	1	0.9	1.8	2.4	2.5	6.6	10.8	7.5	13.4	24.9
		0.3	1.3	2.4	2.3	6.6	12	7.3	13.5	26.4

Table 7.14: Performance of Relaxation Heuristic on Network 2 for 18 scenarios

		Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	0.4	1.1	2.1	2.6	5.9	9.4	7.4	8.1	8.8
	1	0.4	1.2	2.1	3.7	7.6	11.7	12.4	19.2	26.2
B	0	0.6	1.3	1.9	2.6	7.1	11.7	7.8	16.8	26.2
	1	0.4	1	2.1	3.7	6.5	9.4	7.4	10.5	13.4
C	0	0.4	0.7	1.3	2.6	6.6	11.7	7.4	13.7	26.2
	1	0.7	1.5	2.1	3.1	6.9	11.1	7.9	13.6	25.2
		0.4	1.1	2.1	2.6	6.8	11.7	7.4	13.7	26.2

Table 7.15: Performance of Binary Assignment Heuristic on Network 2 for 18 scenarios

		Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
A	0	0.8	1.6	2.4	3.1	4	5.1	11.9	13.9	17.1
	1	0.9	1.9	4.8	5.8	8.5	11.3	28.2	33	38.6
B	0	0.9	2	4.8	3.1	7.1	11.3	16.7	27.1	38.6
	1	0.8	1.6	2.6	4.2	5.4	6.5	11.9	19.9	29.4
C	0	0.8	1.2	1.6	3.1	6.3	11.3	11.9	23.9	38.6
	1	1.3	2.4	4.8	3.4	6.2	10.6	12	23	37.1
		0.8	1.8	4.8	3.1	6.3	11.3	11.9	23.5	38.6

Table 7.16: Performance of Relaxation Heuristic on Network 2 for 108 scenarios

			Gap From LB			Cost Imp. Exp.			Cost Imp. Init.		
			Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Network 1	228 Scenarios	Relaxation	0.5	2.4	4.9	0.9	6	21.3	6.2	13.4	27.5
		Binary	0.4	2.8	7.5	1	6.3	21.4	5.4	13.6	27.2
Network 2	18 Scenarios	Relaxation	0.3	1.3	2.4	2.3	6.6	12	7.3	13.5	26.4
		Binary	0.4	1.1	2.1	2.6	6.8	11.7	7.4	13.7	26.2
	108 Scenarios	Relaxation	0.8	1.8	4.8	3.1	6.3	11.3	11.9	23.5	38.6

Table 7.17: Performance of Heuristics on Network all networks

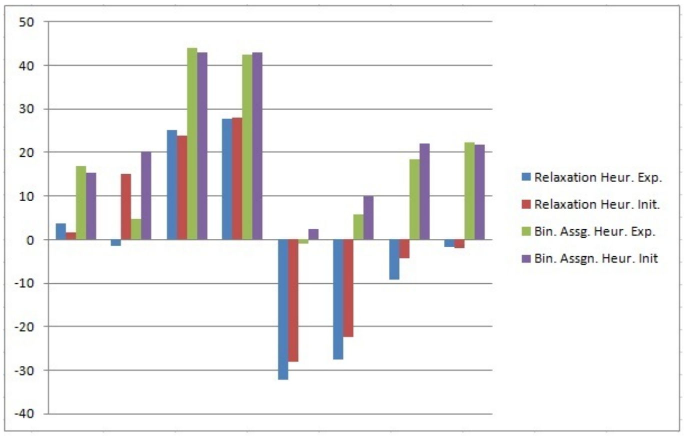

Figure 7.1: % Difference in Number of Disrupted Passengers on Network 2 for 18 scenarios

than one hour on Network 1 when 228 scenarios are considered. The average values for each factor level and network is given on table 7.18, CPU times are given in seconds.

		Relaxation Heuristic			Binary Assignment Heuristic	
		NW 1 228	NW 2 18	NW 2 108	NW 2 18	NW 1 228
A	0	233	663	2497	2188	3939
	1	227	663	2916	2171	3932
B	0	247	693	2724	2168	3937
	1	214	634	2689	2191	3934
C	0	228	725	2689	2201	3928
	1	233	601	2724	2166	3936
		230	663	2707	2181	3934

Table 7.18: CPU Times of Heuristics

The effect of factor A, idle cost, is not very significant on Network 1 for 228 scenarios and Network 2 for 18 scenarios. However, as the number of scenarios and size of the network increases the difference becomes more substantial. When idle time cost takes its high level, computation time increases. Longest CPU time is observed on Network 2 when both factors A and B take their high level and

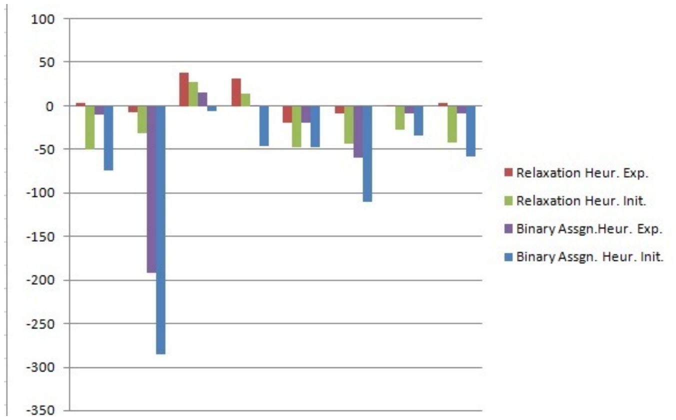

Figure 7.2: % Difference in Number of Disrupted Passengers on Network 1 for 228 scenarios

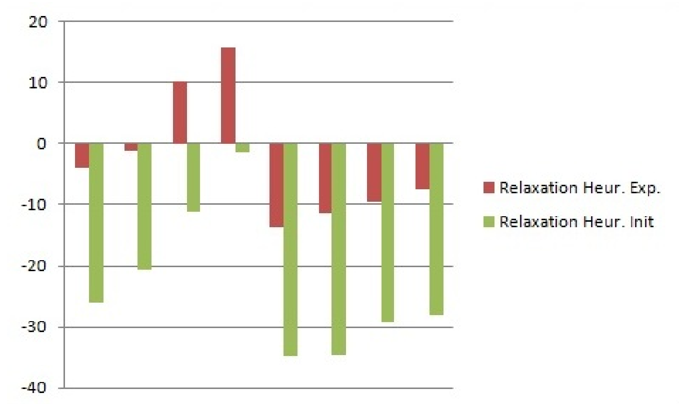

Figure 7.3: % Difference in Number of Disrupted Passengers on Network 2 for 108 scenarios

factor C is set to its low level. The increase in the size of the network and increase in number of scenarios cause the long CPU times for LP relaxation as it can be observed on Network 2 for 108 scenarios. Moreover, the increase in number of scenarios causes to long solution times and solution quality of heuristics decrease in the quality of solution when compared with using expected value of non-cruise times.

7.3 Summary

In this section, the computational study of stochastic model is given on Network 1, which contains 31 flight legs and 9 aircraft, for 18 scenarios is given. Different factor levels are selected to observe the effect of changing parameters on the quality of solution. The result of stochastic model is compared with the using expected values of non-cruise times and initial published schedule. Moreover, the performance of solutions are compared with optimal results obtained for this case.

Moreover, the performance of relaxation heuristic is demonstrated for three other combinations, Network 1 with 228 scenarios, Network 2 with 18 and 108 scenarios and solution of binary assignment heuristic is given for Network 1 with 228 scenarios, Network 2 with 18 scenarios. Cost, number of disrupted passenger and CPU time analyses are presented.

Chapter 8

Conclusions and Future Research

The motivation behind this study and definition of the problem are provided in Chapter 1. In this chapter, a brief summary of the thesis, contributions and future research directions are stated.

8.1 Summary and Contributions

In this study, nonlinear two stage stochastic programming model is developed to obtain robust flight schedule by minimizing the airline's operating, delay and passenger disruption costs. The block time of a flight is examined in two parts as cruise time and non-cruise time. The cruise time is less susceptible to variations and it is controllable within some limits. Hence, in our study the cruise time of flight is considered as a decision variable. However, non-cruise time depends on the weather conditions, airport congestion and security delays. In order to capture this variability in our model, non-cruise time scenarios are generated and incorporated into the problem by two stage stochastic programming. Non-cruise time of flights are determined under each scenario according to their origin and destination airports. The problem is modeled as two stage stochastic programming model in which the published departure times of flights are the first stage decision variables and actual departure, arrival and cruise times, idle and delay times are determined in the second stage. The initial flight durations are adjusted within some limits in order to protect the market share of the airline. Current aircraft and passenger connections are satisfied in the stochastic model. The objective is minimizing the cost of CO_2 emission, fuel consumption, idle time, delay and passenger disruption. Our mathematical model allows us to take the trade-off between idle time insertion, experiencing delay and adjusting the speed of the aircraft into account and produces a robust flight schedule. To the best of our knowledge this is the first study that generates non-cruise time scenarios and develop two stage stochastic model in order to determine flight schedule.

The cost function of CO_2 emission, fuel consumption is nonlinear and this non-linearity is handled by second order cone programming. Moreover, a valid inequality is introduced in order to decrease the solution time. The optimal values can be obtained on small networks for moderate number of scenarios. However, the problem becomes computationally intractable as the size of the network and number of scenarios increase. In order to overcome this obstacle and generate schedules when the problem size increases, two heuristic algorithms are developed. In relaxation heuristic, the integrality restriction of binary variables is relaxed in the stochastic model and the first stage decision variables generated by solving relaxed stochastic model is taken as input for the second stage. The second stage of the stochastic model is decomposed into scenarios. In binary assignment heuristic the relaxed stochastic model is solved and the fractional binary variables are taken as input. The cost of assigning the binary variables to zero or one is computed. The variables which are assigned to zero is forced to take value zero in the stochastic model and the stochastic model is solved for thirty minutes.

Two networks are considered in the solution process. In network 1 there are 31 flight legs and 9 aircraft, in network two 114 flights and 31 aircraft exist. Two instances are created for each network that include different number of scenarios. The instance 1 on network 1 consists 18 scenarios and is solved to optimality. However, for the other instances heuristics are used. The heuristics outperform using expected values of non-cruise times and initial schedule for each instance.

8.2 Future Study

This study can be extended to several directions. In our study, we assume that the non-cruise times of the flights depend on the origin and destination airports. Time of the flight can also be a factor that affects the non-cruise time of the flight. The airports are assumed to be independent of each other in this thesis. However, the congestion of airports might have an impact of the non-cruise time of other airports. Hence, the correlation between airports can be incorporated into scenario generation process. Moreover, the congestion at the beginning of the day might affect the rest of the day. This phenomenon can be included in the model by using multi-stage stochastic programming instead of two stage stochastic programming.

In this study, the cancellation of flight is not considered as an option and delay duration can be very long in some cases. Allowing flight cancellations might also be effective to handle pessimistic non-cruise scenarios. In addition, schedule recovery can be considered in the second stage. Moreover, the path of the aircraft are taken as fixed but the paths of the aircraft can be reconstructed in the first stage by taking the non-cruise time scenarios into account. In this way, more robust schedule can be generated.

Since there is no information on the passenger itineraries in our study, the passenger disruptions are only handled in the objective function with a cost component. However, recovering the passenger itinerary for the passengers who miss their flights is possible if these data are provided.

The delay cost per minute does not depend on the duration of the delay. However, this assumption is not very realistic since long delays might cause more frustration among passengers. The delay cost can be nonlinear or piecewise linear to capture the volatile response of the passengers depending on the duration of the delay.

Bibliography

[1] P. Belobaba, A. Odoni, and C. Barnhart, *The global airline industry*, vol. 23. John Wiley & Sons, 2009.

[2] V. Deshpande and M. Arikan, "The impact of airline flight schedules on flight delays," *Manufacturing & Service Operations Management*, vol. 14, no. 3, pp. 423–440, 2012.

[3] A.T.C.-Modernization, "Nextgen: Near-term achievable goals," in *111th Congress Subcommittee on Aviation Hearing*, 2009.

[4] C. Barnhart and A. Cohn, "Airline schedule planning: Accomplishments and opportunities," *Manufacturing & Service Operations Management*, vol. 6, no. 1, pp. 3–22, 2004.

[5] BTS, "Airline on-time statistics and delay causes." `http://www.transtats.bts.gov/otdelay/OTDelayCause1.asp?pn=1`, 2013. Visited December 2013.

[6] D. S. Fearing, C. Barnhart, A. Odoni, and V. Vaze, "Demand and capacity management in air transportation," *EURO Journal on Transportation and Logistics*, vol. 1, no. 1-2, pp. 135–155, 2012.

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

[8] L. Marla and C. Barnhart, "Robust optimization: Lessons learned from aircraft routing," tech. rep., 2010.

[9] S. Ahmadbeygi, A. Cohn, and M. Lapp, "Decreasing airline delay propagation by re-allocating scheduled slack," *IIE transactions*, vol. 42, no. 7, pp. 478–489, 2010.

[10] V. Chiraphadhanakul and C. Barnhart, "Robust flight schedules through slack re-allocation," *EURO Journal on Transportation and Logistics*, vol. 2, no. 4, pp. 277–306, 2013.

[11] M. S. Aktürk, A. Atamtürk, and S. Gürel, "Aircraft rescheduling with cruise speed control," *Operations Research*, 2014.

[12] A. S. Duran, S. Gurel, and M. S. Akturk, "Robust airline scheduling with controllable cruise times and chance constraints," *to appear in IIE Transactions*, 2014.

[13] M. Sohoni, Y.-C. Lee, and D. Klabjan, "Robust airline scheduling under block-time uncertainty," *Transportation Science*, vol. 45, no. 4, pp. 451–464, 2011.

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

[15] Y. Ageeva and J.-P. Clarke, "Approaches to incorporating robustness into airline scheduling," Master's thesis, Massachusetts Institute of Technology, 2000.

[16] A. Shapiro and A. Philpott, "A tutorial on stochastic programming," *Manuscript. Available at www2. isye. gatech. edu/˜ ashapiro/publications. html*, 2007.

[17] A. Shapiro and P. R. Andrzej, *Stochastic programming.* Elsevier, 2003.

[18] J. R. Birge and F. Louveaux, *Introduction to stochastic programming.* Springer, 2011.

[19] M. Kaut and S. W. Wallace, “Evaluation of scenario-generation methods for stochastic programming,” *Manuscript. Available at http://work.michalkaut.net/papersetc/SGevaluation.pdf*, 2003.

[20] L. Chen and T. Homem-de Mello, “Re-solving stochastic programming models for airline revenue management,” *Annals of Operations Research*, vol. 177, no. 1, pp. 91–114, 2010.

[21] A. Möller, W. Römisch, and K. Weber, “A new approach to o&d revenue management based on scenario trees,” *Journal of Revenue and Pricing Management*, vol. 3, no. 3, pp. 265–276, 2004.

[22] N. Growe-Kuska, H. Heitsch, and W. Romisch, “Scenario reduction and scenario tree construction for power management problems,” in *Power Tech Conference Proceedings, 2003 IEEE Bologna*, vol. 3, pp. 7–pp, IEEE, 2003.

[23] A. Möller, W. Römisch, and K. Weber, “Airline network revenue management by multistage stochastic programming,” *Computational Management Science*, vol. 5, no. 4, pp. 355–377, 2008.

[24] H. Heitsch and W. Römisch, “Scenario tree modeling for multistage stochastic programs,” *Mathematical Programming*, vol. 118, no. 2, pp. 371–406, 2009.

[25] B. Lardeux, O. Goyons, and C.-A. Robelin, “Availability calculation based on robust optimization,” *Journal of Revenue & Pricing Management*, vol. 9, no. 4, pp. 313–325, 2010.

[26] O. Listes and R. Dekker, “A scenario aggregation–based approach for determining a robust airline fleet composition for dynamic capacity allocation,” *Transportation Science*, vol. 39, no. 3, pp. 367–382, 2005.

[27] G. W. Graves, R. D. McBride, I. Gershkoff, D. Anderson, and D. Mahidhara, “Flight crew scheduling,” *Management Science*, vol. 39, no. 6, pp. 736–745, 1993.

[28] J. W. Yen and J. R. Birge, "A stochastic programming approach to the airline crew scheduling problem," *Transportation Science*, vol. 40, no. 1, pp. 3–14, 2006.

[29] S. Yan, C.-H. Tang, and Y.-Z. Hou, "Airport gate reassignments considering deterministic and stochastic flight departure/arrival times," *Journal of Advanced Transportation*, vol. 45, no. 4, pp. 304–320, 2011.

[30] M. O. Ball and G. Lulli, "Ground delay programs: Optimizing over the included flight set based on distance," *Air Traffic Control Quarterly*, vol. 12, no. 1, pp. 1–25, 2004.

[31] A. Mukherjee and M. Hansen, "A dynamic stochastic model for the single airport ground holding problem," *Transportation Science*, vol. 41, no. 4, pp. 444–456, 2007.

[32] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, vol. 2. Siam, 2001.

[33] O. Günlük and J. Linderoth, "Perspective reformulations of mixed integer nonlinear programs with indicator variables," *Mathematical programming*, vol. 124, no. 1-2, pp. 183–205, 2010.

[34] M. S. Aktürk, A. Atamtürk, and S. Gürel, "A strong conic quadratic reformulation for machine-job assignment with controllable processing times," *Operations Research Letters*, vol. 37, no. 3, pp. 187–191, 2009.

[35] O. Şafak, "Fleet type assignment and robust airline scheduling with chance constraints under environmental emission considerations," Master's thesis, Bilkent University, 2013.

$\updownarrow$

Appendix A

Computational Results

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	6:45	6:45	9:30	9:12	0	0
	564	RSW	ORD	10:20	10:25	13:25	13:20	0	35
	1446	ORD	EWR	14:55	15:07	17:40	17:29	0	51
	1411	EWR	ORD	18:45	18:45	21:30	21:16	0	21
N3ETAA	1704	ORD	EWR	6:35	6:35	8:40	8:17	0	0
	1883	EWR	ORD	9:30	9:36	12:10	12:02	0	41
	810	ORD	DCA	13:10	13:19	14:55	14:46	0	37
	2013	DCA	ORD	15:45	15:50	18:00	17:56	0	36
	2013	ORD	LAS	19:00	19:00	23:10	22:59	0	44

Table A.1: Initial schedule is considered and optimistic scenario is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	6:45	6:45	9:30	10:42	72	0
	564	RSW	ORD	10:20	11:20	13:25	16:20	175	0
	1446	ORD	EWR	14:55	17:16	17:40	21:29	229	0
	1411	EWR	ORD	18:45	22:24	21:30	27:10	340	0
N3ETAA	1704	ORD	EWR	6:35	6:35	8:40	10:14	94	0
	1883	EWR	ORD	9:30	10:53	12:10	15:35	205	0
	810	ORD	DCA	13:10	16:14	14:55	19:23	268	0
	2013	DCA	ORD	15:45	19:52	18:00	24:01	361	0
	2013	ORD	LAS	19:00	24:21	23:10	29:37	387	0

Table A.2: Initial schedule is considered and pessimistic scenario is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	6:45	6:45	9:30	9:56	26	0
	564	RSW	ORD	10:20	10:34	13:25	14:28	63	0
	1446	ORD	EWR	14:55	15:24	17:40	17:54	14	0
	1411	EWR	ORD	18:45	18:49	21:30	21:30	0	0
N3ETAA	1704	ORD	EWR	6:35	6:35	8:40	8:44	4	0
	1883	EWR	ORD	9:30	9:30	12:10	12:17	7	8
	810	ORD	DCA	13:10	13:10	14:55	15:33	38	14
	2013	DCA	ORD	15:45	16:02	18:00	19:06	66	0
	2013	ORD	LAS	19:00	19:25	23:10	23:56	46	0

Table A.3: Initial schedule is considered and most likely for ORD,EWR and pessimistic for others is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	7:30	7:30	10:15	9:57	0	0
	564	RSW	ORD	10:44	10:47	13:49	13:42	0	12
	1446	ORD	EWR	14:33	14:41	17:18	17:03	0	3
	1411	EWR	ORD	18:00	18:00	20:45	20:30	0	2
N3ETAA	1704	ORD	EWR	7:20	7:20	9:25	9:02	0	0
	1883	EWR	ORD	10:03	10:10	12:43	12:36	0	30
	810	ORD	DCA	13:27	13:35	15:12	15:02	0	20
	2013	DCA	ORD	15:40	15:44	17:55	17:50	0	13
	2013	ORD	LAS	18:15	18:15	22:25	22:10	0	5

Table A.4: Stochastic model's schedule is considered and optimistic scenario is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	7:30	7:30	10:15	11:28	73	0
	564	RSW	ORD	10:44	12:06	13:49	17:06	197	0
	1446	ORD	EWR	14:33	18:02	17:18	22:17	299	0
	1411	EWR	ORD	18:00	23:11	20:45	28:00	435	0
N3ETAA	1704	ORD	EWR	7:20	7:20	9:25	11:00	95	0
	1883	EWR	ORD	10:03	11:38	12:43	16:21	217	0
	810	ORD	DCA	13:27	17:00	15:12	20:10	298	0
	2013	DCA	ORD	15:40	20:38	17:55	24:49	414	0
	2013	ORD	LAS	18:15	25:09	22:25	30:27	482	0

Table A.5: Stochastic model's schedule is considered and pessimistic scenario is realized

Tail #	Flight #	From	To	Dep.Time	Act.Dep	Arr.Time	Act. Arr. Time	Delay	Idle
N535AA	2460	ORD	RSW	7:30	7:30	10:15	10:41	26	0
	564	RSW	ORD	10:44	11:19	13:49	15:13	84	0
	1446	ORD	EWR	14:33	16:09	17:18	18:40	82	0
	1411	EWR	ORD	18:00	19:34	20:45	22:09	84	0
N3ETAA	1704	ORD	EWR	7:20	7:20	9:25	9:25	0	0
	1883	EWR	ORD	10:03	10:04	12:43	12:48	5	0
	810	ORD	DCA	13:27	13:27	15:12	15:50	39	0
	2013	DCA	ORD	15:40	16:19	17:55	19:23	88	0
	2013	ORD	LAS	18:15	19:43	22:25	24:14	109	0

Table A.6: Stochastic model's schedule is considered and most likely for ORD,EWR and pessimistic for others is realized

Vita

Nil Karacaoğlu was borned in April 11, 1990 in Tokat. She graduated from Aydın Science High School in 2008. She attended Bilkent University with scholarship and graduated from the Industrial Engineering Program of Faculty of Engineering in 2012. She continued her education by pursuing a master degree in the same department. Since then, she has been working with Prof. M. Selim Aktürk and Hande Yaman on her graduate study. She had been on the grant 2210 awarded by The Scientific and Technological Research Council of Turkey (TUBITAK) during her M.S. study.