

Integrated Allocation and Utilization of Airport Capacity to Mitigate Air Traffic Congestion

by

Alexandre Jacquilat

S.M., Applied Mathematics, École Polytechnique (2010)

S.M., Technology and Policy, Massachusetts Institute of Technology (2012)

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

© Massachusetts Institute of Technology 2015. All rights reserved.

Author
Engineering Systems Division
May 18, 2015

Certified by
Amedeo R. Odoni
Professor of Aeronautics and Astronautics, and Civil and Environmental Engineering
Thesis Supervisor

Certified by
Cynthia Barnhart
Chancellor
Ford Professor of Engineering
Committee Member

Certified by
Mort D. Webster
Associate Professor of Energy Engineering, Pennsylvania State University
Committee Member

Certified by
Vikrant Vaze
Assistant Professor, Thayer School of Engineering, Dartmouth College
Committee Member

Certified by
David Gillen
YVR Professor of Transportation Policy, University of British Columbia
Committee Member

Accepted by
Munther A. Dahleh
William A. Coolidge Professor of Electrical Engineering and Computer Science
Acting Director, Engineering Systems Division

Integrated Allocation and Utilization of Airport Capacity to Mitigate Air Traffic Congestion

by

Alexandre Jacquillat

Submitted to the Engineering Systems Division
on May 18, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The combination of air traffic growth and airport capacity limitations has resulted in significant congestion throughout the US National Airspace System, which imposes large costs on the airlines, passengers and society. Absent opportunities for capacity expansion, the mitigation of air traffic congestion requires improvements in (i) the *utilization* of airport capacity to enhance operating efficiency at the tactical level (i.e., over each day of operations), and/or (ii) the *allocation* of airport capacity to the airlines to limit over-capacity scheduling at the strategic level (i.e., months in advance of the day of operations). This thesis develops an integrated approach to airport congestion mitigation that jointly optimizes the utilization of airport capacity and the design of airport capacity allocation mechanisms.

First, we focus on airport capacity utilization. We formulate an original Dynamic Programming model that optimizes, at the tactical level, the selection of runway configurations and the balancing of arrival and departure service rates to minimize congestion costs, for any given schedule of flights. The model integrates the stochasticity of airport operations into a dynamic decision-making framework. We implement exact and approximate Dynamic Programming algorithms that, in combination, enable the real-time implementation of the model. Results show that optimal policies are path-dependent, i.e., depend on prior decisions and on the stochastic evolution of the system, and that the model can reduce congestion costs, compared to advanced heuristics aimed to replicate typical decisions made in practice and to existing approaches based on deterministic queue dynamics.

Second, we integrate the model of airport capacity utilization into a macroscopic queuing model of airport congestion. The resulting model quantifies the relationships between flight schedules, airport capacity and flight delays at the strategic level, while accounting for the way airport capacity utilization procedures can vary tactically to maximize operating efficiency. Results suggest that the model estimates the average departure queue lengths, the variability of departure queue lengths and the average arrival and departure delays at the three major airports in the New York Metroplex relatively well. The application of the model shows that the strong nonlinearities between flight schedules and flight delays observed in practice are captured by the model.

Third, we develop an *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that jointly optimizes scheduling interventions for airport capacity allocation at the strategic level and airport capacity utilization at the tactical level. Scheduling interventions start with a schedule of flights provided by the airlines, and reschedule a selected set of flights to reduce imbalances between demand and capacity, while minimizing interference with airline competitive scheduling. The ICUSM optimizes such interventions, while accounting for the impact of changes in flight schedules on airport operations. It relies on an original modeling architecture that integrates a Stochastic Queuing Model of airport congestion, our Dynamic Programming model of capacity utilization, and an Integer Programming model of scheduling interventions. We develop an iterative solution algorithm that converges in reasonable computational times. Results suggest that substantial delay reductions can be achieved at busy airports through limited changes in airline schedules. It is also shown that the proposed integrated approach to airport congestion mitigation performs significantly better than a typical sequential approach where scheduling and operating decisions are made separately.

Last, we build upon the ICUSM to design, optimize and assess non-monetary mechanisms for scheduling interventions that ensure inter-airline equity and enable airline collaboration. Under the proposed mechanism, the airlines would provide their preferred schedules of flights, their network connections, and the relative scheduling flexibility of their flights to a central decision-maker, who may then consider scheduling adjustments to reduce anticipated delays. We develop a lexicographic architecture that optimizes such interventions based on efficiency (i.e., meeting airline scheduling preferences), equity (i.e., balancing scheduling adjustments fairly among the airlines), and on-time performance (i.e., mitigating airport congestion) objectives. Theoretical and computational results suggest that inter-airline equity can be achieved at no, or small, losses in efficiency, and that accounting for airline scheduling preferences can significantly improve the outcome of scheduling interventions.

Thesis Supervisor: Amedeo R. Odoni

Title: Professor of Aeronautics and Astronautics, and Civil and Environmental Engineering

Acknowledgments

This dissertation is the result of an exciting, instructive and mind-opening journey at MIT.

First and foremost, I would like to express my deepest gratitude to my research advisor, Amedeo Odoni. Amedeo has been a constant source of knowledge, support and guidance since I arrived at MIT. From start to finish, this research has tremendously benefited from his expertise in air transportation and in operations research, as well as from his vision, his openness and his candid encouragements. I am truly thankful for his remarkable dedication to his students, his continuous feedback, and his numerous edits that greatly improved the quality of this dissertation. Beyond his research guidance, I am very fortunate to have found in Amedeo a great mentor who has helped me develop my passions and has made me a better researcher and a better professional.

I would also like to thank the other members of a unique and interdisciplinary doctoral committee. I am greatly indebted to Cynthia Barnhart for her invaluable mentoring. Cindy is one of the busiest persons I know, yet she has always been available when I needed her, she has continuously guided and advised this research, and she has invested significant amounts of time in my personal and professional development. Thank you to Mort Webster for his guidance regarding the strategic directions of this research and the framing of research questions. I have also been fortunate to work with Vikrant Vaze, who invested a lot of time to share his broad expertise and invaluable insights to improve the contents of this thesis. This research has very much benefited from our collaboration, which I hope to sustain in the future. Finally, thanks to David Gillen for expanding the scope of my thinking by raising thought-provoking economic questions. I am grateful for David's mentorship and friendship.

This research has benefited from many interactions with industry practitioners who greatly contributed to the realism and relevance of the models developed in this dissertation. I would like to gratefully acknowledge the Federal Aviation Administration, the Port Authority of New York and New Jersey and the MIT Airline Industry Consortium and its members for funding or advising this research and for inviting me to share this work.

I am grateful to my colleagues and friends from the International Center for Air Trans-

portation for many stimulating discussions about aviation and everything else. I would particularly like to acknowledge Ioannis Simaiakis for helping me develop my research approach and for providing empirical capacity estimates at several US airports, Nikolas Pyrgiotis for helping me identify research opportunities, Michael Wittman for many exciting research conversations in and out of the lab, and Hamsa Balakrishnan for her insightful feedback on various aspects of this research. Thanks also to Joseph Sussman, Christopher Magee and Peter Belobaba for providing exciting teaching opportunities. And many thanks to Philippe Bonnefoy, Vivek Sakhrani and Stephen Zoepf for exciting collaborations which have taught me a lot about domains I was less familiar with.

I am greatly indebted to MIT's Engineering Systems Division, which has been a fantastic academic home for me over the past three years. Thanks to Elizabeth Milnes for making everything possible in E40. To Lita Das, Yinjin Lee, Maite Pena-Alcaraz, Fernando de Sisternes and Abigail Horn for being amazing office mates and for creating a stimulating and fun environment on a daily basis, and to Vivek Sakhrani for the frequent visits and spirited conversations. And to all the members of the vibrant ESD Student Society for the continuous feedback on this research, and for all the fun in and out of the office. I am very fortunate to have been part of this extraordinary and diverse community of passionate colleagues and close friends who have made me a better researcher and a better person.

Finally, I would like to thank my family and my other friends from both sides of the Atlantic. Special thanks to my parents for their long-standing support and for their commitment to creating the best opportunities for me. And, last but very not least, many thanks to my wife, my best friend and my roommate, Claire, without whom none of this would have been possible. Thank you for your understanding and support throughout the tortuous road that led to this dissertation, for your daily refreshing and contagious laughs, and for our exciting past and future adventures together.

Contents

- 1 Introduction 15**
 - 1.1 Congestion in the US National Airspace System 16
 - 1.1.1 Airport Congestion 16
 - 1.1.2 Airport Congestion Mitigation Interventions 19
 - 1.1.3 Research Objectives 24
 - 1.2 Literature Review 25
 - 1.2.1 Models of Airport Congestion 25
 - 1.2.2 Airport Capacity Utilization 27
 - 1.2.3 Airport Demand Management 28
 - 1.3 Contributions of the Thesis 31
 - 1.3.1 Capacity Utilization under Stochastic Operating Conditions 33
 - 1.3.2 An Integrated Model of Airport Congestion 35
 - 1.3.3 An Integrated Approach to Scheduling Interventions 36
 - 1.3.4 Equity and Collaboration in Scheduling Interventions 37
 - 1.4 Organization of the Thesis 38

- 2 Modeling Framework 41**
 - 2.1 Modeling Inputs 42
 - 2.1.1 Airport Capacity 42
 - 2.1.2 Flight Schedules 44
 - 2.2 Congestion Mitigation Interventions 45

2.3	Descriptive Model of Airport Operations	46
2.3.1	Queuing Model of Airport Congestion	46
2.3.2	A Model of Weather Variations	49
2.4	Discussion of the Assumptions of the Models	50
2.5	Experimental Setup	53
3	Capacity Utilization under Operating Stochasticity	59
3.1	Model Formulation	61
3.1.1	Problem Statement	61
3.1.2	State Variables	62
3.1.3	Decision Variables	63
3.1.4	Dynamics of the System	65
3.1.5	Cost Function	69
3.1.6	Dynamic Programming Formulation	69
3.2	Solution Algorithm	70
3.2.1	Experimental Setup	70
3.2.2	Exact Dynamic Programming Algorithm	72
3.2.3	One-Step Look-Ahead Algorithm	73
3.2.4	Evaluation of Performance	75
3.3	Computational Results	78
3.3.1	Optimal Policies	78
3.3.2	Frequency of Decisions	80
3.3.3	Sensitivity of Queue Lengths to Model Parameters	86
3.4	Performance Evaluation	88
3.4.1	Comparison of the Optimal Policy to Heuristics	89
3.4.2	Benefits of the Integration of Queue Stochasticity	92
3.5	Conclusion	93

4	Application to Congestion Modeling	97
4.1	Model Formulation	98
4.1.1	Model Presentation	98
4.1.2	Simplified Control of Service Rates	100
4.2	Model Implementation	103
4.2.1	Measures of Airport On-Time Performance	104
4.2.2	Model of Departure Queue Lengths	107
4.2.3	Model of Arrival and Departure Delays	110
4.2.4	Benefits of the Integrated Approach	113
4.3	Scheduling and On-time Performance Trends	115
4.4	Conclusion	119
5	Integrated Capacity Utilization and Scheduling	121
5.1	Integrated Capacity Utilization and Scheduling Model	122
5.1.1	Model Presentation	122
5.1.2	Model of Scheduling Interventions	126
5.1.3	Queue Length Reduction Constraints	130
5.2	Iterative Solution Algorithm	131
5.2.1	Deterministic Queue Dynamics	131
5.2.2	Collinearity Assumption	132
5.2.3	A Bi-level Iterative Solution Algorithm	133
5.2.4	Size of the Formulation	137
5.3	Computational Results	137
5.3.1	Convergence of the Iterative Algorithm	138
5.3.2	Optimal Schedules and Delays	141
5.3.3	Sensitivity of Displacement to Queue Length Targets	145
5.4	Benefits of Integration	147
5.5	Conclusion	151

6	Inter-airline Equity and Collaboration in Scheduling Interventions	153
6.1	Motivation	154
6.2	Multi-criteria Modeling Architecture	156
6.2.1	Performance Attributes	157
6.2.2	Lexicographic Modeling Approach	159
6.2.3	Solution Architecture	162
6.3	A Theoretical Discussion on Inter-airline Equity	166
6.3.1	Cases of Joint Maximization of Efficiency and Equity	166
6.3.2	Cases of Efficiency/Equity Trade-off	174
6.4	Mechanisms for Airport Scheduling Interventions	177
6.5	Computational Results	180
6.5.1	Inter-Airline Equity: The ESM-IN	181
6.5.2	Network Connectivities: The ESM-SN	184
6.5.3	Credit Allocation: The ECSM-SN	188
6.5.4	Summary	190
6.6	Conclusion	192
7	Conclusion	195
7.1	Summary of Research	196
7.2	Practical Implications	199
7.3	Future Research	202
A	Airport Diagrams	207
B	Tail Number Reconstruction	211
C	Scheduling Interventions at LGA	215

List of Figures

1-1	Enplanements by US airlines [1] and airline operations at 7 airports [43] . . .	17
1-2	Scheduling profiles and capacity estimates at FRA and JFK	22
1-3	Integrated approach to airport congestion mitigation	32
2-1	A schematic representation of modeling inputs and decisions	42
2-2	Representation of the airport as a queuing system	47
2-3	State-transition diagram of the $M(t)/E_k(t)/1$ queuing system	48
2-4	Arrival and departure schedules at JFK, EWR and LGA on busy days in 2007	55
2-5	Capacity estimates [104] and 2007 scheduling levels at JFK, EWR and LGA	57
3-1	Schematic representation of capacity utilization decisions	62
3-2	Mapping between queue states and queue lengths	66
3-3	Representation of a runway configuration change at the beginning of period t	67
3-4	Operational Throughput Envelopes [104] and scheduling levels on 05/25/2007	71
3-5	Representation of the solution architecture	74
3-6	Total expected congestion costs with the three policies, under weather updates	77
3-7	Optimal runway configuration and arrival rate at 12:00 ($\tau_I = 5$ min, $ws_t = 1$)	79
3-8	Use of runway configurations for different values of τ_I	82
3-9	Use of runway configurations with uniform vs. differentiated values of τ_I . .	83
3-10	Sensitivity of expected arrival and departure queues to α ($\tau_I = 5$ minutes) .	87
3-11	Sensitivity of expected arrival and departure queues to τ_I ($\alpha = 1$)	88
4-1	Inputs with Operational Throughput Envelopes and constant service rates .	100

4-2	Breakdown of scheduled and actual flight times	105
4-3	Average departure queue lengths in Summer 2007	109
4-4	Variability of departure queue lengths in Summer 2007	110
4-5	Predicted and actual delays at JFK, EWR and LGA in Summer 2007	112
4-6	Modeled delays with and without the control of service rates in Summer 2007	114
4-7	Average number of flights scheduled per hour in Summer 2007, 2008 and 2010	116
4-8	Predicted and actual average delays from Summer 2007 to Summer 2011 . .	117
5-1	Modeling Architecture	125
5-2	Convergence of the algorithm: displacement and queue lengths per iteration	139
5-3	Original and modified schedules of flights ($A_{MAX} = 11$ and $D_{MAX} = 15$) . . .	144
5-4	Expected queue lengths before and after the scheduling interventions	145
5-5	Sensitivity of the optimal displacement as a function of $A_{MAX} = D_{MAX}$. . .	146
5-6	Integrated and sequential approaches to airport congestion mitigation	148
5-7	Expected queue lengths under the integrated and sequential approaches . . .	151
6-1	A schematic trade space between weighted efficiency and equity	162
6-2	Solution architecture	165
6-3	Trade-off between efficiency and equity due to variations in airline schedules	175
6-4	Trade-off between efficiency and equity due to variations in airline connectivities	176
6-5	Trade-off between efficiency and equity due to variations in flight valuations	177
6-6	ESM-SN Results: Displacement per airline under P1 , P2 , and $\widehat{\mathbf{P3}}(\rho)$	187
6-7	Airline utilities when flight valuations are considered (in full) vs. ignored . .	191
A-1	Airport Diagram: JFK	208
A-2	Airport Diagram: EWR	209
A-3	Airport Diagram: LGA	210
C-1	Flight schedule at LGA on 07/17/2014	216
C-2	Original and modified schedules at LGA	217
C-3	Expected queue lengths at LGA before and after the scheduling interventions	218

List of Tables

1.1	Classification of congestion mitigation interventions	19
3.1	Definition of wind states at JFK: Runways that can be used per wind state .	63
3.2	Set of JFK runway configurations that can be selected per wind state	64
3.3	Total expected congestion costs with the three policies under schedule updates	76
3.4	Policy frequency for six different periods of the day	85
3.5	Expected total congestion costs for different values of τ_I	88
3.6	Relative error of the two heuristics	92
3.7	Relative performance of the optimal stochastic and deterministic policies . .	94
4.1	Runway configuration clusters at JFK, EWR and LGA	101
4.2	Size of the original and simplified controls	102
4.3	Statistics on modeled and actual departure queue lengths	108
4.4	Statistics on modeled and actual delays	111
4.5	Percent of change in demand and delays between 2007 and 2011	118
5.1	Size of the model	137
5.2	Computational results	140
5.3	Optimal displacement and peak queue lengths for the 9 days considered . . .	142
5.4	Displacement and delays for different expected queue length targets	143
5.5	Optimal displacement under the integrated and sequential approaches	149
6.1	Results of the BSM-IN and ESM-IN	183

6.2	ESM-SN results: Inequity Φ , price of equity and efficiency	186
6.3	ECSM-SN results: Displacement per airline, price of equity and efficiency . .	190
C.1	Displacement and delays at LGA for different expected queue length targets	217

Chapter 1

Introduction

The development and growth of air transportation worldwide has been supported by large-scale infrastructure, comprising airports, air traffic control sectors, and air traffic management systems. Over the past decades, however, many air transportation systems have been facing critical challenges to accommodate increasing demand in the face of capacity limitations. This results in significant imbalances between demand and capacity, and ultimately in air traffic congestion, increased greenhouse gas emissions, and lost demand.

This thesis designs, optimizes, and evaluates airport congestion mitigation interventions to promote more efficient, more reliable, and more sustainable air transportation systems. We develop innovative managerial and operational procedures to improve the utilization of airport capacity at the tactical level and the design of scheduling interventions mechanisms allocating airport capacity to the airlines at the strategic level. By “tactical”, we refer to the use of airport infrastructure to process flights over each day of operations. By “strategic”, we refer to the planning of flight schedules well before the day of operations (several months in advance), taking into consideration long-term patterns of capacity availability. As we shall detail, the strategic problem of capacity allocation (flight scheduling) and the tactical problem of capacity utilization (airport operations) are strongly interdependent, and we develop an integrated approach that addresses them jointly. This thesis focuses on the US National Airspace System, and applies this modeling approach to the three airports in

the New York Metroplex: John F. Kennedy International Airport (JFK), Newark Liberty International Airport (EWR) and LaGuardia Airport (LGA). Its results then inform the design of demand management policies at busy airports in and outside the United States.

In this chapter, we first discuss the problem of airport congestion mitigation in the US National Airspace System. In Section 1.2, we review the existing literature, focusing on the available modeling tools that can be used to quantify airport on-time performance, the optimization approaches developed to improve the utilization of airport capacity, and the economic and managerial approaches to airport demand management. In Section 1.3, we present the main contributions of the thesis, namely (i) the development of an integrated approach to airport congestion mitigation; (ii) the development of an integrated modeling architecture that combines different stochastic modeling and optimization components; and (iii) the practical implications of the results for the management and operation of air transportation infrastructure. Finally, Section 1.4 outlines the structure of the thesis.

1.1 Congestion in the US National Airspace System

1.1.1 Airport Congestion

Over the past decades, demand for air traffic has grown significantly, both in the United States and internationally. This trend was initiated by the deregulation of the US airline industry in 1978, and sustained by the growth, urbanization, and globalization of the world's economy, as well as the development of new aircraft technologies, new services offered by the airlines, reductions in airfares, etc. Figure 1-1 shows the number of passenger enplanements by US airlines, as well as the evolution of the number of flights operated by commercial airlines at the seven busiest US airports since 1990. Note, first, that passenger enplanements have grown by over 60% in aggregate since 1990. The declines in demand during the post-9/11 air transportation crisis and the economic downturn in 2008 and 2009 have been relatively small in comparison to the substantial growth in air traffic over the past 25 years. Moreover, this overall demand growth does not impact all individual airports uniformly. In-

stead, traffic at some airports has grown extensively (e.g., at JFK Airport), while traffic at other airports has remained stagnant or declined (e.g., at Chicago’s O’Hare Airport (ORD)). In other words, whereas overall traffic has followed a relatively smooth upward trend, demand at each airport exhibits significant variations from year to year, as a function of variations in local passenger demand, in airline business strategies, in airport access regulations, and in airport performance. These trends underscore that the US National Airspace System as a whole, and some individual airports in particular, have had to scale up to accommodate rapidly growing demand.

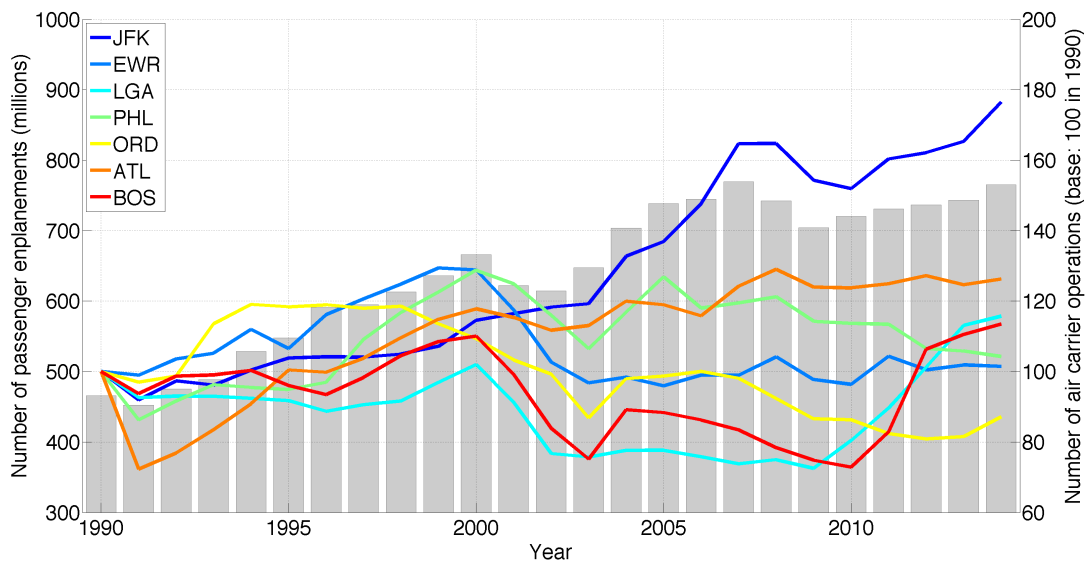


Figure 1-1: Enplanements by US airlines [1] and airline operations at 7 airports [43]

At the same time, the capacity of the US National Airspace System is limited. First, safety procedures impose separation requirements between aircraft at the airports, in terminal airspaces, and in the en-route airspace, and thus constrain throughput throughout the system. Second, system throughput is limited by the available infrastructure. The main bottlenecks of operations are the runway systems of the busiest airports [34]. Airport operations are constrained by the number of runways and the geometric layout of the runway system. Moreover, all runways are typically not simultaneously activated. Instead, airports operate in a given *runway configuration*, defined as the set of runways that are simultaneously

active to operate arrivals and the set of runways that are simultaneously active to operate departures. For instance, only two or three runways of JFK's four runways are generally simultaneously active.

The coupling of demand growth and capacity limitations results in air traffic congestion through the National Airspace System, which materializes in the form of aircraft queues, flight delays and cancellations, and missed connections by passengers. The economic costs of these delays are undoubtedly significant, especially when compared to the meager profit margins that the airline industry, as a whole, has been achieving over the past decade. The nation-wide impacts of air traffic delays in the United States were estimated at roughly \$36 billion in 2007. Of these costs, approximately 25% were direct costs to the airlines, 50% were borne by passengers and 25% represented other welfare losses to society, including lost demand, environmental impacts due to increased emissions, and a negative impact on productivity [5]. In comparison, the total annual profits of the US airline industry have exceeded \$5 billion in only three years since 2000.

Most of flight delays are caused by the imbalances between demand and capacity resulting from airlines scheduling more flights than available capacity at the airports. The Bureau of Transportation Statistics estimated that roughly 50% of system-wide delays can be attributed to heavy traffic volume, broadly defined. This includes delays due to more flights being scheduled than optimal airport capacity, as well as delays due to the combination of scheduling levels close to optimal capacity and capacity shortages related to non-extreme weather conditions or other inefficiencies in airport operations. Another 25% of system-wide delays are due to the propagation of disturbances within the network of airline operations (e.g., late-arriving aircraft), themselves primarily created by imbalances between demand and capacity at upstream airports. The remaining 25% is due to other causes, including inefficiencies in airline operations, mechanical failures and extreme weather conditions [26]. The mitigation of congestion in the National Airspace System can thus be primarily achieved by reducing the imbalances between demand and capacity at the busiest US airports.

1.1.2 Airport Congestion Mitigation Interventions

Airport congestion mitigation interventions fall into three categories, spanning from long-term to short-term interventions: capacity planning, capacity allocation, and capacity utilization. These interventions are outlined in Table 1.1 and detailed below.

Table 1.1: Classification of congestion mitigation interventions

Intervention	Objective	Example
Capacity planning	Increasing system capacity (1 year – 30 years)	Runway construction New ATM technologies
Capacity allocation	Limiting, or not, overscheduling (3 months – 1 year)	Do-nothing (Level 1) Schedule coordination (Level 2 or 3) Congestion pricing, slot auctions
Capacity utilization	Enhancing operating efficiency (1 minute – 1 day)	Air Traffic Flow Management Air Traffic Control

Capacity Planning

First, capacity planning aims to increase system capacity through the expansion of the physical or technological infrastructure of the system. Physical infrastructure expansion may involve brownfield expansion (e.g., the construction of new runways or the expansion of terminal buildings) to increase the capacity of existing airports, or the development of greenfield airports (e.g., the development of multi-airport systems [22]). In the United States, 17 major airport runway expansion projects have been completed between 2000 and 2013 [101]. Outside the United States, recent examples of capacity expansion include the construction of a fourth runway at Frankfurt Airport (FRA) and the ongoing infrastructure expansion plan in the London Airport System. Whereas infrastructure expansion enables significant capacity increases, it is also investment-intensive and spans several decades from conception to completion. Most important, it might be unfeasible in the densest urban areas and terminal airspaces because of environmental, socioeconomic and political constraints.

The capacity of air transportation systems can also be increased through the expansion of their technological infrastructure. Most notably, the Next Generation Air Transporta-

tion System (NextGen) uses new satellite-based aircraft surveillance technologies to better monitor and manage aircraft operations. Its implementation in the US National Airspace System would result in better sequencing and spacing of aircraft in the en-route airspace and in terminal airspaces [42]. The objectives of the Single European Sky ATM Research (SESAR) system in Europe are similar [102]. However, the development and deployment of such technologies are also investment-intensive, require the participation of diverse stakeholders, and may thus take many years to be effective. Moreover, they may not be sufficient to significantly scale the capacity of the busiest airports to meet demand growth.

For these reasons, there is little prospect for short-term quantum capacity increases at the busiest US airports and most of the busiest airports elsewhere. In this case, the mitigation of air traffic delays requires improvements in the allocation and the utilization of air transportation infrastructure capacity.

Capacity Allocation

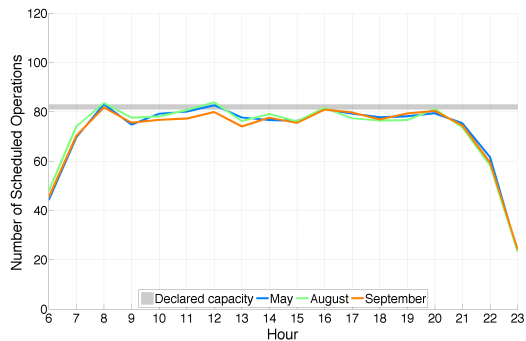
The allocation of airport capacity refers to the *demand management* policies that govern the access to airport infrastructure by commercial airlines and other aircraft operators, and, in turn, the scheduling of flights at the airports. Demand management measures can take the form of unregulated access, or administrative (e.g., slot control) or economic (e.g., congestion pricing, slot auctions) mechanisms. The generic objectives of demand management consist of balancing the need for limiting over-capacity scheduling to prevent delays from reaching undesirable levels, on the one hand, and capturing the economic and social benefits associated with high airport throughput and high scheduling levels, on the other. The “scheduling interventions” studied in this thesis fall within the realm of airport demand management.

Demand management practices differ significantly between airports located in the United States and those elsewhere. The overwhelming majority of busy airports outside the United States are subject to *slot control policies*, or “Level 3 coordination”. Slot-controlled airports declare a value of capacity and allocate administratively a corresponding number of slots to air carriers. Slot allocation takes place through a bi-annual administrative procedure

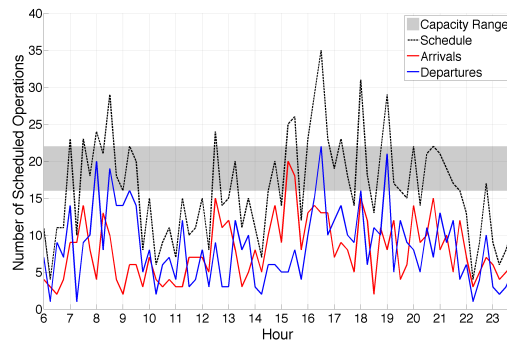
conducted under the aegis of the International Air Transport Association (IATA), relying on a combination of grandfathering rights, use-it-or-lose-it rules, and slot pools for new entrants [58]. The value of the declared capacity typically either remains constant (e.g., “80 scheduled flights per hour”) or varies little throughout the day. Moreover, the value of declared capacity is typically conservative and corresponds to the average number of flights that can be operated under poor weather conditions. In addition, a number of mildly congested airports outside the United States are subject to *schedule facilitation*, or “Level 2 coordination”, under which the airlines submit their scheduling requests to a schedule facilitator, who may then propose some scheduling adjustments to reduce anticipated delays [58]. In contrast, flight schedules have been weakly constrained at US airports since the phase-out of the High Density Rule, effective on January 1, 2007 (49 USCS 41715). At the vast majority of airports, no airport- or government-mandated restriction on runway access is imposed, an approach sometimes referred to as “Level 1 coordination” [85].

To illustrate these differences, Figure 1-2 shows schedules of flights at Frankfurt Airport (FRA) and at JFK Airport from 2007, as well as estimates of the capacities of these airports. As a result of the strict enforcement of the slot limits, the number of hourly flights scheduled at FRA never exceeds the value of the declared capacity, which is set between 81 and 83 scheduled flights per hour (Figure 1-2a). Moreover, strong demand at a schedule-coordinated airport such as FRA results in evenly distributed schedules of flights throughout the day. In contrast, the absence of schedule coordination at JFK in 2007 is reflected in the high variability of the demand. The schedule of flights at JFK is uneven with significant morning and afternoon peaks, and valleys at off-peak hours. At times during the day, more flights are scheduled than even the optimal capacity of the airport, thus resulting almost certainly in imbalances between demand and capacity. Obviously, when capacity is sub-optimal (e.g., under poor weather conditions), these imbalances become even more significant.

These differences in demand management policies and in the resulting scheduling patterns have significant impacts on airport performance. First, airport throughput is typically higher at US airports than at their European counterparts, especially under good weather



(a) Scheduling at FRA



(b) Scheduling at JFK

Figure 1-2: Scheduling profiles and capacity estimates at FRA and JFK

conditions. Moreover, the high scheduling levels at peak hours provide significant economic benefits to the airlines and passengers. In contrast, slot control policies, as currently practiced, may result in the under-utilization of scarce airport capacity when good weather conditions prevail, as well as barriers to entry or growth in competitive markets. On the other hand, on-time performance is generally better at European airports, where flight delays have been found to be both lower and less variable (thus, more predictable) than at US airports. In other words, the approach at slot-controlled airports puts a premium on schedule reliability, while the US approach puts a premium on capacity utilization [80, 86].

Since 2007, the design of demand management rules at busy US airports has been an ongoing policy issue. Limits on the number of flights scheduled per hour or 30 minutes (“flight caps”) were imposed at the 3 airports in the New York Metroplex (at LGA in 2007 and at JFK and EWR in 2008). Flight schedules are then determined through a process similar to schedule facilitation, or Level 2 coordination, with the Federal Aviation Administration (FAA) playing the role of the schedule facilitator. However, the flight caps in New York have been applied on an *ad hoc* basis, often relying on voluntary compliance by the airlines, and have been criticized as being too high to effectively mitigate congestion. It was thus recommended that they be reexamined [87, 49, 34].

Capacity Utilization

Finally, improvements in capacity utilization falls within the realm of Air Traffic Control (ATC) and Air Traffic Flow Management (ATFM) [7, 119, 34]. These two broad categories of interventions aim to design procedures to operate scheduled flights in a way that minimizes the magnitude and/or the costs of flight delays. ATC refers to the operation of each individual aircraft on the ground, in terminal airspaces, and through en-route sectors with the two major objectives of maximizing throughput and avoiding aerial and ground conflicts. These are in-flight interventions that involve the control of aircraft speed and flight trajectories, relying on surveillance data, automation systems, and communications between air traffic controllers and airline operators. ATFM refers to the control of the flows of aircraft at a given airport or in a network of airports to minimize local imbalances between demand and capacity. Whereas ATC interventions are microscopic (i.e., at the level of each individual aircraft) and fall within a scope of seconds to minutes, ATFM interventions are mesoscopic (i.e., at the level of aircraft flows) and take place within a scope of 15 minutes to several hours. Typical ATFM procedures include, ranging from interventions implemented several hours in advance to interventions implemented on a 15-minute basis: (i) the ground holding of aircraft through the initiation of Ground Delay Programs; (ii) the re-routing, speed control and airborne holding of en-route aircraft; (iii) the activation of runways at the airports and the balancing of the number of arrivals and departures operated; and (iv) the control of surface operations and the metering of departing aircraft. Our research falls within the third category, i.e., the utilization of runway infrastructure at a given airport.

Note that ATFM and ATC interventions are applied over the course of each day of operations, i.e., after the schedule of flights has been issued by the airlines and tickets got marketed to air travelers. Therefore, they mostly aim at minimizing the costs of air traffic delays *for a given schedule of flights*. In combination, they have resulted in significant improvements in the efficiency of air traffic operations, but they may be insufficient to mitigate air traffic congestion significantly if the number of flights scheduled exceeds the capacity at some airports or at some air traffic control sectors by any significant margin.

1.1.3 Research Objectives

The main objective of this research is to develop a comprehensive approach to the allocation and utilization of airport capacity in order the design, optimize, and assess airport congestion mitigation interventions in the National Airspace System.

Specifically, this research addresses the two following problems:

- (i) The optimization of airport capacity utilization procedures. We develop a dynamic decision-making support tool that optimizes two of the most important capacity utilization interventions at any given airport, namely the *sequencing of runway configurations* and the *balancing of the arrival and departure service rates*. These are *tactical* interventions that are applied over each day of operations, in each time period when the airport is operating.
- (ii) The design of scheduling interventions mechanisms to allocate airport capacity to the airlines. We consider a scheduling process under which the airlines submit their scheduling preferences to a central authority (e.g., to schedule facilitators at “Level 2” airports, to schedule coordinators at “Level 3” airports), who may then suggest some adjustments in the proposed schedule of flights to reduce imbalances between demand and capacity. We develop decision-making models that optimize such flight rescheduling. These are *strategic* measures that are implemented several months in advance of the day of operations.

The research reported in this thesis is motivated by the interdependencies between flight scheduling and airport operations. For any set of prevailing operating conditions, the optimal capacity utilization decision in some particular time period of the day (i.e., the optimal runway configuration to be used and the optimal combination of arrivals and departures to be served) and the resulting arrival and departure delays will depend not only on the number of arrivals and departures scheduled for that period, but also on the demand (number of arrivals and departures) in the preceding and following time periods. Conversely, the optimal schedule of flights (and the design of demand management mechanisms) will depend

on the way arrivals and departures can be served at the airport, under different operating scenarios. As we shall develop in the next section, these two problems have been mostly treated in isolation in previous research.

1.2 Literature Review

In this section, we review the three most relevant areas of research, namely: (i) descriptive models of airport operations; (ii) models of airport capacity utilization; (iii) models of airport demand management. These three areas have been the focus of extensive research, but limited effort has been made to date to integrate them into a unified framework. First, airport operations have been extensively described by means of *stochastic* queuing models, but models of airport capacity utilization are commonly based on *deterministic* queuing models. Second, models of airport capacity utilization treat the schedule of flights as given, while models of airport demand management consider typically simplified operational settings and fail to capture the endogeneity of airport capacity utilization, i.e., the impact of flight schedules on airport operating procedures and on airport congestion. This thesis addresses these two gaps in the existing literature.

1.2.1 Models of Airport Congestion

Models of airport on-time performance estimate the extent of congestion at a given airport or in a network of airports as a function of flight schedules, airport capacity, and operating procedures. They fall into three categories: microscopic, mesoscopic and macroscopic models. Microscopic models consider each aircraft individually and reproduce precisely the physical layout of the airport and the sequencing of movements [21, 111, 45]. These models can evaluate some Air Traffic Control procedures, but are not well-suited to performing strategic planning under a wide range of scenarios. Mesoscopic models predict runway delays and taxiing times using operational data, such as the runway configuration in use, short-term arrival demand, pushback schedules, etc. [103, 93, 106]. These models have been used to

improve surface operations procedures [107, 66]. However, their heavy reliance on detailed operational data limits their applicability to modeling delays for strategic planning purposes when such data are not yet available. Finally, macroscopic models aggregate operations at the airport level to provide computationally efficient estimates of the relationships between flight schedules, airport capacity and flight delays in support of strategic planning (*e.g.*, to assess the benefits of capacity expansion or demand management). Our research falls within this third category.

Macroscopic models of congestion are based on econometric models [70, 82, 122], deterministic queuing models [53], stochastic queuing models [67, 50, 95], or a combination thereof. In this research, we consider a state-of-the-art stochastic queuing model, specifically an $M(t)/E_k(t)/1$ model that represents the demand and service processes by means of Poisson and Erlang processes, respectively. This model aims to capture the dynamics of formation and propagation of delays over the day as well as the uncertainty and variability associated with airport operations. Previous research has shown that this model approximates well queuing dynamics at major US airports [97, 76, 59]. Moreover, this model has been found to capture the nonlinear relationship between the number of flights scheduled and flight delays observed in practice when airports operate close to capacity [39, 96].

The delay estimates obtained with any queuing model depend critically on the estimates of the rates at which arrivals and departures join the system and are serviced (*e.g.*, the rates of the Poisson and Erlang processes for an $M(t)/E_k(t)/1$ model). Demand rates are determined by the schedule of flights, and may be subject to dynamic updates in the case of operational disturbances. Service rates are constrained by the capacity of the airport. In existing models, service rates are generally kept constant over the day [94, 59] or varied using *ex post* operational data, such as meteorological conditions and reported capacity estimates [54]. Recent clustering techniques generate capacity profiles based on probabilistic weather forecasts that are used *ex ante* in queuing models [74, 27, 84]. These variations in service rates are exogenous and do not depend on operating procedures at the airports.

However, other important variations in arrival and departure service rates are *endoge-*

nous, i.e., they depend on how the airport tactically adapts to the schedules of flights and to observed queue lengths. For instance, if a large number of landings are scheduled, then air traffic managers may decide to enhance the arrival throughput, at the expense of the departure throughput. As well, the arrival throughput may be enhanced if the observed arrival queue is longer than expected, or if the departure queue is shorter than expected. These controls are not taken into account in existing macroscopic models of airport congestion, although they might affect the dynamics and the magnitude of delays. One recent application of a queuing model included variations in arrival and departure service rates that occur in response to changes in daily flight schedules, but these variations are introduced manually and not systematically [96].

1.2.2 Airport Capacity Utilization

The optimization of runway capacity utilization procedures at a given airport has been an important topic in Air Traffic Control (ATC) and Air Traffic Flow Management (ATFM). ATC interventions for congestion mitigation include mostly the optimization of the sequencing and spacing of aircraft on a “real-time” basis [3, 110]. Such interventions take advantage of knowledge of the types of aircraft waiting to land and takeoff at any given time. However, these data are not available when it comes to optimizing flight schedules several months in advance, or even ATFM procedures several hours in advance. Alternatively, ATFM interventions optimize the arrival and departure loads allocated to the different runways of the airport per 15-minute period, or per hour. Our research falls within this second category.

The first study that formulates and solves the trade-off between arrival and departure throughput for congestion mitigation is the seminal paper of Gilbo [46]. He first introduced the concept of the “Capacity Envelope”, which represents the non-increasing relationship between arrival and departure capacities. Then, he developed a Linear Programming model that determines the optimal number of arrivals and departures to serve per 15-minute period. The model minimizes a congestion cost function that depends on arrival and departure queue lengths, under linear capacity constraints. This approach was then extended to account for

arrival and departure fixes [47], for airline’s operational preferences and airline collaboration in the development of Ground Delay Programs [51, 48], and for nonlinear capacity constraints and objective functions [35].

More recent studies addressed the problem of *jointly* selecting runway configurations and arrival and departure service rates by introducing different Capacity Envelopes for different runway configurations. Optimizing these decisions jointly is critical when different runway configurations achieve different arrival and departure throughput. This may occur if, for instance, a particular configuration uses one runway to operate arrivals and two runways to operate departures, while an alternative configuration uses two runways to operate arrivals and one runway to operate departures. Bertsimas et al. developed a Mixed Integer Program that determines the optimal schedule of runway configurations and of arrival and departure service rates for an entire day of operations [18]. This model was extended to account for marginally decreasing penalties associated with runway configuration switches [120]. Li and Clarke developed an alternative formulation that dynamically controls runway configurations under stochastic wind conditions [73]. This literature has demonstrated that the optimal selection of runway configurations and of the balance of arrivals and departures serviced can lead to significant congestion cost savings.

However, all existing approaches to the selection of runway configurations and the balancing of arrival and departure service rates are based on *deterministic* queue dynamics. This contrasts with the literature on airport congestion modeling (Section 1.2.1) that relies on *stochastic* queue dynamics extensively. In turn, the existing approaches may not be able to account for the strong non-linearities in the system when the airport operates close to its capacity, which are captured only by a stochastic queuing model.

1.2.3 Airport Demand Management

The literature on airport demand management has been motivated by (i) the inefficiencies in the current IATA slot allocation process, which constrains airline competition and creates barriers to entry [83, 31, 4, 34], and (ii) the significant costs of air traffic congestion in the

US National Airspace System (Section 1.1.1). This has been an important topic both in transportation economics and in operations research/management science.

Economic studies have proposed airport congestion pricing relatively early to correct delay externalities [28], but the design of congestion tolls is complicated by several features of the airline industry. First, market concentration may provide incentives for delay internalization by the airlines, suggesting that optimal tolls decrease with airline market share [24, 38]. Second, the presence of market power may result in suboptimal output, suggesting that optimal tolls decrease with market concentration [90]. Third, the hub-and-spoke structure of air traffic operations creates the scope for positive network externalities, suggesting that optimal tolls decrease with the extent of hubbing at the airport [77]. From an empirical standpoint, the econometric evidence on the self-internalization of delays by the airlines as a function of market share, market concentration, and hubbing is mixed and ongoing [32, 81, 33, 100, 2]. Given these complexities, a first-best outcome can be achieved in theory through a differentiated tolling scheme (i.e., a congestion pricing mechanism that imposes different congestion tolls to different airlines), but such a mechanism may be undesirable in practice. The literature has then focused on the evaluation of alternative demand management schemes (e.g., uniform congestion prices, slot control, slot auctions) [25, 12, 30]. These studies provide important insights on the economic performance of demand management. However, they typically consider simplified operational settings and do not capture the complexity and variability of airport operations and of the networks of flights that airlines operate.

A more relevant stream of research involves modeling the effects of demand management on airline schedules and airport congestion to improve practices and policies at busy airports. At schedule-coordinated airports, research is primarily concerned with improving the efficiency of the slot allocation process, defined as the ability of the process to meet airline scheduling preferences. Mixed Integer Programs have been proposed to optimize the coordination of flight schedules and accommodate airline preferences better through primary slot allocation [123] or secondary slot trading [89].

At US airports, evidence suggests that airport performance could be improved through

scheduling interventions. First, the nonlinearities in the relationship between flight schedules and flight delays indicates that small scheduling changes can have significant impacts on on-time performance [38, 94, 59]. Second, the comparison of actual schedules of flights to a schedule obtained from the perspective of a single hypothetical monopolistic airline suggests that air traffic delays are not unavoidable under current levels of passenger demand and current level-of-service (e.g., service frequency, number of connections, etc.) [72, 117]. Third, recent evidence indicates that the busiest US airports are over-scheduled, i.e., that the costs of airport congestion may exceed the benefits of high scheduling levels. Swaroop et al. quantified the trade-off between reductions in flight delays and the costs imposed by scheduling constraints on air travelers, and found that small reductions in allocated capacity could improve passenger welfare [114]. Additionally, Vaze and Barnhart used a game-theoretic framework of airline frequency competition and found that slot control and congestion pricing could mitigate congestion and improve airline profitability [115, 116]. Fourth, Pyrgiotis and Odoni simulated the effects of scheduling limits at the busiest US airports on airline schedules by minimizing the displacement from an original schedule of flights planned in the absence of demand management. They demonstrated that large delay reductions could be achieved under “mild” scheduling constraints while keeping airline networks of flights and passenger itineraries unchanged [96]. In short, evidence suggests that even modest scheduling interventions could have significant benefits on the US National Airspace System.

However, determining the “optimal” level of scheduling interventions (e.g., optimal scheduling limits) remains an open question. The aforementioned studies evaluate typically the impacts of *given scheduling limits* on airline schedules, passenger itineraries, and airport congestion. These limits are generally determined *ex ante*, and do not consider the complexity of operating capacity constraints. Moreover, they are typically kept constant throughout the day, while airline scheduling, passenger demand, and airport queuing dynamics exhibit intra-day variations. Recent stochastic programming models look for optimal scheduling profiles (i.e., the optimal number of scheduled flights per hour, aggregated across all airlines) at slot-controlled airports given weather-related capacity uncertainty, as a function of

time-dependent flight valuations, delay costs and cancellations costs [29]. But the quantification of the benefits (delay reductions) and the costs (scheduling adjustments) of demand management is rendered difficult by the complex interactions between airline schedules of flights and airport operations. The integrated approach to airport scheduling interventions and airport operations developed in this research aims to address this difficulty.

1.3 Contributions of the Thesis

The main contribution of this thesis is the development of the first integrated approach to airport congestion mitigation that jointly optimizes the utilization of airport capacity and the design of scheduling interventions mechanisms for airport capacity allocation. This is illustrated in Figure 1-3. The approach uses, as a starting point, a Stochastic Queuing Model of airport operations that quantifies flight delays as a function of demand and service rates, themselves a function of flight schedules and airport capacity, respectively. It integrates this *descriptive* model of airport operations into *decision-making* models of congestion mitigation. It optimizes, first, airport capacity utilization procedures at the tactical level, subject to capacity constraints. It then optimizes scheduling interventions for airport capacity allocation at the strategic level, subject to scheduling constraints, network connectivity constraints and delay reduction targets.

The development of this integrated approach follows a four-step structure:

1. We develop a new decision-making model of airport capacity utilization under stochastic airport queuing dynamics and stochastic operating conditions. It is formulated as a Dynamic Programming model.
2. We integrate this model of capacity utilization into a macroscopic Stochastic Queuing Model of airport congestion.
3. We develop an *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that integrates the Dynamic Programming model of airport capacity utilization and the

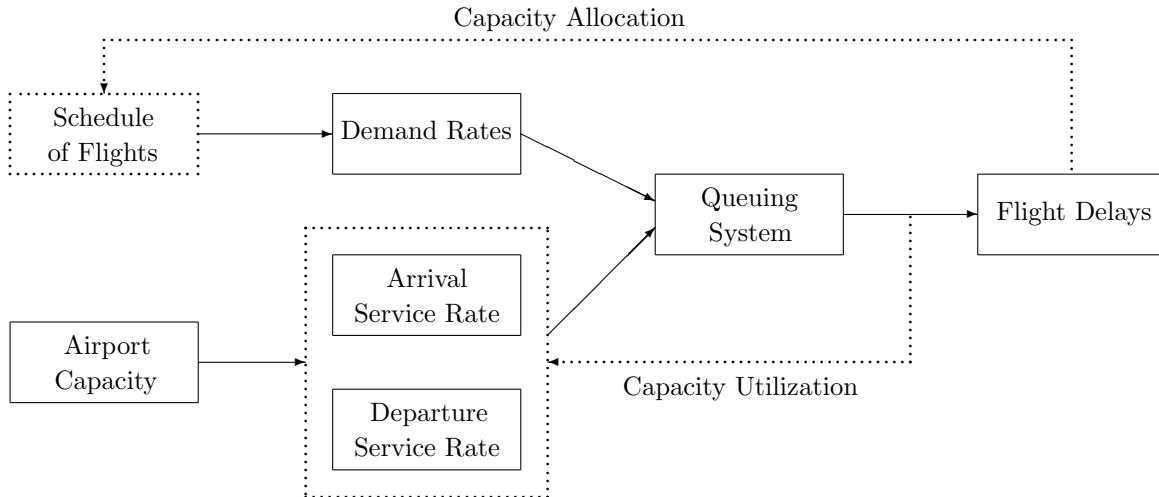


Figure 1-3: Integrated approach to airport congestion mitigation

Stochastic Queuing Model of airport congestion into an Integer Programming model of scheduling interventions.

4. We apply this integrated modeling architecture to design and evaluate alternative scheduling interventions mechanisms.

From a methodological standpoint, this thesis provides a computationally tractable approach to address a class of problems where scheduling decisions spanning several future time periods must be made in advance, and operating decisions are made subsequently in real time. Given their different time scales, these decisions are based on different types of information. Scheduling decisions are produced before operating uncertainty is resolved, thus relying solely on historical/statistical information about the tactical operation of the system. In contrast, operating decisions are made sequentially at a later time as information is dynamically revealed. The approach developed here integrates a descriptive model of the system (here, a Stochastic Queuing Model of congestion), a Dynamic Programming control of resource utilization, and an Integer Programming model of scheduling.

From a practical standpoint, this thesis provides decision-making support for optimizing delay-mitigating runway capacity utilization procedures and scheduling interventions at congested airports. The tactical capacity utilization part selects jointly and dynamically the

runway configuration to be used and the balance of arrival and departure service rates to be applied over each day of operations to minimize congestion costs, as uncertainty regarding operating conditions gets resolved over time. The decision-making tool developed in this thesis can be integrated into Air Traffic Management systems. The scheduling interventions part proposes minimal modifications in airlines’ preferred schedules of flights several months in advance of the day of operations, while anticipating how capacity utilization procedures will react to such scheduling changes at the tactical level. This thesis provides a flexible and technically sophisticated approach to optimize scheduling modifications aiming to reduce the number of flights scheduled at peak hours while minimizing interference with airline competitive scheduling, without relying on an arbitrary, administrative notion of “declared capacity”. At “Level 2” airports and at congested US airports, it can be applied directly to support and optimize the rescheduling of flights. At schedule-coordinated, “Level 3” airports, it could significantly improve the efficiency of the slot allocation process. The implementation of this approach relies on a collaborative environment between the central decision maker (e.g., the Federal Aviation Administration (FAA) in the United States, or the airport schedule facilitators or coordinators outside the United States) and the airline community. The latter would provide their scheduling preferences, and the former would then propose the scheduling adjustments. Such a collaborative environment already exists to coordinate successfully Ground Delay Programs at the day-to-day, tactical level in the United States [118, 9], and could be extended to this type of scheduling interventions.

1.3.1 Capacity Utilization under Stochastic Operating Conditions

First, we formulate and solve the problem of runway capacity utilization under airport operating stochasticity. This intends to bridge the gap in the literature between descriptive models of airport congestion based on stochastic queue dynamics and decision-making models of runway capacity utilization based on deterministic queue dynamics.

We formulate an original Dynamic Programming model of airport capacity utilization. The model takes as inputs airport capacity estimates and the schedule of flights for a given

day of operations. At the beginning of each 15-minute period of the day, it selects jointly the runway configuration and the balance of arrival and departure service rates to minimize congestion costs, as a function of a state variable that includes: (i) the arrival queue length, (ii) the departure queue length, (iii) the runway configuration previously in use, (iv) weather conditions and (v) wind conditions. The first two state components, i.e., the arrival and departure queue lengths, depend endogenously and stochastically on the control that is exercised, i.e., on the selected arrival and departure service rates.

We develop exact and approximate Dynamic Programming solution algorithms that, in combination, enable the on-line implementation of the model. The solution architecture relies on a two-step optimization approach that uses *a priori* as well as real-time information. In the first step, the exact Dynamic Programming algorithm provides operating policies for the entire day of operations, based on inputs available before the beginning of that day. However, this policy might no longer be optimal in the face of such operating disturbances as schedule updates and updated weather forecasts. We therefore implement, as a second step, a fast approximate algorithm based on a one-step look-ahead that greatly accelerates the execution of the model. Computational experiments show that this two-step approximation scheme yields congestion costs within 2% of the optimal solution and thus results in close-to-optimal policies.

We then apply this approach to JFK Airport. We show that the optimal policy is path dependent, i.e., depends on prior decisions and on the prior stochastic evolution of the system over the day. This underscores the theoretical and practical need for integrating operating stochasticity into the decision-making framework. We also show that the model can result in significant congestion cost savings at busy airports, as compared to current operating procedures and existing approaches in the literature. First, we compare the model to two advanced heuristics that aim to imitate typical decisions made in practice. We estimate the resulting congestion cost savings at 20% to 30%. Second, we compare the model to an alternative model based on deterministic queue dynamics and we estimate the benefit of integrating queue stochasticity into the decision-making framework at 5% to 20%.

1.3.2 An Integrated Model of Airport Congestion

Then, we develop an original approach to airport congestion modeling that integrates our tactical Dynamic Programming model of airport capacity utilization (Section 1.3.1) into a strategic Stochastic Queuing Model of airport congestion. This intends to address the gap in the literature between models of capacity utilization that optimize the variations in arrival and departure service rates at the tactical level to best operate scheduled flights, and models of airport congestion that are typically based on single-value estimates of airport capacity.

We integrate an endogenous control of arrival and departure service rates into a stochastic and dynamic queuing model of airport congestion. The resulting integrated model of airport congestion quantifies the relationships between flight schedules, airport capacity and flight delays at the strategic level, given the way service rates can be selected over the day of operations by air traffic managers to maximize airport efficiency at the tactical level, under capacity constraints. It can then be applied to quantify airport congestion under different airport capacity or flight scheduling scenarios.

We formulate an approximate version of the stochastic control of runway configurations and of arrival and departure service rates (Section 1.3.1). This control captures the trade-off between arrival and departure capacity, and approximates the selection of runway configurations. The purpose of this simplification is to approximate the arrival and departure queue lengths that minimize congestion costs at the tactical level, without accounting for all the operational details that need to be considered in the full control. This ensures that the resulting integrated model of airport congestion is computationally efficient.

We validate our model as a means of quantifying airport on-time performance through extensive comparisons with operational data at the three airports in the New York Metroplex (JFK, EWR and LGA). To the best of our knowledge, this represents the first attempt to compare the results of a macroscopic model of airport congestion to historical records of operations over a large sample of days. We develop estimates of on-time performance from flight-level data. We show that our model approximates expected departure queue lengths and expected delays at the three airports relatively well. We also show that it provides good

approximations of the range, hence of the variability, of departure queue lengths.

Last, we apply the model to study recent trends in scheduling and on-time performance at the New York airports. Results suggests that the large delay reductions observed between 2007 and 2011 can be largely attributed to the comparatively small reduction in demand over the period. This underscores the nonlinear relationships between flight schedules and delays, and motivates the consideration of scheduling interventions for congestion mitigation.

1.3.3 An Integrated Approach to Scheduling Interventions

Third, we develop and apply an original approach to airport congestion mitigation that jointly optimizes airport scheduling interventions at the strategic level and airport capacity utilization procedures at the tactical level. This bridges the gap between existing models of flight scheduling and demand management that typically do not consider the impact of flight schedules on airport operations and models of airport and airspace operations that treat flight schedules as given.

We develop the Integrated Capacity Utilization and Scheduling Model (ICUSM). The model optimizes scheduling interventions, given the endogeneity of airport congestion (i.e., the impact of flight schedules on flight delays), and the endogeneity of airport capacity utilization (i.e., how operating procedures can be modified to minimize congestion costs at the tactical level for any given schedule of flights). It relies on an integrated modeling architecture that combines (i) a strategic Integer Programming model of scheduling interventions; (ii) a tactical Dynamic Programming model of capacity utilization; and (iii) a Stochastic Queuing Model of airport congestion.

We introduce an iterative solution algorithm that solves simultaneously the schedule optimization problem and the capacity utilization problem. The algorithm relies on the integration of a *deterministic* queuing model into the Integer Program of scheduling. This provides a modified schedule of flights, under which we approximate the optimal capacity utilization policies. In the last stage, we quantify the resulting expected delays using a *stochastic* queuing model. By iterating among these three steps, we approximate the optimal

schedule of flights and optimal capacity utilization policies under *stochastic* queue dynamics. Extensive computational experiments show that the algorithm converges in a small number of iterations and in short and reasonable computational times.

We present an extensive application of the ICUSM to JFK Airport. Computational results suggest that substantial delay reductions can be achieved through modest modifications to airline schedules that do not disrupt aircraft or passenger itineraries. In particular, peak arrival and departure delays can be reduced by over 30% and 50%, respectively, while (i) no flight originally scheduled is eliminated; (ii) aircraft and passenger connections are maintained; (iii) the scheduled time of 75%-90% of flights landing or taking off at JFK is not modified; and (iv) no flight is displaced by more than 30 minutes. This result capitalizes on the non-linear relationships between flight schedules and airport on-time performance.

Finally, we compare the performance of the integrated approach developed in this thesis with that of a sequential approach under which flight schedules and operating procedures are optimized consecutively. The comparison shows that the benefits of integration can be significant.

1.3.4 Equity and Collaboration in Scheduling Interventions

Finally, we extend the ICUSM (Section 1.3.3) to design mechanisms for scheduling interventions at busy US airports that, for the first time, account for inter-airline equity and enable airline collaboration. These mechanisms (i) rely on non-monetary transfers exclusively; (ii) enable the airlines to provide their preferences regarding which flights to reschedule; and (iii) balance scheduling adjustments equitably among the airlines.

We propose a set of performance attributes for scheduling interventions. We identify efficiency (defined as the mechanism's ability to meet airline scheduling preferences), inter-airline equity (defined as the mechanism's ability to balance scheduling adjustments fairly among the airlines), and on-time performance (defined as the mechanism's ability to mitigate airport congestion) as three performance attributes. We develop quantitative indicators for each of them, using a unified framework of scheduling interventions. We then extend the

ICUSM to formulate a tractable lexicographic architecture to characterize the trade space between efficiency, equity, and on-time performance in airport scheduling interventions.

We provide a theoretical discussion of the trade-off between efficiency and equity. First, we show that, under some conditions, efficiency and equity can be jointly maximized. We then describe how a trade-off between efficiency and equity may arise from inter-airline variations in flight schedules, network connectivities and flight valuations.

We propose non-monetary, credit-based mechanisms for airport scheduling interventions. These mechanisms enable the airlines to provide to a central decision-maker their preferred schedule of flights, the network connections that need to be maintained, and the relative scheduling flexibility of their flights through credit allocation. These inputs are then used to generate a modified schedule, based on efficiency, equity, and on-time performance objectives.

Finally, we generate computational scenarios and results at JFK Airport to analyze the performance of the proposed mechanisms. We first show that, under current scheduling conditions, inter-airline equity can be achieved in airport scheduling interventions at no (or minimal) efficiency losses. We then show that airlines' network connectivities can be preserved and that accounting for airlines' preferences regarding the relative scheduling flexibility of their flights can significantly improve the outcome of scheduling interventions.

1.4 Organization of the Thesis

The remainder of this thesis consists of 6 chapters. In Chapter 2, we present the modeling architecture developed in this thesis and the experimental setup at the airports in the New York Metroplex. Chapters 3 and 4 focus on the problem of capacity utilization. We formulate, solve, and implement our Dynamic Programming model of capacity utilization in Chapter 3, and we develop an approximation thereof that we integrate into a Stochastic Queuing Model of airport congestion in Chapter 4. The focus of this thesis then shifts to the problem of scheduling interventions in Chapters 5 and 6. Chapter 5 develops the ICUSM that integrates the Dynamic Programming model of capacity utilization (Chapter 3) and the Stochastic Queuing Model of airport congestion (Chapter 4) into an Integer Programming

model of scheduling interventions. Chapter 6 then uses this integrated approach to design, optimize, and evaluate mechanisms for scheduling interventions that incorporate inter-airline equity considerations and enable airline collaboration. Finally, Chapter 7 summarizes the findings of this thesis, discusses their implications for policy and practice, and identifies opportunities for future research.

Chapter 2

Modeling Framework

The main objective of this research is to develop operating and scheduling procedures for airport congestion mitigation. Our modeling approach takes as inputs: (i) estimates of airport capacity in various operating conditions at the airport, estimated statistically using historical data, and (ii) flight scheduling data, provided by the airlines. The models developed in this thesis aim to optimize: (i) capacity utilization policies (i.e., the dynamic control of runway configurations and of arrival and departure service rates), and (ii) scheduling interventions (i.e., the rescheduling of flights to reduce peak-hour scheduling levels, while minimizing interference with airline competitive scheduling). Figure 2-1 shows the modeling inputs (Figure 2-1a) and congestion mitigation interventions (Figure 2-1b), which we describe in more detail in Sections 2.1 and 2.2, respectively. Section 2.3 then introduces the Stochastic Queuing Model of airport congestion that we use to quantify airport on-time performance, as well as a simple Markovian model of weather conditions. Section 2.4 discusses the assumptions and the scope of the modeling framework developed in this thesis. Finally, Section 2.5 presents the experimental setup at the airports in the New York Metroplex.

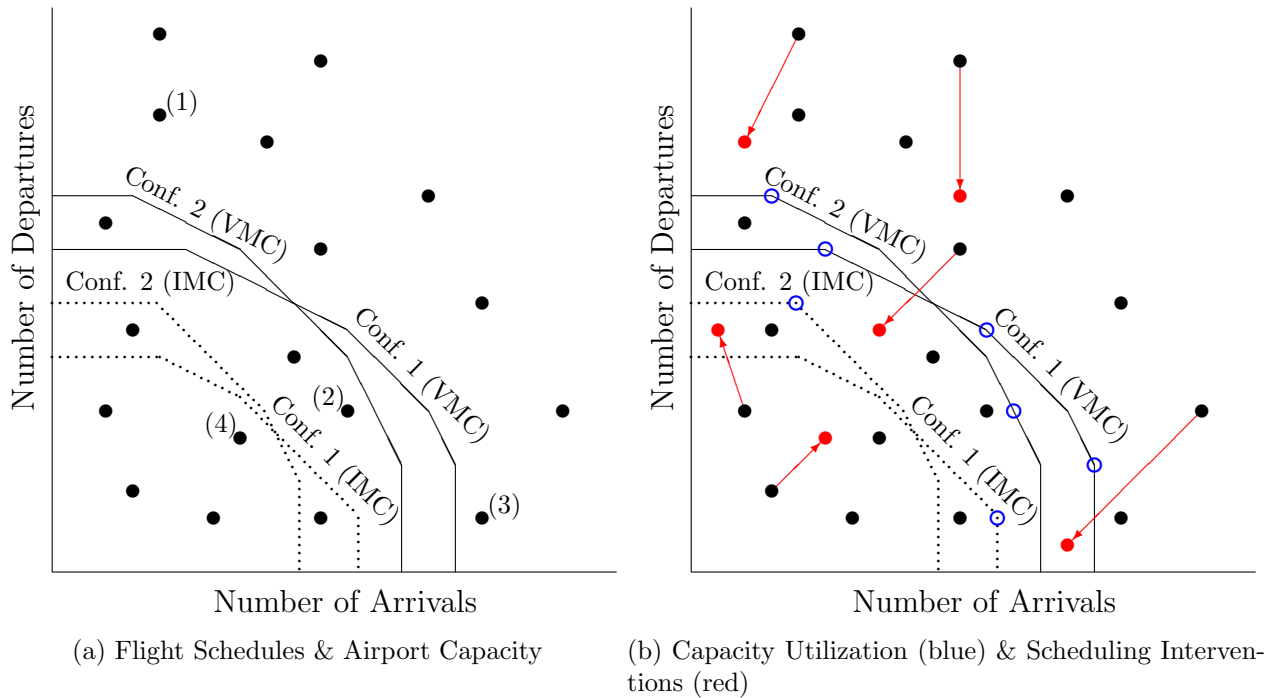


Figure 2-1: A schematic representation of modeling inputs and decisions

2.1 Modeling Inputs

2.1.1 Airport Capacity

Airport capacity is defined as the expected number of movements that can be performed per unit of time, under continuous demand [34]. It depends on a number of operational factors. First, the capacity of US airports is subject to weather variations. Whereas airports outside the United States typically operate under radar-based “Instrument Flight Rules”, US airports usually operate under “Visual Flight Rules”, weather permitting, i.e., the pilots are often requested to maintain visually a safe separation from preceding aircraft during the final approach. This results in lower separations and higher throughput under “Visual Meteorological Conditions” (VMC) than under “Instrument Weather Conditions” (IMC). In the remainder of this thesis, we use VMC and IMC as surrogates of “good” and “poor” weather conditions, respectively. Second, airport capacity depends on the proportion of arrivals and departures operated. There exists at many airports a trade-off between arrival

and departure capacities. This is obvious at single-runway airports, where the longer the runway is occupied by arrivals, the less time is available to accommodate departures, and vice versa. But even when several runways are active, the layout of the runway system may be such that departures and arrivals interact, resulting once again in a trade-off. Third, arrival and departure capacities depend on the runway configuration in use, i.e., the set of simultaneously active runways. Note that airports with two or more runways can have multiple configurations depending on the geometric layout of the runways. For instance, New York’s JFK Airport and Boston’s Logan Airport have 4 and 6 runways, respectively, and can be operated in any one of more than 10 and 20 configurations, respectively.

In order to capture these dependencies, we represent airport capacity by means of piecewise linear *Operational Throughput Envelopes*. These envelopes determine the relationship between the *average* number of landings and the *average* number of takeoffs that can be operated per unit of time for a given runway configuration and in given weather conditions (VMC or IMC). This representation of airport capacity was proposed by Simaiakis [104], who also estimated empirically these envelopes at several major US airports. It extends the concept of “Capacity Envelope” introduced by Gilbo [46], which represents the relationship between the *maximal* number of arrivals and departures that can be operated per unit of time. Other factors impacting airport capacity (e.g., aircraft types, sequencing of arrivals and departures, human factors in operations, environmental considerations, etc.) are not explicitly represented in our characterization of airport capacity, but they are taken into account in the statistical estimation of the Operational Throughput Envelopes. Finally, note that weather conditions are exogenous (we present a model of weather variations in Section 2.3.2), while the runway configuration in use and the balance of serviced arrivals and departures are determined by air traffic handling procedures.

Figure 2-1a shows the VMC and IMC Operational Throughput Envelopes associated with two hypothetical runway configurations. Consistently with the discussion above, three observations are noteworthy. First, a larger average throughput can be achieved in VMC than in IMC. Second, the relationship between the arrival service rate and the departure

service rate for a given runway configuration is non-increasing, which captures the trade-off between arrival and departure throughput. Third, the relationship between arrival and departure capacity at the airport depends on the runway configuration in use. In the example shown in Figure 2-1a, Configuration 1 can achieve the largest arrival throughput when few departures are operated, while Configuration 2 achieves the largest departure throughput. This may occur, for instance, if Configuration 1 uses two active runways to operate arrivals and one to operate departures while Configuration 2 uses two active runways to operate departures and one to operate arrivals.

2.1.2 Flight Schedules

The second set of inputs in the model is the schedule of flights. Its exact definition depends on the scope of the decision-making problem considered. In the case of tactical capacity utilization decisions, the schedule of flights corresponds to that planned for the day of operations considered. In the case of strategic scheduling interventions, a starting point is the schedule of flights that is preferred by the airlines, e.g., a schedule of flights provided by the airlines to a central decision-maker in charge of applying the scheduling interventions (e.g., the FAA in the United States, schedule facilitators or coordinators outside the United States). Note, finally, that these scheduling inputs may also include connections between pairs of flights to enable aircraft, passenger, or crews to connect.

These scheduling inputs are shown in Figure 2-1a as a hypothetical scatter plot of the numbers of landings and takeoffs scheduled at a given airport per time period (e.g., per 15 minutes, hour). For example, Dots (1), (2), (3), and (4) may indicate, respectively, the demand during the periods 07:00 - 07:15, 07:15 - 7:30, 07:30 - 07:45, and 07:45 - 08:00. The remaining dots show scheduled demand in subsequent periods. Using this representation, all the dots in the upper-right corner corresponds to imbalances between demand and capacity, i.e., periods of the day with more flights scheduled than available capacity at the airport. This is what our models of congestion mitigation will aim to reduce.

2.2 Congestion Mitigation Interventions

Figure 2-1b shows the two sets of decisions considered in this thesis. First, capacity utilization decisions aim to optimize where, on the available Operational Throughput Envelopes, the airport should operate at any period of the day. This involves determining which runway configuration to use and how to balance the number of arrivals and departures to operate in order to minimize the imbalances between demand and capacity. This is represented by the blue circles in Figure 2-1b. However, even the “optimal” utilization of airport capacity may not be sufficient to keep delays within desirable bounds if the number of flights scheduled at the airport exceeds its capacity by any significant margin. This motivates the consideration of scheduling interventions to mitigate airport congestion. Scheduling interventions involve modifying the schedule of flights to reduce peak-hour scheduling levels, i.e. to bring dots in the upper-right corner closer to capacity. This is represented by the red arrows in Figure 2-1b. Note that this can be offset by an increase in off-peak scheduling levels.

As previously mentioned, the scheduling and capacity utilization decisions have different time scales and are based on different types of information. Capacity utilization decisions are made dynamically over the course of each day of operations, as a function of observed queue lengths and of observed operating conditions (e.g., weather, winds). In contrast, the scheduling interventions are made months in advance, well before operating uncertainty is resolved. They involve determining the schedule of flights for an entire day of operations, with coupling constraints between different time periods. Thus, optimizing capacity utilization is a sequential decision-making problem that we formulate as a Dynamic Program, while optimizing scheduling interventions is a combinatorial decision-making problem that we formulate as an Integer Program.

Figure 2-1b illustrates the interdependencies between scheduling and operations, thus the need for the *joint* optimization of scheduling interventions and capacity utilization policies. Specifically, any change in flight schedules may lead to changes in how the airport should be tactically operated. Consider a period of the day during which a large number of takeoffs has been scheduled and let us assume that, in order to reduce departure delays,

the departure peak is lowered through the scheduling interventions. In that case, the optimal utilization of airport capacity may also change. Given the trade-off between arrival and departure throughput, characterized by the Operational Throughput Envelope, it may indeed be beneficial to lower the departure service rate and increase the arrival service rate to best serve the changed balance between arrival and departure demand. In other words, the balance of arrivals and departures *scheduled* impacts the optimal balance of arrivals and departures *operated*.

2.3 Descriptive Model of Airport Operations

The design of airport congestion mitigation procedures requires a *descriptive* model of airport operations. We use in this thesis a stochastic and dynamic queuing model of airport congestion, described in Section 2.3.1. Section 2.3.2 describes a Markovian model of weather dynamics at airports.

2.3.1 Queuing Model of Airport Congestion

We characterize in this research the airport as a queuing system, illustrated in Figure 2-2. Aircraft join the queue when they are ready to land or to depart, and service is provided by the runway system. Departing aircraft are queuing on the ground, primarily on the taxiways. Arriving aircraft are queuing in the terminal airspace, in the en-route airspace, or at the origin airport if a Ground Delay Program is implemented.

We represent the arrival queue and the departure queue as two distinct $M(t)/E_k(t)/1$ queuing systems. The demand processes and the service processes are respectively modeled as Poisson processes and Erlang processes of order k . The smaller the value of k , the more variable the service processes. Note that this model is stochastic and dynamic: both demand and service are time-varying random processes. In combination they are intended to capture the uncertainty and the variability associated with the actual queuing processes. For instance, the number of aircraft that demand the usage of the runway system during a

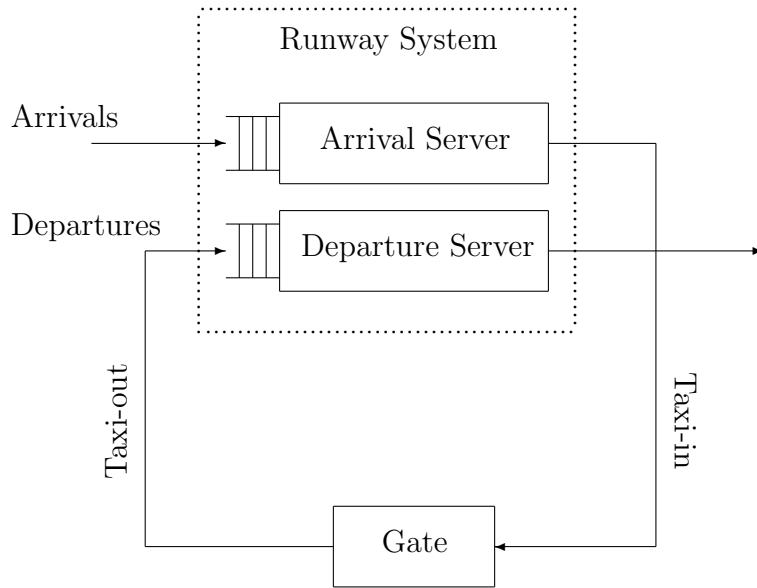


Figure 2-2: Representation of the airport as a queuing system

particular period depends on several uncertain factors, including airline operations, passenger operations in terminal buildings, on-time performance at other airports, etc. As well, the number of landings and takeoffs that can be operated over a period depends on the aircraft mix, the sequencing of landings and takeoffs, human factors in airport operations, etc.

We divide each day of operations into 15-minute periods. The demand and service rates are modeled as constant over any 15-minute period t and are thus denoted by λ_t and μ_t , respectively. The arrival and departure demand rates, i.e., the expected number of arriving and departing aircraft that will demand the usage of the runway system per period, are determined by the schedule of arrivals and departures at the considered airport. The arrival and departure service rates, i.e., the expected number of arrivals and departures that can be operated per period at the considered airport, are constrained by airport capacity. The greater the imbalances between λ_t and μ_t , the greater the resulting queue lengths. Our model of capacity utilization will control the service rates μ_t , while our model our scheduling interventions will impact λ_t .

Note that arrival and departure queues are not independent from each other since the arrival and departure service rates are subject to the same weather-related constraints and are

negatively correlated: increasing the arrival throughput reduces the departure throughput, and conversely. However, we assume that the stochastic evolution of the arrival queue is independent from that of the departure queue, e.g., for given values of the arrival and departure service rates, the arrival queue might be longer than expected, while the departure queue might be shorter than expected.

The state-transition diagram of the $M(t)/E_k(t)/1$ queuing system is shown in Figure 2-3. It relies on the characterization of an Erlang process of order k and rate μ as the succession of k independent and Markovian “stages of work” that are completed at rate $k\mu$. The state of the system is defined by the number of remaining stages of work, denoted by i . We introduce a time index s that varies continuously over the day and we denote by $P_i^S(s)$ the probability of being in state i at time s . Equation (2.1) shows the system of Chapman-Kolmogorov first-order differential equations describing the evolution of the state probabilities $P_i^S(s)$ over time period t , i.e., between $(t-1)S$ and tS (where $S = 15$ min denotes the length of each period). The practical queue capacity is denoted by N (i.e., each of the arrival and departure queues can hold up to N aircraft in the model). The system is assumed to be empty at the beginning of the day.

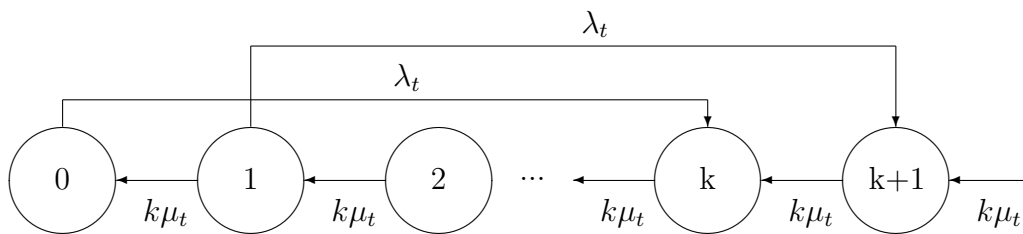


Figure 2-3: State-transition diagram of the $M(t)/E_k(t)/1$ queuing system

$$\begin{aligned}
\frac{dP_0^S(s)}{ds} &= -\lambda_t P_0^S(s) + k\mu_t P_1^S(s) \\
\frac{dP_i^S(s)}{ds} &= -(\lambda_t + k\mu_t)P_i^S(s) + k\mu_t P_{i+1}^S(s) & \forall i \in \{1, \dots, k\} \\
\frac{dP_i^S(s)}{ds} &= \lambda_t P_{i-k}^S(s) - (\lambda_t + k\mu_t)P_i^S(s) + k\mu_t P_{i+1}^S(s) & \forall i \in \{k+1, \dots, (N-1)k\} \\
\frac{dP_i^S(s)}{ds} &= \lambda_t P_{i-k}^S(s) - k\mu_t P_i^S(s) + k\mu_t P_{i+1}^S(s) & \forall i \in \{(N-1)k+1, \dots, kN-1\} \\
\frac{dP_{kN}^S(s)}{ds} &= \lambda_t P_{k(N-1)}^S(s) - k\mu_t P_{kN}^S(s)
\end{aligned} \tag{2.1}$$

We discuss the determination of the practical queue capacity N in Chapters 3 and 4. The value of N should be chosen sufficiently large to approximate a queue with infinite queue capacity, i.e., to avoid underestimating delays. At the same time, large values of N also mean larger computational times to run the model.

2.3.2 A Model of Weather Variations

Weather variations are modeled as a time-varying two-state Markov chain. We denote by p_t (resp. q_t) the transition probability from state *VMC* to state *IMC* (resp. from *IMC* to *VMC*) over period t , i.e., the probability that period $t+1$ is in *IMC* (resp. *VMC*) given that period t is in *VMC* (resp. *IMC*). The time-dependence of p_t and q_t is intended to capture potential variations in weather dynamics over the course of the day and to incorporate weather forecasts. The transition matrix of the weather Markov chain, denoted by P_t^{wc} , is given by:

$$P_t^{wc} = \begin{array}{cc} & \begin{array}{cc} VMC & IMC \end{array} \\ \begin{array}{c} VMC \\ IMC \end{array} & \begin{pmatrix} 1-p_t & p_t \\ q_t & 1-q_t \end{pmatrix} \end{array}$$

The transition probabilities p_t and q_t can be estimated in two different ways, which correspond to two different levels of analysis. First, at the tactical level, we can use weather forecasts to update the probabilities p_t and q_t using latest information (see Chapter 3). Sec-

ond, at the strategic level (i.e., before weather forecasts are made available), we calibrate this model using historical records of weather conditions in order to obtain a statistical characterization of weather variations at the airports. This calibration is detailed in Section 2.5 and applied in Chapters 4 to 6.

2.4 Discussion of the Assumptions of the Models

We discuss next the major assumptions of our model of airport operations, and the scope of the decision-making models for congestion mitigation developed in this thesis.

Model of Airport Operations

Modeling Assumption 1 *Delay dynamics at a given airport are modeled by means of an $M(t)/E_k(t)/1$ queuing model.* This model has been proposed early to model airport delays [67, 57] and applied widely since then. It offers the advantage of being easy to characterize as a Markov process, thus analytically tractable. It relies, however, on specific assumptions on the demand and service processes. On the demand side, the use of the Poisson process has been challenged because of its memoryless property. Alternative characterizations based on pre-scheduled flight times have been proposed [84]. On the service side, the analysis of surveillance data suggests that the Erlang distribution approximates well inter-service times, but alternative characterizations (e.g., a displaced exponential distribution) may provide a better fit [105]. Altogether, the $M(t)/E_k(t)/1$ queuing model has been shown to estimate well the magnitude and dynamics of airport queues observed in practice [76, 97, 59]. Given that our objective is *not* to predict on-time performance with perfect accuracy, but to integrate a model of airport operations into decision-making models for congestion mitigation, this level of adequacy is reasonable. We also perform comparisons between the model's outcome and actual delays observed at the New York airports in Chapter 4.

Modeling Assumption 2 *Weather variations are modeled by the simple, aggregate two-state Markovian model described in Section 2.3.2.* This model relies on current practice

that defines distinct flight rules under VMC and IMC. Additional weather states could be considered to enable more fine-grain characterization of the weather dynamics at the airport [74]. However, this would increase the complexity of the decision-making problems. It would also make the corresponding estimates of the Operational Throughput Envelopes (see Figure 2-1a) less statistically reliable, as each would be based on a smaller data sample. Finally, this model was shown to capture the weather dynamics at the New York airports relatively well [59]. Given the purpose of this simple weather model being *not* to describe exactly weather conditions at an airport on any given day but rather to approximate weather profiles in our decision-making models, this level of adequacy is acceptable.

Modeling Assumption 3 *Delays are considered at only one airport; the propagation of delays in a network of airports is ignored.* Operational disturbances, however, do propagate throughout the network of air traffic operations. First, delays propagate between consecutive flights flown by the same aircraft. Second, flight delays lead to schedule updates, and thus to changes in later delay dynamics at the airports. In recent research, *descriptive* models of airport congestion have been built to quantify air traffic delays in a network of airports [94, 95]. However, the computational requirements of *decision-making* models requires a more restricted scope for the problem. Note, nonetheless, that this assumption is conservative, as any congestion mitigation intervention at a given airport is likely to have ripple effects that will *also* reduce the propagation of delays in a network of airports [96]. Thus, if anything, this assumption underestimates the benefits of the congestion mitigation interventions developed in this thesis.

Models of Congestion Mitigation

Scope Assumption 4 *The congestion mitigation interventions that are considered focus on one airport's runway system.* Several observations are noteworthy. First, the capacity utilization interventions (Chapter 3) can be applied at several airports' runway systems simultaneously. Second, the scheduling interventions that we consider in this thesis (Chapters 5 and 6) do account for the network structure of air traffic operations, but rely on the

assumption that congestion-mitigating scheduling interventions are applied at only one airport. Our framework could be extended to consider such interventions at several airports simultaneously, but this would come at the cost of increased computational requirements. Third, we do not consider other congestion-mitigating interventions applied on the surface of an airport (e.g., gate pushback metering) or in a network of airports (e.g., ground holding, re-routing, speed control). This is a strong assumption as these interventions are undoubtedly *not* independent on runway operations. It is motivated by the fact that runway operations at busy airports are the main bottleneck of system-wide operations, thus runway congestion represents the main source of system-wide delays.

Scope Assumption 5 *Our model of scheduling interventions considers airlines’ schedules of flights, but not other decision variables in the airline planning process.* Our models are based on airlines’ preferred schedules of flights, but airlines’ profit-maximizing decisions also include the choice of aircraft sizes, pricing and revenue management strategies, etc. Also, economic theory has indicated the potential internalization of delays by dominant airlines at concentrated airports [24], which we also leave out of scope of this research. These choices are reasonable because this research does not look for any welfare-maximizing equilibrium. Instead, it starts with a schedule of flights proposed by the airlines for a given day of operations, and looks for “mild” delay-reducing scheduling adjustments. We assume that these small adjustments do not impact other decisions made by the airlines. In fact, our objective of minimizing such scheduling adjustments and the scheduling and network connectivity constraints that we will consider also aim to minimize the impacts of scheduling interventions on other decisions made by the airlines.

Scope Assumption 6 *We do not consider passenger demand and passenger welfare explicitly in our model of scheduling interventions.* Improvements in airport on-time performance certainly increases passenger welfare. But changes in the available schedule of flights also impact passengers’ *schedule delay* (defined as the difference between passengers’ ideal and scheduled travel times [36]), and hence passenger welfare. In this research, we do not con-

sider passengers' preferred travel times explicitly. Instead, we use the airlines' preferred schedule of flights as a proxy for passenger demand. Given the mildness of the scheduling adjustments that are considered, we assume that their effect on passengers' schedule delay is relatively minor, and definitely outweighed by the corresponding reductions in flight delays. This assumption is motivated by the results from Swaroop et al. [114]. Finally, as we do not consider the impacts of the mild scheduling interventions considered in this thesis on airlines' profit-maximizing decisions, we also assume that their effects on other factors that impact passenger welfare (e.g., airfares) are of second order.

Scope Assumption 7 *We do not consider the potential for airlines' strategic behaviors.*

In reality, any mechanism that involves the gathering of airline scheduling preferences may involve some strategic behavior. For instance, it may provide incentives for the airlines to request the scheduling of more flights at peak hours than they would schedule in the absence of any scheduling intervention, in order to gain a competitive scheduling advantage or to deter entry [79]. Vaze and Barnhart accounted for the potential of gaming in airline frequency competition [115, 116], but the time unit of these studies was a full day of operations, and thus too coarse to account for the timetabling of flights and for intra-day delay dynamics. The difficulty with integrating airlines' intra-day strategic behaviors lies in the complexity of capturing the impact of marginal scheduling changes and of airport congestion mitigation on airlines' operating profitability and on passengers' behaviors. In this thesis, we strive to minimize the opportunities for airline strategic behaviors through the mechanisms' objectives of minimizing interference with airline competitive scheduling and through the design of some aspects of the mechanisms, but the explicit consideration of the game-theoretic aspects of the submission of airline scheduling preferences is left for future research.

2.5 Experimental Setup

This thesis analyzes the problem of congestion mitigation at the three primary airports in the New York Metroplex: John F. Kennedy International Airport (JFK), Newark Liberty

International Airport (EWR) and LaGuardia Airport (LGA). Appendix 1 provides the diagram of each of these 3 airports. Their runway systems comprise 4 runways (JFK), 3 runways (EWR), and 2 runways (LGA), respectively. The New York airport system is a prime example of one facing high and growing demand and capacity limitations due to a dense urban area and a dense airspace. In turn, JFK, EWR and LGA are the most congested airports nationwide. Moreover, delays originating from New York's airports propagate throughout the National Airspace System, which reinforces the need for addressing the problem of airport congestion in the New York area. Most of the results shown in this thesis are obtained at JFK Airport. This choice of JFK as the study airport is motivated by the facts that (i) the range of the arrival and departure capacities of its alternative runway configurations creates the scope for seeking potential delay reductions through improvements in capacity utilization procedures; and (ii) its peaked schedule of flights offers opportunities for potential delay reductions through scheduling interventions.

Data on flight schedules, airport operations and flight delays are obtained from the Aviation System Performance Metrics (ASPM) database [41]. Details on the use of this database for our purposes are provided in [59] and in Appendix B. In order to capture the variability of flight schedules over the year, we apply our models of congestion mitigation on the 9 days that correspond to the 9 deciles of the distribution of the number of daily flights at JFK in 2007. Figure 2-4 shows the schedule at JFK, EWR and LGA on busy days in 2007 (specifically, on the days that correspond to the 9th deciles of the distributions of the number of daily flights at these airports in 2007). Note, first, that these three airports face high demand. The number of flights scheduled at peak hours is over 30 per 15-minute period at JFK and EWR and around 25 at LGA, while the VMC capacity of each of these airports is of the order of 20 movements per 15 minutes. Second, the schedules at JFK and EWR are highly uneven, with a morning peak and an extended afternoon peak, as well as valleys around 12 noon and in the evening. In contrast, the schedule at LGA is more evenly distributed over the day, as a result of the enforcement of hourly flight caps at LGA in 2007, but still exhibits significant variability on a 15-minute basis. Third, the proportion of arrivals and departures

can vary over the course of the day. This is most significant at JFK, where more departures than arrivals are scheduled in the morning, while the reverse is true in early afternoon.

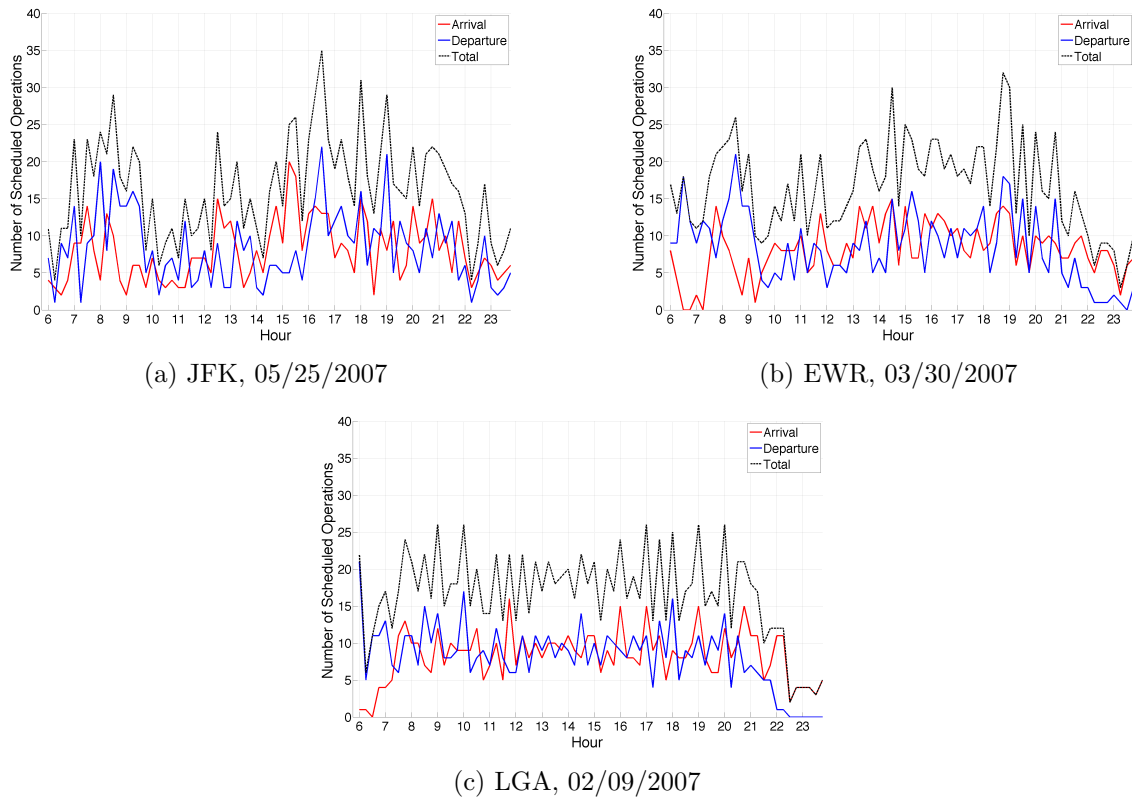


Figure 2-4: Arrival and departure schedules at JFK, EWR and LGA on busy days in 2007

We describe next the major runway configurations at the airports. In this thesis, we use the standard nomenclature that denotes each runway configuration by the set of active runways on which landings are operated, followed by the set of active runways on which takeoffs are operated. We consider the following configurations at JFK:

- (i) Configurations with two arrival runways and one departure runway: $13L, 22L|13R$ and $31L, 31R|31L$. They can achieve the largest arrival rates.
- (ii) Configurations with one arrival runway and two departure runways: $22L|22R, 31L$ and $4R|4L, 31L$. They achieve the largest departure rates when few landings are operated.
- (iii) Configurations with two independent parallel runways: $13L|13R$ and $31R|31L$. They achieve a larger departure rate than configurations with two arrival runways and one

departure runway when few landings are operated, since departures are less constrained by aircraft landing and taxiing in, but a lower throughput when a large number of landings are operated.

- (iv) Configurations with two more closely spaced runways: $22L|22R$ and $4R|4L$. They achieve the lowest service rates. This is mainly due to the location of the runways relative to the terminal complex, since aircraft taxiing in have to cross the departure runway.

Similarly, we consider the following configurations at EWR: (i) Configurations with two arrival runways and one departure runway: $11, 22L|22R, 4R, 11|4L$, and $4R, 29|4L$; (ii) A configuration with one arrival runway and two departure runways: $22L|22R, 29$; and (iii) Configurations with one arrival runway and one departure runway: $22L|22R, 4R|4L$, and $29|22R$. Finally, at LGA, all major configurations use one arrival runway and one departure runway: $31|4, 22|31, 22|13$, and $4|13$.

The VMC and IMC of the Operational Throughput Envelopes of each runway configurations at each airport are obtained from Simaiakis [104] with operational data from the year 2007. Figure 2-5 shows the VMC and IMC envelopes of the runway configurations that achieve the *largest* throughput at JFK (Figure 2-5a), EWR (Figure 2-5b) and LGA (Figure 2-5c). The figure also shows the scatter plot of scheduled arrival and departure counts per 15-minute period in 2007, the size of the dots being proportional to the frequency of the observations. Note, first, that scheduling levels often exceed the capacity of each of the three airports, even in VMC. This is likely to create significant delays, even under optimal operating conditions. Obviously, the imbalances between demand and capacity are even larger in IMC. Note, also, that the relative capacity of the different configurations is similar under VMC and IMC. For instance, the configurations that achieve the largest departure throughput under VMC is likely to also achieve the largest departure throughput under IMC.

We estimate the parameter k of the Erlang random variables that describe the service time by the ratio of its squared sample mean over its sample variance. This ensures that the theoretical and empirical distributions of the service times have the same mean and the

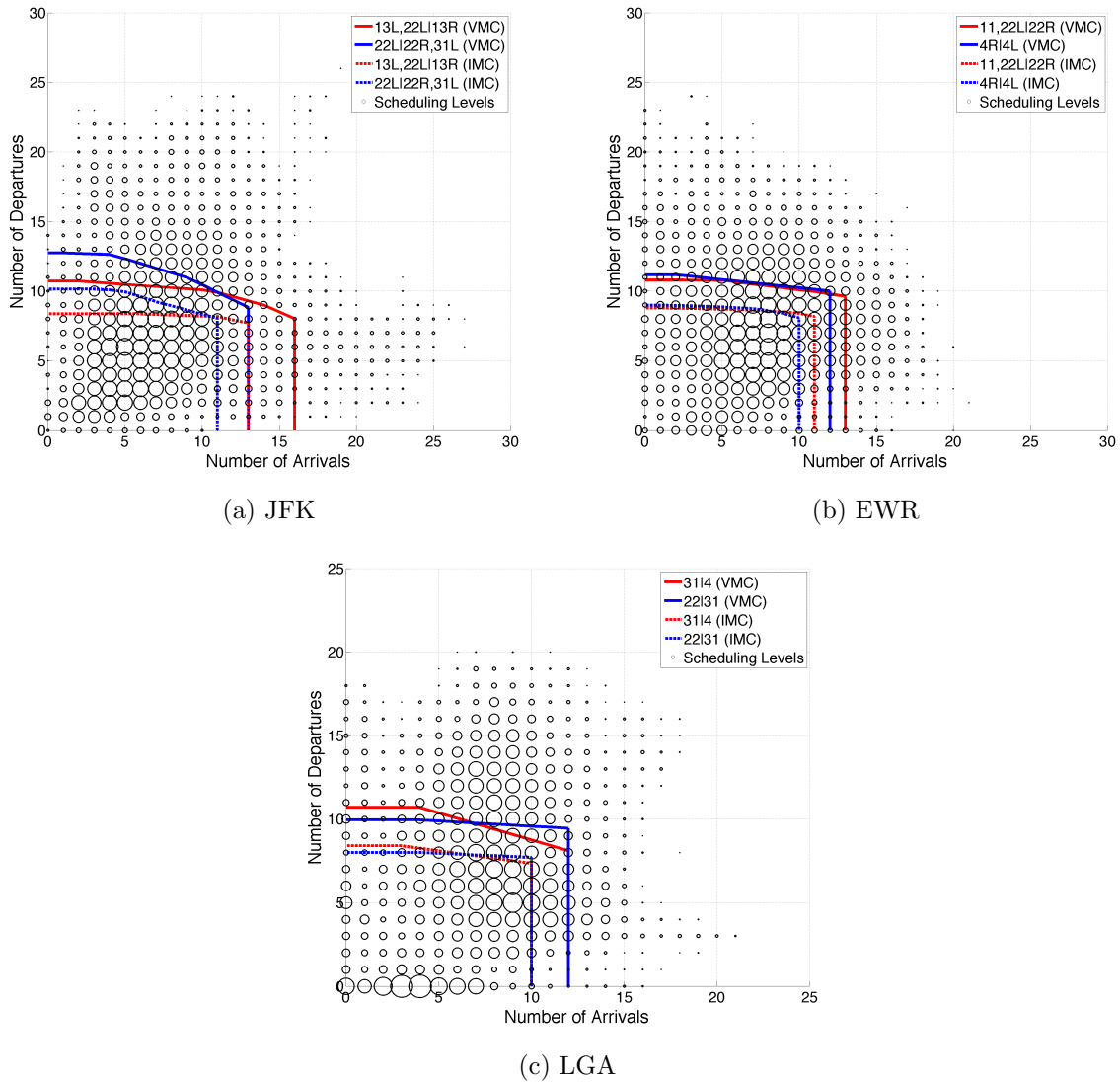


Figure 2-5: Capacity estimates [104] and 2007 scheduling levels at JFK, EWR and LGA

same variance. This leads to the use of $k = 3$ for the cases described here [59].

Finally, the calibration of the model of weather variations using historical weather records is done as follows. We consider two categories of days: *all-VMC days* that have only VMC periods, and *VMC/IMC days* that have some VMC and some IMC periods. The probability that a given day is all-VMC is unbiasedly estimated by the empirical proportion of days that have only VMC periods. During VMC/IMC days, we assume that all values of p_t and q_t are identical (a reasonable assumption at the New York airports), which we denote by

p and q , respectively. We estimate p (resp. q) by its maximum likelihood estimator, i.e., the empirical ratio of the number of transitions from VMC to IMC (resp. from IMC to VMC) over the number of periods in VMC (resp. in IMC). The initial state of the system is sampled from the stationary distribution of the Markov chain, i.e., a VMC/IMC day starts in state VMC (resp. IMC) with probability $\frac{p}{p+q}$ (resp. $\frac{q}{p+q}$). We showed in previous research that this model captures the aggregate weather dynamics at the New York airports reasonably well. Specifically, the model predicts that the number of consecutive periods in VMC (resp. IMC) for VMC/IMC days follows a geometric distribution with parameter p (resp. q). We obtained a very good fit between these model-derived distributions and the empirical distributions, as confirmed by a chi-squared test [59]. Note, again, that this model only applies to cases where airport on-time performance is quantified at the strategic level; as detailed in Chapter 3, our tactical model of capacity utilization can incorporate alternative calibrations of weather variations that are based on real-time forecasts.

The models presented in this thesis are implemented using a combination of three programming languages. First, ASPM data are processed using R 2.15.1. Second, the Dynamic Programming model of capacity utilization and the Stochastic Queuing Model of airport congestion are implemented in MATLAB 8.1. We use the built-in differential equation solver ode45 to solve Equation (2.1). Third, the Mixed Integer Programs developed for scheduling interventions are implemented in GAMS 24.0 using CPLEX 12.5. All computational results are obtained on an Intel(R) Core(TM) i7 running at 2.6 GHz 16 GB RAM. We looked for solutions to the Mixed Integer Programs within an optimality gap of 1%. If none was found after 30 minutes, we accepted the solution obtained at that time.

Chapter 3

Capacity Utilization under Operating Stochasticity

This chapter develops a decision-making model that optimizes the utilization of airport capacity at the tactical level (i.e., over each day of operations) to minimize congestion costs, for any given schedule of flights. It provides the first decision support tool to improve the selection of runway configurations and the balancing of arrival and departure service rates that considers the stochasticity of airport queue dynamics and of operating conditions (i.e., weather and winds) in these decisions. In practice, these are two of the major decisions faced by air traffic controllers and they are primarily made on the basis of experience. The main drivers of these decisions include inertia, wind speed and directions, arrival and departure schedules, noise abatement procedures, configuration switch proximity and inter-airport coordination [98].

Under stochastic airport operations, the arrival and departure queues observed at later time periods are *not* known with certainty in advance. We thus develop a *dynamic* decision-making framework. For instance, let us consider a period of the day when more arrivals than departures are scheduled. Existing approaches, based on deterministic queue dynamics, might suggest the use of a runway configuration that utilizes most of available capacity to operate arrivals rather than departures during that period (e.g., a configuration with two

arrival runways and one departure runway). If, however, the departure queue is longer than expected at the beginning of the period, while the arrival queue is shorter than expected, then it might be beneficial to choose an alternative runway configuration to enhance the departure throughput and alleviate ground congestion. A dynamic approach might thus yield operational benefits in the face of queue uncertainty.

Any decision-making support tool must be applicable in real time. Indeed, airport operations are subject to real-time disturbances that impact operating procedures. This might include schedule updates arising from the occurrence of upstream delays in previous flight legs, en-route congestion or the initiation of Ground Delay Programs. This might also include changes in weather and wind forecasts. For this reason, we develop a solution architecture that enables the dynamic revision of capacity utilization policies. Such revisions are performed almost instantaneously as new real-time information becomes available and are thus well suited to the actual problem faced by air traffic controllers.

The contributions of this chapter fall into four categories:

- We formulate an original Dynamic Programming model of airport capacity utilization under stochastic queue dynamics and stochastic operating conditions.
- We develop exact and approximate solution algorithms that, in combination, enable the real-time implementation of the model.
- We characterize the optimal capacity utilization policies at JFK Airport. We show that the optimal policy is path dependent, i.e., depends on prior decisions and on the prior stochastic evolution of the system over the day. This underscores the theoretical and practical need for integrating operating stochasticity into the decision-making framework.
- We show that the model can result in significant congestion cost savings at busy airports, as compared to current operating procedures and existing decision-making models based on deterministic queue dynamics.

A paper based on this chapter can be found in [63].

The remainder of this chapter is organized as follows. In Section 3.1, we formulate the Dynamic Programming (DP) model of airport capacity utilization. In Section 3.2, we present exact and approximate solution algorithms that, in combination, enable the on-line implementation of the model in real-time. We characterize the optimal capacity utilization policies at JFK Airport in Section 3.3 and quantify the benefits of the decision-making model in Section 3.4. Section 3.5 concludes.

3.1 Model Formulation

We formulate the dynamic control of runway configurations and of arrival and departure service rates as a finite-horizon Dynamic Programming (DP) model. A day of operations, between 6 a.m. and 12 a.m., is divided into $T = 72$ periods of length $S = 15$ minutes. We index these time periods by $t = 1, \dots, T$. Observations and decisions are made at the beginning of each period. Operations between 12 a.m. and 6 a.m. are not considered since they are typically not capacity-constrained and decisions are based on noise abatement procedures and other environmental concerns.

3.1.1 Problem Statement

The model takes as inputs (a) the schedule of landings and takeoffs on a given day and (b) capacity estimates, i.e., the VMC and IMC Operational Throughput Envelopes of each runway configuration (see Chapter 2). During each 15-minute period of the day, we denote by x_t (resp. y_t) the number of aircraft scheduled to land (resp. to take off) at the considered airport. Figure 3-1 shows hypothetical schedules and capacities (similarly to Figure 2-1a).

The problem of jointly selecting a configuration and the arrival and departure service rates, under dynamic and stochastic conditions, can be illustrated as follows. Let us consider a period of the day with a large number of scheduled departures. It might be optimal to operate in Configuration 2 during this period to maximize the departure throughput (e.g., by operating with the service rates corresponding to Point A in VMC or to Point B in IMC).

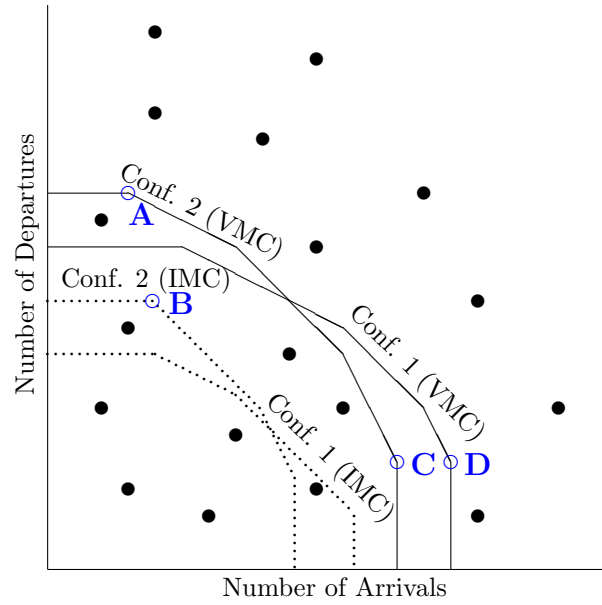


Figure 3-1: Schematic representation of capacity utilization decisions

If, at a later period, a large number of arrivals are expected to be operated, then it might be optimal to enhance the arrival throughput. This can be done by staying in Configuration 2 (e.g., by operating in Point C) or by switching to Configuration 1 to increase the arrival throughput further (e.g., by operating in Point D). The best choice is not obvious for at least two reasons. First, there is a trade-off between the potential efficiency enhancements resulting from configuration switches and the costs of operating inefficiency during the time it takes to switch configurations. Second, the stochasticity of airport operations makes it challenging to anticipate the future evolution of the system.

3.1.2 State Variables

The decisions in each period depend on the observed extent of congestion, on the runway configuration in use and on operating conditions. Operating conditions, including weather and winds, are assumed to be observed at the beginning of each time window and not to change over one 15-minute period. At the beginning of period t , the state is described by the following variables:

- Arrival Queue Length a_{t-1} : Number of arriving aircraft queuing at the end of the

previous period

- Departure Queue Length d_{t-1} : Number of departing aircraft queuing at the end of the previous period
- The runway configuration in use at period $t - 1$, denoted by RC_{t-1}
- Weather conditions for the next period, denoted by $wc_t \in \{VMC, IMC\}$
- The wind state ws_t , which determines the set of runway configurations that *can* be used at any time. According to FAA-specified safety requirements, a runway can only be used if the tailwind and crosswind do not exceed the thresholds of 5 knots and 20 knots, respectively [40]. A wind state is defined as the set of wind vectors, i.e., the set of wind strengths and wind directions, such that the same set of runways satisfies the corresponding threshold requirements. In other words, a wind state is defined by the runways that meet the threshold requirements and by the runways that do not. This approach follows the procedure developed by Li and Clarke [73]. Table 3.1 defines the 13 wind states at JFK. For instance, State 1 corresponds to a situation of calm winds, so that flights can be operated on all runways. In contrast, State 9 corresponds to the case of strong winds from the South, in which case runways $4L$, $4R$, $31L$ and $31R$ face above-threshold tailwinds.

Table 3.1: Definition of wind states at JFK: Runways that can be used per wind state

Wind States	1	2	3	4	5	6	7	8	9	10	11	12	13
$4L, 4R$	✓	✓	✓	✓	✗	✓	✓	✗	✗	✓	✗	✗	✗
$22L, 22R$	✓	✓	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗
$13L, 13R$	✓	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗	✓	✗
$31L, 31R$	✓	✗	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓
Proportion (%)	10.6	7.6	16.0	8.9	15.6	2.5	13.6	8.1	8.8	10.3	0.8	0.5	5.7

3.1.3 Decision Variables

The decision in each period t consists of two components:

- The runway configuration for the next period RC_t . This choice is constrained by the wind state: a runway configuration can only be used if each of its runways meets the threshold requirements. The set of runway configurations that can be selected when the wind state is ws_t is denoted by $\mathcal{RC}(ws_t)$. For each one of JFK's 8 main runway configurations, Table 3.2 indicates the set of wind states for which the configuration can be used. For instance, configuration $13L, 22L|13R$ can be used only in wind states that allow runways $13L$, $13R$ and $22L$ to be used, i.e., in Wind States 1, 2, 5 and 8 (see Table 3.1). The selection of the runway configuration and the observation of the weather state determine the Operational Throughput Envelope for the next period.

Table 3.2: Set of JFK runway configurations that can be selected per wind state

Wind States	1	2	3	4	5	6	7	8	9	10	11	12	13
$13L, 22L 13R$	✓	✓	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗
$31L, 31R 31L$	✓	✗	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓
$22L 22R, 31L$	✓	✗	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗
$4R 4L, 31L$	✓	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗
$13L 13R$	✓	✓	✗	✓	✓	✓	✗	✓	✗	✗	✗	✓	✗
$31R 31L$	✓	✗	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓
$22L 22R$	✓	✓	✓	✗	✓	✗	✗	✓	✓	✗	✓	✗	✗
$4R 4L$	✓	✓	✓	✓	✗	✓	✓	✗	✗	✓	✗	✗	✗

- The rates at which arrivals and departures are served, respectively denoted by μ_t^a and μ_t^d . These service rates are controlled among the outermost set of achievable service rates for the selected runway configuration and the observed weather conditions. For instance, in Configuration 2 in Figure 3-1 and in VMC weather, the decision-maker can decide to operate with the arrival and departure service rates corresponding to point A, to point C, or to any other point on the VMC envelope. The decision-maker in fact selects the arrival service rate for the next period $\mu_t^a \in \{0, \dots, A_{RC_t, ws_t}\}$. The upper bound of this choice A_{RC_t, ws_t} corresponds to the largest arrival throughput that can be realized in the selected runway configuration and observed weather conditions. In turn, the departure service rate μ_t^d is determined by the Operational Throughput

Envelope corresponding to the selected runway configuration and the observed weather conditions. Hence, we can write: $\mu_t^d = \Phi_{RC_t,wc_t}(\mu_t^a)$, where Φ_{RC_t,wc_t} denotes the function that associates to the arrival service rate the value of the departure service rate when the airport operates in runway configuration RC_t and in weather conditions wc_t .

3.1.4 Dynamics of the System

Queue Dynamics

Arrival and departure queue lengths are quantified by means of $M(t)/E_3(t)/1$ queuing models as a function of demand and service rates (see Section 2.3.1). The arrival and departure service rates are determined by the number of landings and takeoffs, respectively, scheduled per 15-minute period. The arrival and departure service rates are controlled by the decision-maker. The state transition probabilities P^S are determined by Equation (2.1). Note that each “state” of the queue corresponds to number of “stages of work” in the system. Each queue length greater than 0 corresponds to k “stages of work”.

Since decision-makers cannot observe the fine-grain state of the queues, i.e., the number of remaining “stages of work”, but observe the queue lengths instead, we proceed by aggregation. In other words, we map the state transition probabilities into queue length transition probabilities (Figure 3-2). This reduces the number of states of the arrival and departure queuing systems from $kN + 1$ to $N + 1$. In turn, this reduces the size of the state space of the DP model by a factor of k^2 and thus improves its tractability considerably.

In order to compute the queue length transition probabilities from the state transition probabilities, we assume that no aircraft is being served at the beginning of any period, i.e., the system is in one of the states $\{mk, 0 \leq m \leq N\}$. The motivation for this choice is that the arrival and departure service rates are typically much greater than 1 per period. We denote the queue length transition probabilities over period t by $Q_{m,n}^t$, i.e., if q_t denotes the queue length at the end of period t , then $Q_{m,n}^t = P(q_t = n | q_{t-1} = m)$. The probabilities $Q_{m,n}^t$ obviously depend on the demand and service rates λ_t and μ_t over period t and they are

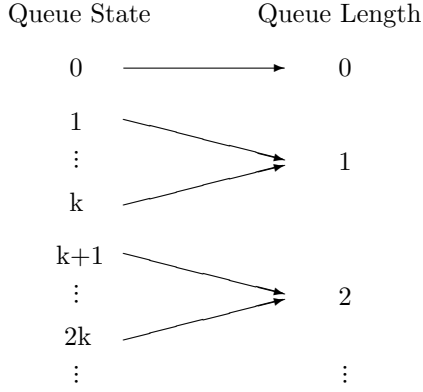


Figure 3-2: Mapping between queue states and queue lengths

derived from the continuous state probabilities P^S according to the following relationship:

$$\begin{aligned}
 Q_{m,0}^t &= P_0^S(tS) \\
 Q_{m,n}^t &= \sum_{l=1}^k P_{(n-1)k+l}^S(tS), \quad \forall n = 1, \dots, N
 \end{aligned} \tag{3.1}$$

where the state probabilities P^S are obtained by solving the system of differential equations (2.1), with the initial conditions:

$$P_i^S((t-1)S) = \begin{cases} 1, & \text{if } i = mk \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

We store in a look-up table the queue length transition probabilities $Q_{m,n}^t$, for all m, n, λ_t, μ_t , so we do not have to re-solve this system of equations at each iteration of the DP algorithm.

Runway Configuration Changes

A challenge in the development of any decision-making support tool to improve the control of runway configurations is to represent the costs of switching runway configurations. Switching configurations is a challenging operating procedure that requires extensive coordination among several airport stakeholders, including airlines and air traffic controllers. Most importantly, queuing aircraft need to be re-routed, which may lead to operational inefficiencies and to the interruption of runway operations for some time. This creates a trade-off between

potential operating enhancements resulting from configuration changes, on the one hand, and the costs of operating the configuration changes, on the other.

Several different ways to represent the costs associated with runway configuration changes have been used in the literature. The most straightforward way is to introduce a penalty cost c that is incurred each time a switch is implemented [73]. A second option is to introduce a time period of idleness during which the configuration switch is operated [18]. Finally, one could model configuration switches through reduced airport capacity (instead of zero capacity) during a longer period [120].

We choose to represent here this cost by a time period of idleness of the runway system, of length denoted by τ_I , during which arriving and departing aircraft may join the queue at an unchanged rate (determined by scheduling levels) but no flight is serviced. In other words, we assume that, if the decision-maker decides to change runway configurations at the beginning of the t^{th} period and chooses arrival and departure service rates $\mu^{(a)}$ and $\mu^{(d)}$ from the Operational Throughput Envelope of the new configuration, then the arrival and departure service rates are both equal to 0 during τ_I . After this, the arrival and departure service rates are, respectively, equal to $\mu^{(a)}$ and $\mu^{(d)}$ until the end of the 15-minute period. This situation is depicted in Figure 3-3.

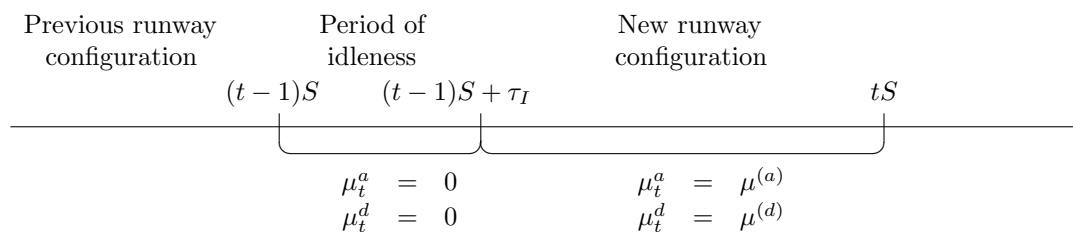


Figure 3-3: Representation of a runway configuration change at the beginning of period t

Modeling runway configuration changes through a time period of idleness offers several advantages. First, it is a simple and transparent way to capture the trade-off between potential operating enhancements and the operating costs of configuration changes. If efficiency can be improved by changing configurations at a given time, this comes at the cost of temporary idleness and consequent initial build-up of the arrival and departure queues. The

attractiveness of runway configuration changes depends on the duration of idleness τ_I . Second, it models the effects of configuration changes as increases in arrival and departure queue lengths and does not require the introduction of an artificial penalty cost c . Third, it enables the quantification of the sensitivity of optimal capacity utilization policies and arrival and departure queue lengths to the costs of configuration changes. Last, it can be efficiently integrated into our DP decision-making model, by modifying the queue length transition probabilities $Q_{m,n}^t$ (Section 3.1.4) to reflect the period of idleness following any configuration switch. Our DP framework, in fact, enables fine-grain calibration of the parameter τ_I , while previous approaches considered a duration of idleness equal to (or proportional to) the length of the decision-making period (e.g., $S = 15$ min) [18].

Finally, this model of configuration changes enables the consideration of additional complexities at no computational cost. For instance, the duration of idleness resulting from a configuration change may vary as a function of the “proximity” of the two consecutive configurations [98]. Indeed, simply activating an additional runway for arrivals or departures (e.g., activating Runway 4R when 4L is already in use) might be less disruptive than a move to a very different configuration that requires a change in the entire flow of arriving and departing aircraft. For this reason, we denote by $\tau_I^{RC_1,RC_2}$ the duration of idleness following the change from runway configuration RC_1 to runway configuration RC_2 .

Note that, in practice, the operational cost of configuration changes might also depend on other system characteristics. For instance, it might increase with the extent of congestion at the time of the change, i.e., with the number of arriving and departing queuing aircraft. As well, it might be larger under IMC than under VMC. These additional complexities can also be captured in our model at no computational cost by making the duration of the idleness τ_I a function of additional state components, e.g., a_{t-1} , d_{t-1} and w_{c_t} . The framework presented here could also be extended to capture alternative characterizations of the operating costs of runway configuration changes, e.g., with a marginally decreasing impact on capacity [120].

Weather and Wind Dynamics

Weather and wind dynamics are modeled by means of Markov chains. As detailed in Chapter 2, weather variations are modeled as a time-varying two-state Markov chain (*VMC* and *IMC*). A similar Markov chain is introduced to characterize the wind dynamics. It is defined by the transition probability from State i to State j , for all pairs of wind states (i, j) (see Table 3.1).

3.1.5 Cost Function

The control strategy aims to minimize congestion costs, which are typically modeled as a non-decreasing function of the queue length with non-decreasing marginal costs. We consider a quadratic cost function of the arrival and departure queue lengths, since the expected total delay scales quadratically with the number of queuing aircraft. Moreover, the costs associated with arrival queues are weighted by a factor $\alpha \geq 1$. This is to capture the fact that arrival delays and departure delays may have different costs, as arriving aircraft can be more challenging and expensive to hold in queue than departing aircraft.

The cost function is written as follows:

$$\alpha \sum_{t=1}^T a_t^2 + \sum_{t=1}^T d_t^2. \quad (3.3)$$

3.1.6 Dynamic Programming Formulation

As described in Section 3.1.3, we denote by $\mathcal{RC}(ws)$ the set of runway configurations that can be selected in wind state ws and by $A_{RC,wc}$ the maximal arrival rate that can be handled in runway configuration RC and in weather conditions $wc \in \{VMC, IMC\}$. We denote the cost-to-go function by $J_t(a_{t-1}, d_{t-1}, RC_{t-1}, wc_t, ws_t)$, which represents the expected total cost of being in state $(a_{t-1}, d_{t-1}, RC_{t-1}, wc_t, ws_t)$ at the beginning of period t . The decision-maker minimizes the sum of the expected congestion costs experienced at the end of period t (i.e., $\alpha E[a_t^2] + E[d_t^2]$) and the future congestion costs from period $t + 1$ onward (i.e.,

$E [J_{t+1} (a_t, d_t, RC_t, wc_{t+1}, ws_{t+1})]$). The Bellman equation is written as follows:

$$J_t(a_{t-1}, d_{t-1}, RC_{t-1}, wc_t, ws_t) = \min_{\substack{RC_t \in \mathcal{RC}(ws_t) \\ \mu_t^a \in [0, A_{RC_t, wc_t}]}} \left(\alpha E [a_t^2] + E [d_t^2] \right. \\ \left. + E [J_{t+1} (a_t, d_t, RC_t, wc_{t+1}, ws_{t+1})] \right), \forall t = 1, \dots, T \quad (3.4)$$

$$J_{T+1}(a_T, d_T, RC_T, wc_{T+1}, ws_{T+1}) = 0 \quad (3.5)$$

The arrival queue a_t at the end of period t depends on the number of scheduled arrivals x_t during period t , on the duration of idleness $\tau_I^{RC_{t-1}, RC_t}$, if any, on the arrival rate μ_t^a during period t and on the previous period's arrival queue length a_{t-1} . Similarly, the departure queue d_t depends on the variables y_t , $\tau_I^{RC_{t-1}, RC_t}$, μ_t^d and d_{t-1} . A summary of the dependencies described in Section 3.1.4 is provided below, where solid lines denote system evolution and dashed lines denote constraints on the decisions.

$$\begin{aligned} ws_t & \text{ ---} \rightarrow RC_t \\ RC_{t-1}, RC_t & \longrightarrow \tau_I^{RC_{t-1}, RC_t} \\ RC_t, wc_t & \text{ ---} \rightarrow \mu_t^a, \mu_t^d = \Phi_{RC_t, wc_t}(\mu_t^a) \\ x_t, \tau_I^{RC_{t-1}, RC_t}, wc_t, \mu_t^a, a_{t-1} & \longrightarrow a_t \\ y_t, \tau_I^{RC_{t-1}, RC_t}, wc_t, \mu_t^d, d_{t-1} & \longrightarrow d_t \end{aligned}$$

3.2 Solution Algorithm

3.2.1 Experimental Setup

We apply the model at JFK Airport (Section 2.5). Unless otherwise specified, we show results obtained with the schedule on 05/25/2007, which corresponds to the 9th decile of the distribution of the number of daily flights at JFK in 2007. Figure 3-4 shows the Operational Throughput Envelopes of JFK's main runway configurations, and the scatter plot of scheduled arrivals and departures per 15-minute period on 05/25/2007 (similarly to Figure 2-5a).

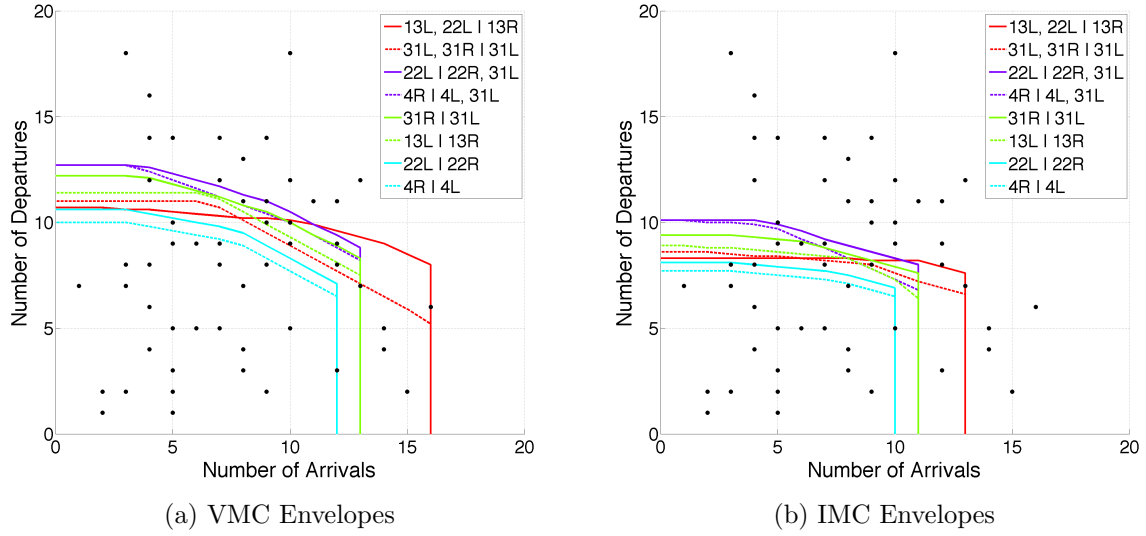


Figure 3-4: Operational Throughput Envelopes [104] and scheduling levels on 05/25/2007

We calibrate our Markovian weather model as follows. We, first, compute the optimal policy assuming that the airport operates in VMC throughout the day, i.e., we set the weather transition probabilities equal to $p_t = 0$ and $q_t = 1$ for all periods t . Note that, by doing so, we do derive the optimal policy if the system is in State IMC at any period t , but this decision is made under the assumption that the airport will operate in VMC from period $t + 1$ to period T . We then update the policies if a weather forecast predicts that the airport will operate under IMC at some periods of the day. This procedure is detailed in Section 3.2.3 and tested in Section 3.2.4.

Wind transition probabilities are estimated by their maximum likelihood estimator using historical records of operations [41]. In other words, we estimate the transition probability from State i to State j by the ratio $\frac{n_{ij}}{n_i}$, where n_{ij} (resp. n_i) designates the number of transitions from State i to State j (resp. the number of periods in State i). This follows the procedure developed by Li and Clarke [73]. Note that the model can easily be extended to incorporate wind forecasts.

Unless otherwise specified, we use the same duration of idleness $\tau_I^{RC_1, RC_2}$ for all pairs of runway configurations $RC_1 \neq RC_2$. To simplify notation, we denote it by τ_I in the remainder

of the chapter. This assumption can be easily modified to introduce differing values of this parameter for different configuration pairs. We provide an example in Section 3.3.1. The purpose of our experimental tests is to capture the trade-off between potential operating enhancements and the operating costs of configuration changes, for different values of τ_I , as described in Section 3.1.4. Sensitivity analyses will quantify the impact of τ_I on optimal policies and expected queue lengths. In practice, the parameters τ_I can be calibrated by airport operators on a case-by-case basis.

We consider a value of $\alpha = 1$, i.e., we assume that arrival and departure delays have identical costs. We investigate the sensitivity of optimal policies and expected queue lengths to α in Section 3.3.3.

Finally, we set the value of the practical queue capacity at $N = 30$. Note that a small value of N may lead to underestimation of the expected arrival and departure queues. On the other hand, a large N will increase computational times, as the number of states scales quadratically with N . We have, in fact, tested a JFK scenario with a larger value of $N = 40$, which implies an almost 75 percent increase in the number of queuing system states from 961 ($=31^2$) to 1681 ($=41^2$) and, consequently, greatly increases computational times. Only marginal differences were observed in the optimal policies, compared to $N = 30$. A capacity of $N = 30$ for (each of) the arrival queue and the departure queue therefore appears to be sufficiently large to minimize the effects of the finite queue size on capacity utilization policies for the JFK application.

3.2.2 Exact Dynamic Programming Algorithm

First, an exact DP algorithm is implemented using the solution concept of backward induction [13, 14]. The optimal policy in the final period (i.e., between 23:45 and 24:00) is computed for all possible states, i.e., for all possible arrival and departure queue lengths that can be observed at 23:45, for all possible runway configurations that can be used in the previous period (i.e., between 23:30 and 23:45) and for all possible weather and wind conditions. This provides optimal costs in the final period as a function of the state of the system

at the beginning of the final period. This cost is then used to compute optimal policies in the second-to-last period, as a function of the state of the system at 23:30. This process is repeated until the optimal policies for all periods have been derived.

The exact DP algorithm is executed in approximately 90 minutes on a laptop computer. The optimal policy for the entire day of operations can thus be easily obtained off-line, i.e., before the beginning of the day. This policy is based on the original model parameters, including the original schedule of flights and the original weather forecast (e.g., assuming that the airport will operate under VMC).

However, the policy determined off-line might no longer be optimal in the face of such dynamic disturbances as schedule updates (e.g., upstream delays, en-route congestion, the initiation of Ground Delay Programs, surface congestion, etc.) or changes in weather forecasts. In these instances, the policy might need to be dynamically revised over the course of the day using real-time information. But given its computational requirements, the exact DP algorithm is likely to exceed the time frame of actual decision-making by air traffic controllers. For this reason, we implement an approximate algorithm, described below, to perform the dynamic revisions. This greatly accelerates execution and consequently enables the policies to be updated in real time.

3.2.3 One-Step Look-Ahead Algorithm

In this section, we develop an approximate solution algorithm based on a one-step look-ahead. The solution architecture is shown in Figure 3-5. In the upper part, we consider the policy obtained off-line with the exact DP algorithm (Section 3.2.2), based on the original model parameters (e.g., the original schedule of flights and the original weather forecast). We denote this policy by $(\widehat{RC}_t, \widehat{\mu}_t^a)$ and its associated expected cost-to-go function by \widehat{J} .

In the lower part, we consider an operational disturbance (e.g., a schedule update, a new weather forecast, etc.) which leads to a change in the model's inputs and parameters (e.g., x_t , y_t , p_t , q_t). Ideally, one would apply the exact DP algorithm with the updated model parameters. This would yield the optimal revised policy, which we denote by $(\overline{RC}_t, \overline{\mu}_t^a)$.

However, the computational times of the exact DP algorithm prevents this policy from being obtainable in real time. We therefore derive a revised policy with a one-step look-ahead algorithm based on the cost-to-go function \hat{J} . In other words, we choose, at the beginning of period t , the policy that minimizes expected total costs, assuming that costs from period $t + 1$ onward are given by \hat{J} . The policy for period t , denoted by $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$, is determined as follows [15]:

$$\left(\widetilde{RC}_t, \widetilde{\mu}_t^a\right) = \arg \min_{\substack{RC_t \in \mathcal{RC}(ws_t) \\ \mu_t^a \in [0, A_{RC_t, wc_t}]}} \left(\alpha E [a_t^2] + E [d_t^2] + E \left[\hat{J}_{t+1} (a_t, d_t, RC_t, wc_{t+1}, ws_{t+1}) \right] \right) \quad (3.6)$$

The execution of this algorithm is almost instantaneous and thus well suited to the actual problem faced by air traffic controllers. In turn, this fast approximation scheme can be implemented in real time when new information becomes available. The performance of this approach depends on how well the look-ahead policy, $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$, approximates the optimal revised policy, $(\overline{RC}_t, \overline{\mu}_t^a)$. We evaluate this performance in the next section.

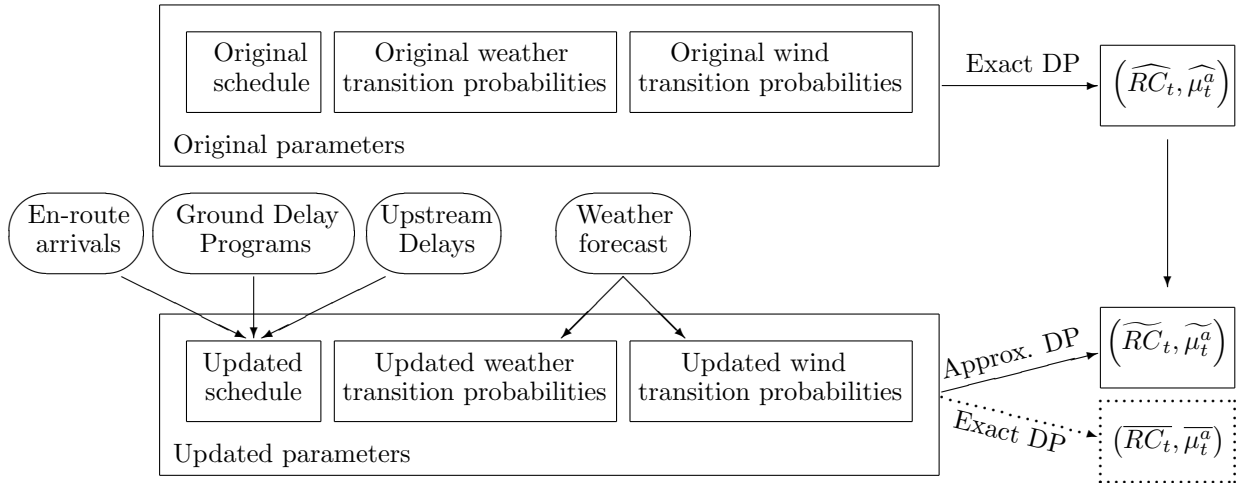


Figure 3-5: Representation of the solution architecture

3.2.4 Evaluation of Performance

In order to test the performance of our approach, we first simulate schedule disturbances and we then simulate weather forecasts. We can simulate wind forecasts similarly. In each instance, we compare the expected congestion costs resulting from the application of (a) the optimal policy with the updated model parameters, $(\overline{RC}_t, \overline{\mu}_t^a)$, which might be too computationally time-consuming to be determined in real time, (b) the optimal policy with the original model parameters, $(\widehat{RC}_t, \widehat{\mu}_t^a)$, and (c) the approximate policy produced by the one-step look-ahead algorithm with the updated model parameters (Equation (3.6)), $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$. We report the average error of policies $(\widehat{RC}_t, \widehat{\mu}_t^a)$ and $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$, defined as the relative increase in congestion costs as compared to policy $(\overline{RC}_t, \overline{\mu}_t^a)$.

The purpose of this comparison is to show that we obtain close-to-optimal policies under various demand and weather scenarios that might arise in practice at the airports. Note that the goal here is *not* to develop an accurate model of schedule disturbances or weather variations. Instead, we simulate realistic examples of such disturbances and we evaluate the performance of the approximation scheme developed in Section 3.2.3.

We simulate a schedule disturbance as follows: at each period, we introduce a schedule perturbation randomly sampled from the integers within ε of the original number of scheduled arrivals and departures. For instance, if 10 arrivals and 15 departures are originally scheduled during a given period, then we uniformly sample, for a value of $\varepsilon = 20\%$, the updated number of arrivals (resp. departures) from the five integers between 8 and 12 (resp. from the seven integers between 12 and 18). The expected total number of flights in the updated schedule is identical to that of the original schedule. For each value of ε considered, we simulate 10 schedules of flights and we report in Table 3.3 the average relative error of each policy.

First, note that the optimal policy computed off-line with the original schedule, $(\widehat{RC}_t, \widehat{\mu}_t^a)$, performs reasonably well, even after schedule updates. Even for the largest schedule perturbations, this policy still results in expected congestion costs within 5% to 10% of the optimal congestion costs. Moreover, the one-step look-ahead algorithm significantly improves the performance over the original policy. Indeed, policy $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$ results in expected congestion

Table 3.3: Total expected congestion costs with the three policies under schedule updates

	Optimal Policy	Available Policies	
	$(\overline{RC}_t, \overline{\mu}_t^a)$	$(\widehat{RC}_t, \widehat{\mu}_t^a)$	$(\widetilde{RC}_t, \widetilde{\mu}_t^a)$
Schedule Considered	Updated	Original	Updated
Algorithm	Exact	Exact	Look-Ahead
Available on-line?	No	Yes	Yes
$\varepsilon = 10\%$	Baseline	+0.44%	+0.12%
$\varepsilon = 20\%$	Baseline	+1.33%	+0.38%
$\varepsilon = 30\%$	Baseline	+2.70%	+0.82%
$\varepsilon = 40\%$	Baseline	+4.09%	+1.16%
$\varepsilon = 50\%$	Baseline	+6.45%	+1.96%

costs that exceed optimal costs by only 1% to 2%, for different levels of schedule perturbations. In addition, results suggest that the look-ahead algorithm performs consistently well for different schedule updates. For instance, for the 20 simulated schedules corresponding to $\varepsilon = 40\%$ and $\varepsilon = 50\%$, the error of policy $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$ ranges from 0.63% to 2.76%.

We proceed similarly to test the performance of the policies $(\widehat{RC}_t, \widehat{\mu}_t^a)$ and $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$ in the case of a change in the weather forecast. As described in Section 3.2.1, the original policy $(\overline{RC}_t, \overline{\mu}_t^a)$ was obtained for an “all-VMC” day, i.e., with weather transition probabilities equal to $p_t = 0$ and $q_t = 1$ for all periods t . We now introduce weather variations. First, we simulate 20 deterministic weather scenarios for the entire day of operations and define the corresponding values of p_t and q_t —note that, in the case of a deterministic forecast, p_t and q_t take only the values 0 and 1. Then, we simulate 20 deterministic weather forecasts for half the day of operations (i.e., between 6 a.m. and 3 p.m.). We set the corresponding values of p_t and q_t for the half-day (again, equal to 0 or 1). For the remainder of the day (i.e., between 3 p.m. and midnight), we estimate the values of p_t and q_t by their maximum likelihood estimator using historical records of operations (see Section 2.5). This aims to capture situations where weather forecasts are available for a limited time only and later forecasts exhibit uncertainty. In both cases, we add a 21st scenario corresponding to an “all-IMC” day (or half-day). This represents the largest weather forecast error.

Since expected arrival and departure queue lengths are larger under deteriorated weather,

the practical queue capacity $N = 30$ might be insufficient for an “all-IMC” day with heavy traffic. We therefore consider, for this particular set of tests, the schedule of flights on 09/18/2007, which corresponds to the median of the distribution of the number of daily flights at JFK in 2007. We sort the two sets of 20 weather forecasts by increasing error of the look-ahead algorithm and report the results in Figure 3-6.

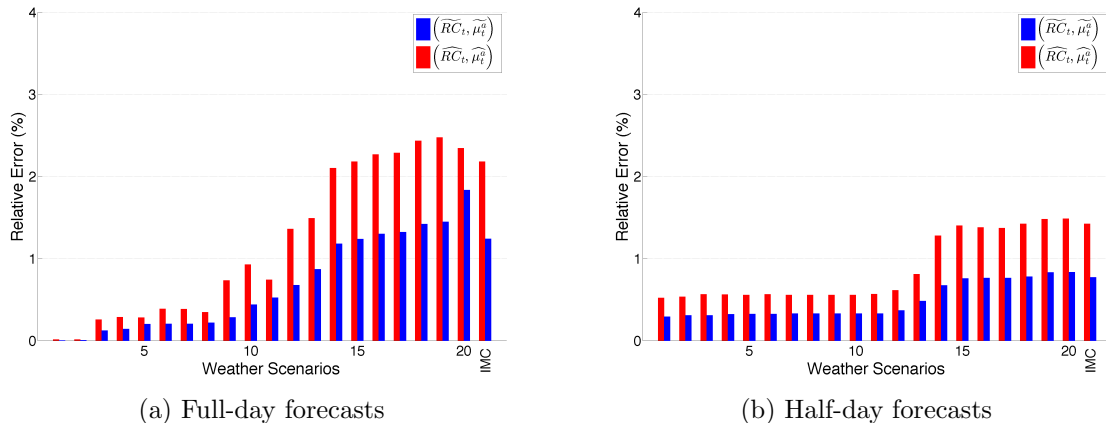


Figure 3-6: Total expected congestion costs with the three policies, under weather updates

Note, first and foremost, that the error of the two available policies, i.e., $(\widehat{RC}_t, \widehat{\mu}_t^a)$ and $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$ is extremely small. The error of the original policy $(\widehat{RC}_t, \widehat{\mu}_t^a)$ is within 1% to 3% under all weather scenarios. Again, the one-step look-ahead algorithm results in a significant performance improvement and the cost of the updated policy $(\widetilde{RC}_t, \widetilde{\mu}_t^a)$ is within 2% of the optimal congestion costs in all scenarios considered. Moreover, the error of the two policies seems much smaller in this case than in the case of a schedule update. This might be explained by the fact that the VMC and IMC Operational Throughput Envelopes (Figure 3-4) are relatively similar to each other, so optimal policies are similar in VMC and IMC. Finally, note that the “all-IMC” scenario does not yield the largest error, which might suggest that *changes* in weather conditions might cause larger policy updates than *errors* in weather forecasts.

In conclusion, the solution architecture shown in Figure 3-5, combining an exact DP algorithm and one-step look-ahead algorithm provides a fast and accurate approximation of

optimal policies. It therefore provides a flexible on-line decision-making tool to help minimize congestion costs by dynamically controlling runway configurations and arrival and departure service rates using real-time information on flight schedules and meteorological conditions.

3.3 Computational Results

In this section we present the results of the application of the exact DP algorithm to JFK. Section 3.3.1 characterizes the optimal policies. Section 3.3.2 shows the frequency of use of different runway configurations and of balances of arrival and departure service rates over the day when the optimal policy is applied. Section 3.3.3 discusses the sensitivity of expected queue lengths to several model parameters. In order to isolate the effects of the schedule of flights and of queue stochasticity on optimal policies and queue lengths, we restrict the presentation of our computational results to an all-VMC day (i.e., $p_t = 0$ and $q_t = 1, \forall t$), but similar results are obtained when different weather forecasts are considered. All results shown are obtained with the schedule on 05/25/2007, unless otherwise specified.

3.3.1 Optimal Policies

The optimal policy derived from the exact DP algorithm is a function that determines the runway configuration and the arrival and departure service rates at each period of the day and in any state of the system. Figure 3-7 plots the contours of the optimal arrival rate μ_t^a and the optimal runway configuration RC_t for the period t that begins at 12:00, for a value of $\tau_I = 5$ minutes, when the airport operates in Wind State 1 (i.e., all runway configurations can be used), as a function of the arrival and departure queue lengths that are observed at the beginning of the period (i.e., at 12:00). Note that the departure rate is not shown explicitly in the figure, but it is uniquely determined through the VMC Operational Throughput Envelope of the selected runway configuration. During the period considered, 8 landings and 4 takeoffs were scheduled, and, more generally, more arrivals than departures were scheduled between 11:45 and 13:00 (see Figure 2-4a). Two cases are considered: Figure 3-7a (resp.

Figure 3-7b) shows the optimal policy at 12:00 when the airport operated in the previous period (i.e., between 11:45 and 12:00) in runway configuration $13L, 22L|13R$ (resp. in runway configuration $22L|22R, 31L$).

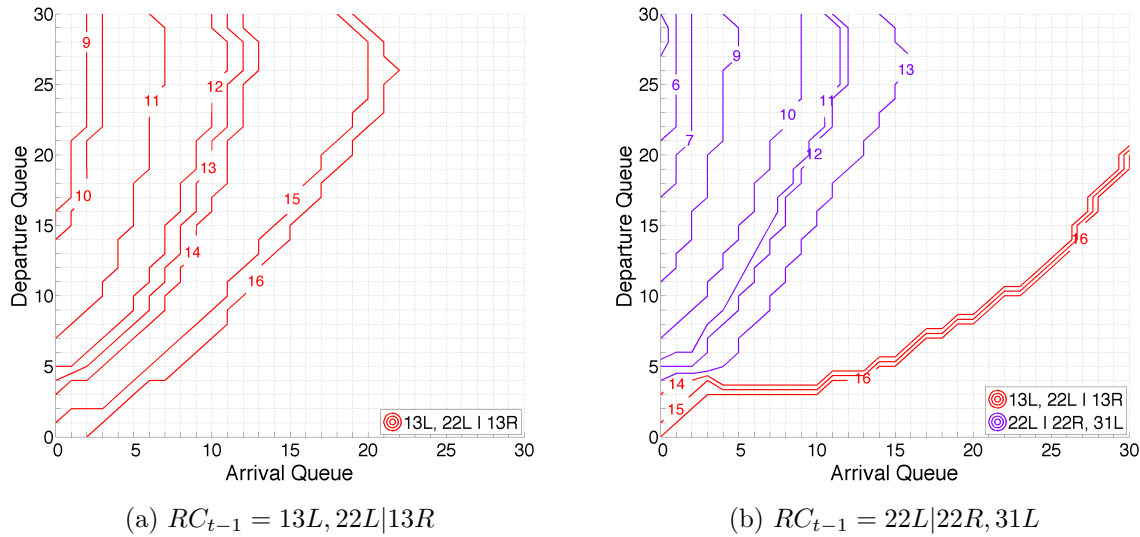


Figure 3-7: Optimal runway configuration and arrival rate at 12:00 ($\tau_I = 5$ min, $wst = 1$)

Several observations can be made about the optimal policy. First, the optimal arrival rate is non-decreasing as a function of the arrival queue length and non-increasing as a function of the departure queue length, with the exception of some “boundary effects” when queue lengths approach the practical queue capacity N . In other words, the longer the arrival queue, the more available capacity should be utilized to operate arriving aircraft. Moreover, the optimal policy depends on the runway configuration in use. For example, when the airport operates in configuration $13L, 22L|13R$, the optimal policy is to stay in this configuration, which activates two runways to operate arrivals and one runway to operate departures, in order to serve a larger number of arriving aircraft during the considered period (Figure 3-7a). In contrast, if the airport operates in configuration $22L|22R, 31L$, which activates two runways to operate departures and one runway to operate arrivals, it may be optimal to switch to configuration $13L, 22L|13R$ if the departure queue is small enough or, equivalently, if the arrival queue is large enough, or to stay in configuration $22L|22R, 31L$ otherwise

(Figure 3-7b). In addition, whereas the selected arrival rate increases quite smoothly as a function of the arrival queue length in the case where $RC_{t-1} = 13L, 22L|13R$ (Figure 3-7a), it increases discontinuously from 13 to 16 when $RC_{t-1} = 22L|22R, 31L$ (Figure 3-7b).

To illustrate these points, let us consider the evolution of the optimal arrival rate as a function of the arrival queue length observed at the beginning of period t , when the departure queue length is equal to 10 aircraft. If configuration $13L, 22L|13R$ was in use in period $t - 1$ (Figure 3-7a), the decision-maker operates in the same configuration in period t . The longer the arrival queue observed at the beginning of the period, the larger the selected arrival rate. When 12 or more arriving aircraft are queuing, then the decision-maker selects the largest arrival rate available (16 in this case). In contrast, if configuration $22L|22R, 31L$ was in use in period $t - 1$ (Figure 3-7b), then the decision-maker stays in configuration $22L|22R, 31L$ if the arrival queue length is sufficiently small. The optimal arrival service rate increases quite smoothly from 9 when no arriving aircraft are queuing to 13 when 7 arriving aircraft are queuing. As the arrival queue length increases from 7 to 21, the optimal policy remains invariant: the decision-maker selects the largest arrival rate that can be achieved under the runway configuration in use (13 in this case). If, however, 22 or more arriving aircraft are queuing, then the decision-maker switches to configuration $13L, 22L|13R$ in order to increase the arrival rate (to 16 in this case). In other words, when the arrival queue exceeds a certain threshold, then it might become beneficial to switch to another configuration that enhances the arrival throughput, in this case to configuration $13L, 22L|13R$, if the operational benefits associated with the switch become large enough to offset the costs associated with the time period of idleness following the runway configuration change.

3.3.2 Frequency of Decisions

In this section, we compute the state probabilities of the system over the day when the optimal policy is applied and we report the frequency of decisions at each period. Figure 3-8 shows the frequency of use for each of the four sets of runway configurations defined in Section 3.2.1 at each period of the day, for different values of τ_I . As expected, the

frequency of use for these different configurations depends on the throughput they achieve (Figure 2-5). Since the three-runway configurations achieve the largest service rates, they are the most frequently used ones. In contrast, the two-runway configurations are mostly used in adverse wind conditions when the airport can only operate on a small subset of runways. Configurations including the two widely-spaced parallel runways, 13/31, are used more frequently than configurations with the two more closely-spaced parallel runways, 4/22, because of the significant difference in the capacity of the corresponding configurations.

Moreover, the exact timing of use of the different configurations depends on the arrival and departure schedules, as well as the evolution of the operations through the day. Importantly, note that no runway configuration is used 100% of the time at any period of the day. This indicates that the stochasticity of the system has an impact on the optimal control, as suggested by Figure 3-7. Finally, the use of different runway configurations depends on the value of the parameter τ_I . Indeed, in the case where $\tau_I = 0$ min, runway configuration changes are very frequent to make the best possible use of available capacity. For larger values of τ_I , however, the cost associated with the idleness of the runway system is more likely to exceed the operational benefits associated with switching from one configuration to another and consequently runway configuration changes become less frequent. For instance, if $\tau_I = 0$, the decision-maker should operate, whenever possible, in a configuration with two departure runways between 13:00 and 14:00 to best serve the departure peak at that time (see Figure 2-4). As τ_I increases, decisions trade off congestion costs with increasing switching costs, depending on the observed number of queuing aircraft on the ground and in the air. As a result, the frequency of a switch between 13:00 and 14:00 becomes smaller. When $\tau_I = 10$ min, then it is almost always optimal to stay in a configuration with two arrival runways and one departure runway for the entire period between 11:00 and 17:00.

As mentioned in Section 3.1.4, the modeling framework developed in this chapter enables us to introduce differentiated costs of runway configuration changes as a function of the proximity of configurations. We compare in Figure 3-9 the use of runway configurations when the duration of idleness following a runway configuration change, τ_I , is uniform across

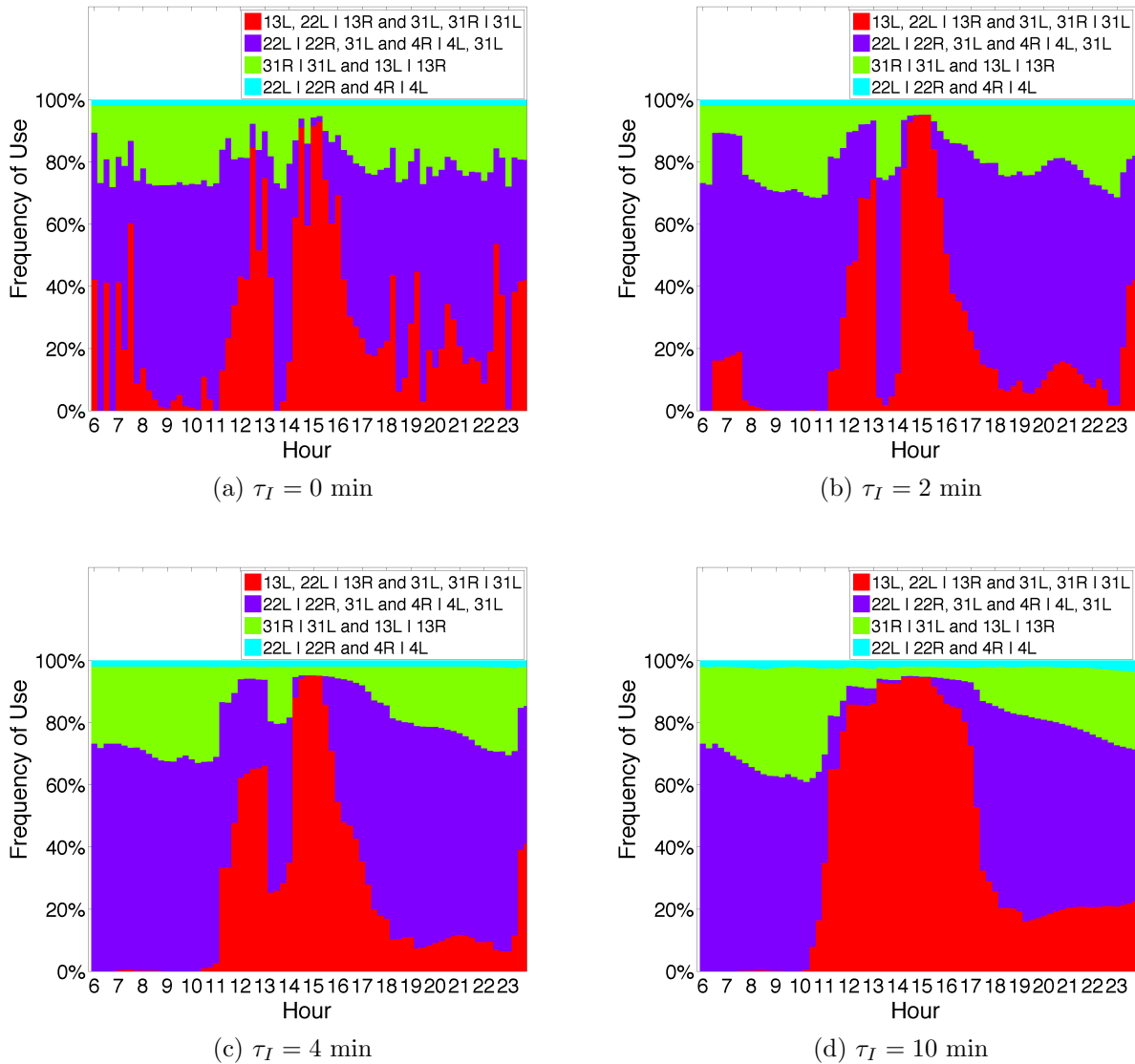


Figure 3-8: Use of runway configurations for different values of τ_I

runway configuration changes (Figure 3-9a) to the case where the values of τ_I vary with the runway configuration change considered (Figure 3-9b). Specifically, we assume in Figure 3-9b values of τ_I equal to (a) 1 minute if the switch merely disturbs operations by simply adding or removing a third runway (e.g., when switching between $31L, 31R|31L$ and $31R|31L$), (b) 5 minutes if the switch involves a 90-degree reorientation of the flow of aircraft (e.g., from $22L|22R, 31L$ to $31R|31L$) and (c) 10 minutes if the switch involves a 180-degree reorientation

of the flow of aircraft (e.g., from $31L, 31R|31L$ to $13L|13R$). In contrast, we assume in Figure 3-9a a single value of τ_I equal to 3 minutes, which is approximately the average value of τ_I in the differentiated case. Any differences between Figures 3-9a and 3-9b can thus be attributed essentially to the differences in the *distribution* of τ_I across configuration changes.

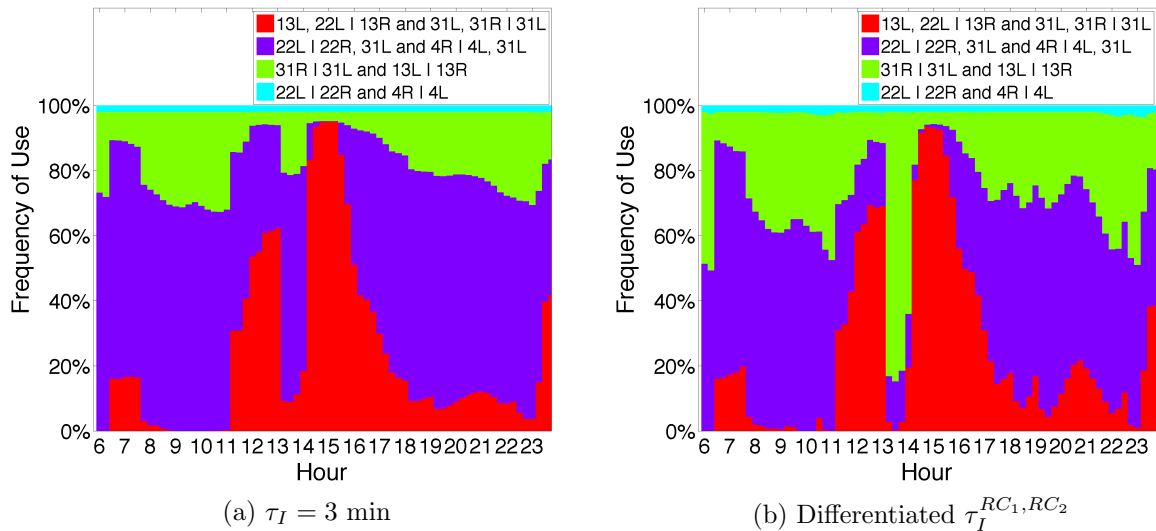


Figure 3-9: Use of runway configurations with uniform vs. differentiated values of τ_I

As expected, differentiated costs of configuration changes can impact significantly the optimal policies. For instance, it may be optimal to switch from a configuration with two arrival runways and one departure runway at 13:00 to a configuration that enables the selection of a larger departure rate in order to best serve the large number of takeoffs between 13:00 and 14:00. Even though the departure throughput could be maximized by using a configuration with two departure runways, it may be optimal to operate during this period in a configuration with two independent parallel runways (i.e., $13L|13R$ or $31R|31L$). This is because the costs of switching from configuration $13L, 22L|13R$ to $13L|13R$ or from $31L, 31R|31L$ to $31R|31L$ may be lower than the costs of switching to $22L|22R, 31L$ or to $4R|4L, 31L$. In this case, it may be optimal to increase moderately the departure throughput by using a “closer” configuration than to increase the departure throughput to a greater extent at a larger operational cost. In turn, the distribution of τ_I across configuration changes

can affect congestion costs. We estimate that the application of the optimal policy obtained with the same value of τ_I for all configuration changes in the case where the values of τ_I vary with configuration changes leads to an increase in total expected congestion costs by 5.19%, as compared to the application of the optimal policy obtained with differentiated values of τ_I . Results from these tests suggest that the fine-grain and flexible calibration by airport operators of the operational costs associated with runway configuration changes enabled by our DP model can significantly improve the efficiency of capacity utilization policies.

In addition to selecting runway configurations (Figures 3-8 and 3-9), the model introduced in this chapter controls the arrival and departure service rates at each period of the day. The frequency of these decisions is shown in Table 3.4 for seven different periods and the six most frequently used configurations. After a particular runway configuration has been selected, any variations in selected arrival and departure service rates are solely motivated by differences in prior stochastic queue evolution, and depend neither on the runway configuration previously in use nor on wind conditions. As seen in the table, for some periods of the day (e.g., at 6:45 and 15:15) the choice of arrival and departure service rates depends only weakly on the prior evolution of arrival and departure queues. In these cases, the main control exercised is the selection of the runway configuration, primarily determined by previous runway configurations and wind-related constraints, but the optimal balance of arrivals and departures does not vary substantially from one sample to another. For instance, the decision-maker selects most frequently the largest arrival rate available at 15:15 and an arrival rate equal to 8 or 9 landings per 15-minute period at 6:45. In contrast, in many other cases, the optimal balance of arrivals and departure is highly variable and depends on the observed extent of congestion at the time of the decision (e.g., at 7:45, 10:30, 11:45, 17:00 and 22:15). In such cases, both the optimal runway configuration and the optimal arrival and departure service rates might depend on the prior evolution of the system.

These results show the path-dependency of the optimal policy. At each period, the optimal runway configuration and service rates depend on the state of the system at the time of the decision (Figure 3-7), which itself depends on previous decisions and on the prior

Table 3.4: Policy frequency for six different periods of the day

Policy			Period t						
RC_t	μ_t^a	μ_t^d	6:45	7:45	10:30	11:45	15:15	17:00	22:15
13L, 22L 13R	16	8.0	–	–	2%	4%	42%	3%	–
	15	8.5	–	–	–	5%	–	–	–
	14	9.0	–	–	–	8%	–	4%	1%
	13	9.3	–	–	–	4%	–	1%	–
	11	9.9	–	–	–	5%	–	7%	1%
	10	10.1	–	–	–	4%	–	9%	2%
	9	10.2	–	–	–	2%	–	3%	3%
31L, 31R 31L	16	5.2	–	–	–	2%	52%	8%	–
	12	7.7	–	–	–	3%	–	–	–
	11	8.3	–	–	–	3%	–	–	–
	7	10.7	–	–	–	1%	–	3%	–
22L 22R, 31L	13	8.8	–	8%	5%	8%	–	9%	4%
	12	9.4	–	5%	1%	3%	–	4%	7%
	10	10.5	–	7%	2%	3%	–	5%	4%
	9	11.0	49%	10%	7%	2%	–	11%	8%
	8	11.3	–	3%	–	–	–	–	–
	7	11.7	1%	7%	7%	1%	–	3%	5%
	6	12.0	–	3%	3%	–	–	1%	2%
	5	12.3	–	1%	3%	–	–	–	1%
4	12.6	–	1%	6%	–	–	–	1%	
4R 4L, 31L	13	8.2	–	4%	2%	4%	–	6%	2%
	10	10.0	–	5%	4%	4%	–	7%	6%
	8	10.8	20%	2%	3%	1%	–	1%	3%
	7	11.2	2%	3%	3%	1%	–	1%	2%
	6	11.6	–	2%	3%	–	–	1%	2%
	4	12.4	–	1%	4%	–	–	–	2%
31R 31L	13	8.3	–	3%	1%	2%	3%	2%	1%
	10	10.0	–	3%	1%	1%	–	1%	2%
	9	10.5	4%	5%	4%	1%	–	1%	5%
	8	10.8	13%	–	1%	–	–	–	–
	7	11.2	2%	3%	5%	1%	–	–	3%
	4	12.1	–	1%	5%	–	–	–	1%
13L 13R	7	11.1	6%	2%	1%	–	–	–	3%
	6	11.4	–	1%	6%	–	–	–	2%

evolution of the system. This includes, first, some deterministic components, such as the runway configuration in use. It also includes exogenous stochastic components, such as the evolution of weather and wind conditions. For instance, configuration $4L|4L, 31L$ achieves a slightly lower departure throughput than configuration $22L|22R, 31L$. Nonetheless, it may be used when the balance of arrivals and departures requires the use of a configuration with two arrival runways and one departure runway and when strong winds from the North prevent runways $22L$ and $22R$ from being used. Last, and perhaps most important, optimal policies depend on endogenous stochastic components, such as the length of the arrival and departure queues. These lengths depend on previous decisions, as well as on the stochastic evolution of the system. This stochasticity gives rise to some variability in the optimal control at any period of the day (see Figures 3-8 and 3-9 and Table 3.4): both the runway configurations and the arrival and departure service rates depend on prior system evolution. This underscores the importance of considering queue stochasticity and its endogeneity to capacity utilization decisions within our decision-making framework, because of the impact this may have on optimal decisions. We further estimate the benefits of accounting for queue stochasticity in Section 3.4.2.

3.3.3 Sensitivity of Queue Lengths to Model Parameters

We have performed several sensitivity analyses to evaluate how the optimal policies and the optimal arrival and departure queue lengths vary with several model parameters. First, Figure 3-10 shows optimal expected arrival (Figure 3-10a) and departure (Figure 3-10b) queue lengths for different values of the weight α of the cost of arrival queuing and for a duration of idleness following runway configuration changes equal to $\tau_I = 5$ minutes. In other words, it shows the effect of an increase of the relative cost of arrival queues, compared to departure queues, on expected arrival and departure queue lengths. As expected, larger values of α lead to policies that increase arrival throughput at the expense of the departure throughput, and therefore to shorter expected arrival queues and, conversely, to longer expected departure queues. Arrival queues seem more sensitive to variations in α than departure queues.

Variations in α from 1 to 2 induce changes in peak expected arrival queue length of the order of 2 to 3 aircraft in queue (i.e., over 30%) and changes in peak expected departure queue length of the order of 1 to 2 aircraft in queue (i.e., 5% to 10%). This difference might be due to the slope of the Operational Throughput Envelope at JFK being lower than 1 (Figure 2-5), so variations in arrival rates induce smaller variations in departure rates.

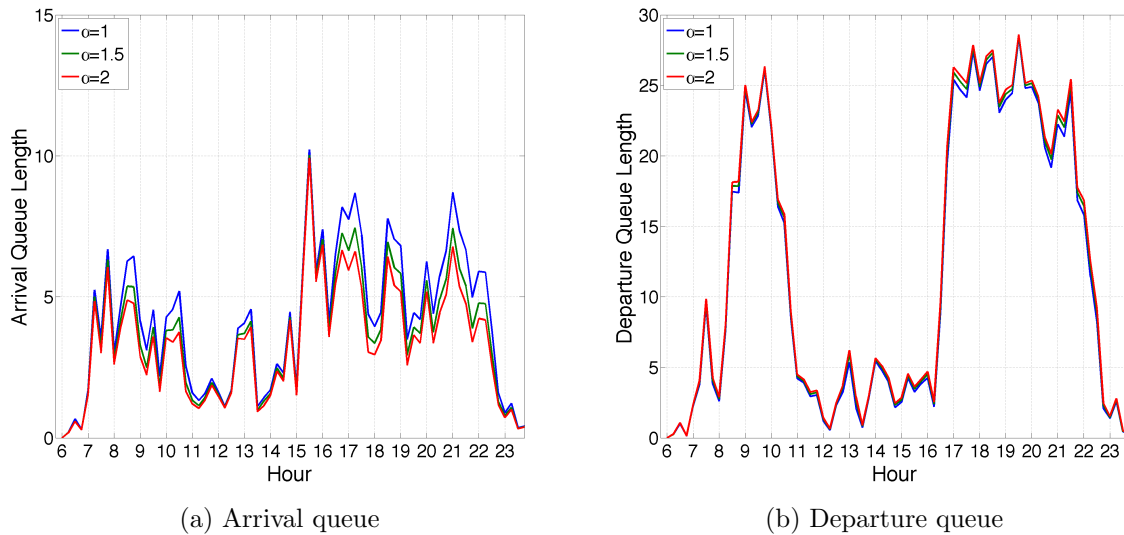


Figure 3-10: Sensitivity of expected arrival and departure queues to α ($\tau_I = 5$ minutes)

As well, variations in the duration of idleness following a runway configuration change, τ_I , induce changes in the efficiency of operations and in the optimal policy (see Section 3.3.2). Figure 3-11 shows the sensitivity of expected arrival (Figure 3-11a) and departure (Figure 3-11b) queue lengths to the value of τ_I . As τ_I increases from 0 to 15 minutes, the expected arrival queue length might increase by 1 to 2 aircraft and the expected departure queue length might increase by 2 to 3 aircraft. This translates into increases in the optimal expected congestion costs, reported in Table 3.5. Total expected costs increase by 25% to 40% when τ_I increases from 0 to 15 minutes. Therefore, the efficiency with which airports can switch between runway configurations has a significant impact on expected airport congestion costs.

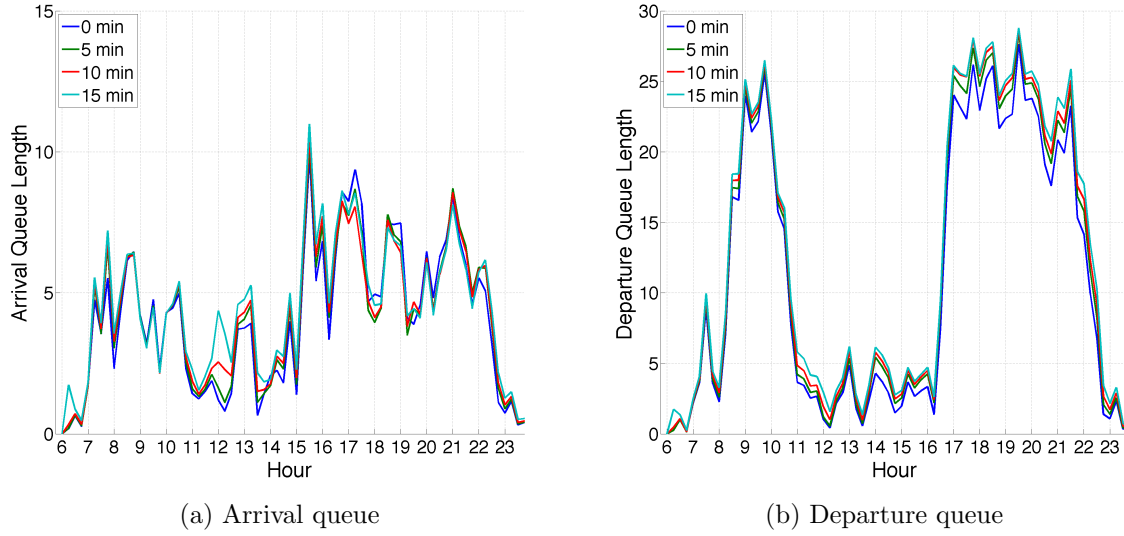


Figure 3-11: Sensitivity of expected arrival and departure queues to τ_I ($\alpha = 1$)

Table 3.5: Expected total congestion costs for different values of τ_I

Day	Flights	Expected total cost			
		$\tau_I = 0$ min	$\tau_I = 5$ min	$\tau_I = 10$ min	$\tau_I = 15$ min
02/04	1,027	Baseline	+18.15%	+28.37%	+37.16%
01/10	1,059	Baseline	+17.57%	+29.13%	+39.35%
02/08	1,085	Baseline	+18.31%	+29.93%	+39.23%
01/25	1,097	Baseline	+17.38%	+27.79%	+36.38%
09/18	1,113	Baseline	+13.67%	+23.52%	+32.17%
10/15	1,125	Baseline	+15.93%	+25.45%	+33.81%
06/01	1,133	Baseline	+14.43%	+24.19%	+34.18%
07/07	1,160	Baseline	+10.58%	+17.38%	+24.89%
05/25	1,172	Baseline	+12.19%	+19.22%	+26.23%

3.4 Performance Evaluation

We now quantify the benefits resulting from the control of runway configurations and of arrival and departure service rates developed in this chapter. We first compare the optimal policy to two advanced heuristics that aim to imitate typical decisions made in practice. We then compare it to an alternative model based on deterministic queue dynamics to quantify the benefits of integrating queue stochasticity into the decision-making framework.

3.4.1 Comparison of the Optimal Policy to Heuristics

We compare the optimal policies obtained with the DP model to two heuristics. The heuristics can be viewed as representative of the kind of “reasonably smart” capacity utilization policies one might apply in practice, in the absence of an advanced model such as the DP model presented here. If our model can outperform the heuristics by a significant margin, this would suggest that large congestion cost savings might result from the model’s implementation as a tool for tactical decision-making. Comparisons of the associated congestion costs provide estimates of the potential savings that may result from such implementation.

Note that, ideally, one would compare the policies recommended by the model to actual decisions made in practice. However, such comparison is complicated by several factors. First, identifying arrival and departure queue lengths and service rates from available databases is subject to significant uncertainty. Second, little data on real-time information available to decision-makers in practice (e.g., dynamic schedule updates) is reported. Third, establishing statistically reliable comparisons and isolating the benefits of our DP model requires the consideration of many days of operations, which is complicated by the stochasticity of the system and variations in the schedules of flights and operating conditions from one day to another. For these reasons, we approximate actual decisions through the two heuristics described below. This approach is similar to the one adopted in the literature [35, 18]. Note, however, that some differences with actual decisions might also arise from practical factors that were not included in the model. Discussions with air traffic managers are therefore required to design iterative improvements of our model and of its implementation.

The two heuristics we consider were developed from discussions with air traffic managers at JFK. Both adjust the arrival rate as a function of arrival demand, which seems to be common practice at the busiest airports. This might be motivated by the fact that, at JFK, the departure rate is not very sensitive to the arrival rate. Indeed, as previously discussed, the departure rates shown by the Operational Throughput Envelopes of the main runway configurations at JFK decrease by less than 1 for every unit increase in the arrival rates (Figure 2-5). Therefore, increasing the arrival rate to match arrival demand induces a

relatively small loss in departure throughput. Moreover, this might be motivated by practical reasons as well, since it is more challenging operationally to hold arriving aircraft in queue in the air than to hold departing aircraft in queue on the ground.

The first heuristic assumes no cost of changing runway configurations. At the beginning of each period, the decision-maker computes the *effective arrival demand*, denoted by ad_t and defined as the expected number of aircraft that will request to land in the period. It is simply equal to the sum of the arrival queue at the beginning of period t , i.e., a_{t-1} , and the expected number of aircraft that will join the arrival queue in period t , i.e., x_t . He then attempts to satisfy arrival demand by selecting an arrival rate equal to the effective arrival demand, if possible. If the maximal arrival rate of each of the runway configurations that can be used under observed wind conditions is smaller than the effective arrival demand, then he selects the largest arrival rate that can be possibly chosen. He then operates in the runway configuration that maximizes the expected departure throughput for the selected arrival rate. This heuristic is presented in Algorithm 1.

Algorithm 1 Heuristic 1: Arrival priority and no cost of runway configuration changes

```

for  $t = 1, \dots, T$  do
  for  $a_{t-1}$  do
    • Compute effective arrival demand:  $ad_t = a_{t-1} + x_t$ 
    for  $wc_t, ws_t$  do
      for  $rc \in \mathcal{RC}(ws_t)$  do
        • Define candidate arrival rate with  $rc$ :  $ar_t(rc) = \min(A_{rc,wc_t}, ad_t)$ 
      end for
        • Determine the set of candidate runway configurations  $\mathcal{RC}_{ar_t}(ws_t) \subset \mathcal{RC}(ws_t)$  that minimize the quantity  $|ad_t - ar_t(rc)|$  and define new arrival rate  $\mu_t^a$  as the corresponding value of  $ar_t(rc), rc \in \mathcal{RC}_{ar_t}(ws_t)$ 
        • Choose runway configuration  $RC_t = \arg \max_{rc \in \mathcal{RC}_{ar_t}(ws_t)} \{\Phi_{rc,wc_t}(\mu_t^a)\}$ 
        • Define new departure rate:  $\mu_t^d = \Phi_{RC_t,wc_t}(\mu_t^a)$ 
      end for
    end for
  end for
end for

```

The second heuristic relies on a similar approach except that a large cost of switching runway configurations is assumed. The decision-maker does not change runway configurations

unless he has to do so, because of changing wind conditions. If the current runway configuration can still be used for the next period, then he chooses the arrival rate that matches the effective arrival demand as closely as possible and the departure rate is subsequently determined through the Operational Throughput Envelope of the runway configuration. If, however, the current runway configuration can no longer be used, then he follows the policy determined according to the first heuristic. This heuristic is presented in Algorithm 2.

Algorithm 2 Heuristic 2: Arrival priority and large cost of runway configuration changes

```

for  $t = 1, \dots, T$  do
  for  $RC_{t-1}$  do
    • Determine the set of wind states  $\mathcal{WS}_{RC_{t-1}}$  in which  $RC_{t-1}$  can be feasibly used
      for  $ws_t \in \mathcal{WS}_{RC_{t-1}}$  do
    • Choose same runway configuration:  $RC_t = RC_{t-1}$ 
      for  $a_{t-1}, wc_t$  do
    • Compute effective arrival demand:  $ad_t = a_{t-1} + x_t$ 
    • Define new arrival rate:  $\mu_t^a = \min(A_{RC_t, wc_t}, ad_t)$ 
    • Define new departure rate:  $\mu_t^d = \Phi_{RC_t, wc_t}(\mu_t^a)$ 
      end for
    end for
    for  $ws_t \notin \mathcal{WS}_{RC_{t-1}}$  do
    • Choose policy from Heuristic 1 in all states
      end for
    end for
  end for

```

Table 3.6 reports the relative difference between the optimal expected congestion costs and the expected congestion costs resulting from the application of each of these two heuristics, for different values of the duration of idleness τ_I . As expected, Heuristic 2 performs better for the larger values of τ_I (e.g., $\tau_I = 10$ min), while Heuristic 1 performs better for the smaller values of τ_I (e.g., $\tau_I = 0$ min). This is consistent with the design of these heuristics. But in all the cases considered ($\tau_I = 0$ min, $\tau_I = 5$ min and $\tau_I = 10$ min), the *better* of the two heuristics results in significantly greater congestion costs than the optimal policy, by an estimated 20% to 30%. These results suggest that the optimal control might result in substantial cost savings. Finally, the estimated 20% to 30% reduction in congestion costs is much larger than the 1% to 2% error of the one-step look-ahead algorithm imple-

mented in Section 3.2.3. This suggests that, when real-time disturbances are considered, the approximate scheme developed in this chapter would still result in significant operational improvements compared to the two advanced heuristics considered in this section.

Table 3.6: Relative error of the two heuristics

Day	Flights	Exact	Heuristic 1		Heuristic 2	
			$\tau_I = 0$ min	$\tau_I = 5$ min	$\tau_I = 5$ min	$\tau_I = 10$ min
02/04	1,027	Baseline	+35.18%	+56.61%	+45.28%	+36.88%
01/10	1,059	Baseline	+35.34%	+75.58%	+51.48%	+41.82%
02/08	1,085	Baseline	+37.60%	+68.60%	+46.84%	+37.17%
01/25	1,097	Baseline	+34.60%	+63.13%	+44.37%	+35.79%
09/18	1,113	Baseline	+28.23%	+57.18%	+37.86%	+29.76%
10/15	1,125	Baseline	+30.19%	+57.29%	+38.13%	+30.70%
06/01	1,133	Baseline	+29.65%	+65.19%	+45.97%	+37.75%
07/07	1,160	Baseline	+19.07%	+41.27%	+29.95%	+25.01%
05/25	1,172	Baseline	+25.02%	+43.70%	+31.48%	+26.08%

The comparison of the optimal control to heuristics suggests that the joint control of runway configurations and of arrival and departure service rates can improve the efficiency of airport operations substantially. As mentioned in Chapter 1, the annual costs of air traffic delays in the United States were estimated at over \$30 billion for the year 2007 [5], 50% to 75% of which are attributed to mismatches between demand and capacity [26]. Given the disproportionate distribution of delays across airports and the propagation of these delays through the National Aviation System, the implementation of the control developed in this chapter at a few of the busiest airports in the United States could generate very significant cost savings for airlines, passengers and society.

3.4.2 Benefits of the Integration of Queue Stochasticity

Finally, we quantify the benefits of integrating the stochasticity of queue dynamics into the decision-making framework. To this purpose, we implement an alternative version of the model developed in this chapter, but with deterministic queue dynamics. In this case, the

queue transition probabilities (see Section 3.1.4) are simply defined as follows:

$$Q_{m,n}^t = \begin{cases} 1, & \text{if } n = m + \lambda_t - \mu_t \\ 0, & \text{otherwise} \end{cases} \quad (3.7)$$

Using (3.7), we derive the optimal policy under deterministic queue dynamics and we compare it to the optimal policy based on stochastic queue dynamics that we obtained previously. Note that both policies consider the same model of weather and wind dynamics. Any difference is therefore due to the consideration, or not, of queue stochasticity. We then simulate the evolution of the system resulting from the application of the optimal deterministic and stochastic policies, under stochastic queue dynamics.

We compare in Table 3.7 the congestion costs obtained with each of the two policies, for different values of τ_I . These results suggest that accounting for the stochasticity of queue dynamics in the design of the operating policy might yield significant congestion cost savings, which we estimate from our tests at 5% to 20%. Note that the benefits of integrating queue stochasticity seem larger with the larger values of τ_I (e.g., $\tau_I = 5$ min and $\tau_I = 10$ min) than with $\tau_I = 0$ min. This might be explained by the fact that, with $\tau_I = 0$, the deterministic model is able to atone, to some extent, for its lack of consideration of stochasticity. It does this by making more frequent changes (at no cost) to runway configurations in response to unexpected changes in the arrival and departure queues.

In conclusion, the stochasticity of queue dynamics has a significant impact on the optimal policy. Ignoring this stochasticity can result in significant operating inefficiencies, estimated at 5% to 20%. This is consistent with the path-dependency of the optimal policy shown in Section 3.3.1.

3.5 Conclusion

In this chapter, we presented an original decision-making framework that dynamically controls runway configurations and the arrival and departure service rates at a major airport

Table 3.7: Relative performance of the optimal stochastic and deterministic policies

Day	Flights	$\tau_I = 0$ min		$\tau_I = 5$ min		$\tau_I = 10$ min	
		Stoch.	Determ.	Stoch.	Determ.	Stoch.	Determ.
02/04	1,027	Baseline	+9.06%	Baseline	+16.70%	Baseline	+14.31%
01/10	1,059	Baseline	+8.24%	Baseline	+14.25%	Baseline	+11.92%
02/08	1,085	Baseline	+8.05%	Baseline	+16.30%	Baseline	+12.51%
01/25	1,097	Baseline	+7.91%	Baseline	+16.45%	Baseline	+13.22%
09/18	1,113	Baseline	+6.70%	Baseline	+13.77%	Baseline	+11.48%
10/15	1,125	Baseline	+7.76%	Baseline	+17.16%	Baseline	+14.92%
06/01	1,133	Baseline	+7.73%	Baseline	+18.19%	Baseline	+16.27%
07/07	1,160	Baseline	+5.60%	Baseline	+11.50%	Baseline	+11.42%
05/25	1,172	Baseline	+6.58%	Baseline	+13.49%	Baseline	+12.84%

while taking into account the stochasticity of queue dynamics and of operating conditions. This addresses the problem of dynamically utilizing airport capacity to serve arriving and departing aircraft under operating uncertainty. We developed an efficient Dynamic Programming (DP) formulation that minimizes airport congestion costs and embeds a realistic stochastic model of queue dynamics and of weather and wind-related uncertainty.

We applied the exact DP algorithm to a realistic setting at JFK Airport and showed that, based on information available before a particular day of operations, the optimal *a priori* operating policy for that day can be obtained within reasonable computational times. However, during the day in question, the policy thus obtained may no longer be optimal due to unforeseen disturbances, such as schedule updates or changes in the weather forecast. To address this issue, we implemented a one-step look-ahead approximate algorithm that greatly speeds up the execution of the DP. Results suggest that this two-step approximation scheme yields near-optimal policies. The fast approximate algorithm therefore enables the on-line implementation of our model using real-time information.

The application of the model at JFK yielded several insights. First, optimal policies are path-dependent in the sense that they depend on the prior evolution of the system, including prior decisions and the stochastic evolution of arrival and departure queues over the day. This underscores the impact of system stochasticity on optimal policies. As a result, we showed that integrating the stochasticity of queue dynamics into the decision-making framework can

yield significant congestion cost savings, estimated at 5% to 20%. Moreover, comparisons of optimal policies to two advanced heuristics, aimed to imitate actual operating procedures, suggest that congestion costs might be significantly reduced through the control of runway configurations and of service rates developed in this chapter. Our results at JFK indicate the potential for congestion cost savings of as much as 20% to 30% at the busiest airports.

The model and the algorithms presented provide an effective decision-making tool to mitigate airport congestion at the tactical level. Its implementation may provide substantial operational benefits to airport stakeholders. Moreover, the model provides a better understanding of how airport capacity utilization procedures depend on the schedule of flights and on the stochastic evolution of the arrival and departure queues at the airport. In the remainder of this thesis, we integrate this model of capacity utilization into macroscopic models of airport congestion (Chapter 4) and into models of scheduling interventions (Chapters 5 and 6) at the strategic level.

Chapter 4

Application to Congestion Modeling

This chapter presents an original approach to airport congestion modeling that integrates our control of capacity utilization procedures into a strategic queuing model of airport congestion. As we have seen in Chapter 3, tactical decisions regarding the selection of runway configurations and the balancing of arrival and departure service rates can vary significantly as a function of the schedule of flights (which is largely set months in advance), of the capacity of the airport (which is relatively stable over long periods of time), and of the dynamics of the formation and propagation of delays over the day of operations. However, these intra-day variations in the rates at which aircraft are serviced are typically *not* considered in strategic queuing models of airport congestion, which are usually based on a single-value of airport capacity (e.g., 20 movements per 15-minute period) or on single-values of arrival and departure capacities (e.g., 10 arrivals and 10 departures per 15-minute period). The integrated model of airport congestion developed in this chapter quantifies the relationships between flight schedules, airport capacity and flight delays at the strategic level, given the way service rates can be selected over the day of operations by air traffic managers to maximize airport efficiency at the tactical level, under capacity constraints. In turn, it provides a fast and flexible tool for forecasting the evolution of delays at different airports under different demand and capacity scenarios.

The contributions of this chapter fall into four categories:

- We provide an integrated approach to airport congestion modeling that combines an endogenous control of arrival and departure service rates into a stochastic and dynamic queuing model.
- We develop an approximation of the control of arrival and departure service rates to ensure that the resulting model of airport congestion is computationally efficient.
- We show that our model approximates well the extent of congestion observed in practice at JFK, EWR and LGA.
- The application of the model suggests that the large delay reductions observed between 2007 and 2011 can be largely attributed to the comparatively small scheduling changes observed over the period.

A paper based on this chapter can be found in [62].

The remainder of this chapter is organized as follows. Section 4.1 presents the integrated queuing model of airport congestion with endogenous control of arrival and departure service rates. Section 4.2 compares the results of the model to historical records of operations at JFK, EWR and LGA. Section 4.3 applies the model to analyze recent trends in scheduling and on-time performance at these airports. Section 4.4 concludes.

4.1 Model Formulation

4.1.1 Model Presentation

Our model of airport congestion aims to quantify the magnitude of delays and their evolution over a day as a function of flight schedules and airport capacity at the *strategic* level (i.e., using information available *before* a day of operations). As detailed previously, flight schedules determine the demand rates into the queuing system, and airport capacity constrain the service rates to be applied over the day of operations. These service rates are not determined in advance, but are dynamically adjusted by air traffic managers over the day

to maximize the efficiency of airport operations. The control of capacity utilization policies described in Chapter 3 captures the endogenous relationship between arrival and departure queue lengths, on the one hand, and arrival and departure service rates, on the other. In this chapter, we combine our Stochastic Queuing Model of airport congestion and our Dynamic Programming model of capacity utilization to develop an integrated model that quantifies the relationship between flight schedules, airport capacity, and arrival and departure queue lengths, given how airport capacity is utilized at the tactical level.

Figure 4-1 illustrates schematically the difference between our integrated approach and “traditional” approaches to airport congestion modeling. Figure 4-1a shows the representation of airport capacity by means of Operational Throughput Envelopes (see Figure 2-1a). Figure 4-1b shows, in contrast, the representation of capacity as constant arrival and departure service rates used in most existing queuing models of airport congestion. The arrival and departure service rates are typically estimated by their long-term expected values, i.e., weighted averages computed by taking into consideration the frequency of use of each runway configuration and what service rate is typically achieved in each configuration. This representation can account for exogenous variations in airport capacity (e.g., weather variations). However, it fails to capture the impact on airport capacity of endogenously selecting (as is done in practice) the runway configuration and the balancing of the arrival and departure loads, as well as the impact of the schedule of arrivals and departures and of operations in the preceding and following time periods on capacity utilization procedures.

The computational requirements of the Dynamic Programming model of capacity utilization developed in Chapter 3, however, prevent it from being used repeatedly with different flight schedules or different capacity estimates and thus limit its applicability in support of strategic planning. Therefore, we develop in this chapter an approximate version of the control to ensure computational efficiency of the model. The purpose of this simplification is to approximate the arrival and departure queue lengths that minimize congestion costs at the tactical level, without accounting for all the operational details that need to be considered in the full control.

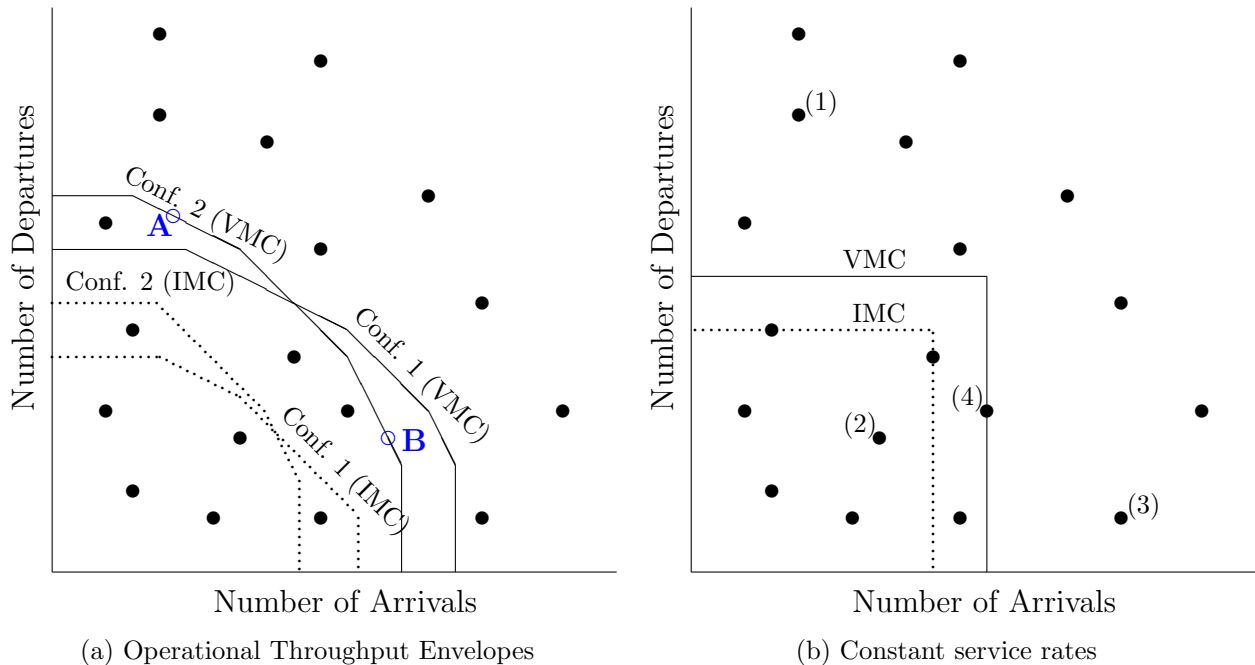


Figure 4-1: Inputs with Operational Throughput Envelopes and constant service rates

4.1.2 Simplified Control of Service Rates

The simplified control is obtained by grouping runway configurations into clusters of “similar” configurations. The objective of this clustering approach is to approximate how the trade-off between arrival and departure throughput varies with runway configurations, but at a more aggregate level than in the full control. Specifically, we cluster runway configurations based on the number of runways they use to operate landings and takeoffs and on their physical layout. We estimate the average capacity of the airport for each cluster of runway configurations, i.e., we consider a *single* Operational Throughput Envelope for each cluster. Table 4.1 shows the clusters considered at the three New York airports, based on the number of runways they use to operated arrivals and departures. For instance, at JFK, the first of the listed “Type 2” configurations utilizes two runways (13L and 22L) to operate arrivals and one runway (13R) to operate departures.

To simplify the model further, we assume that the schedule of use of runway configurations clusters is exogenously determined in advance. It is obtained from the full control with a

Table 4.1: Runway configuration clusters at JFK, EWR and LGA

	“Type 1” Conf.	“Type 2” Conf.	“Type 3” Conf.
Characterization	1 arrival runway 1 departure runway	2 arrival runways 1 departure runway	1 arrival runway 2 departure runways
JFK	31R 31L 13L 13R	13L, 22L 13R 31L, 31R 31L	22L 22R, 31L 4R 4L, 31L
EWR	22L 22R 4R 4L 29 22R	11, 22L 22R 4R, 11 4L 4R, 29 4L	22L 22R, 29
LGA	31 4 22 31 22 13 4 13	—	—

representative schedule of flights and from the actual patterns of runway configuration usage at the airports. Specifically, we assume that JFK operates in a “Type 3” configuration in the morning and in the late afternoon, and in a “Type 2” configuration between 11:45 and 17:00, in order to best serve the arrival peak in this time window. As well, we assume that EWR operates in a “Type 1” configuration in the morning and in a “Type 2” configuration from 10:30 onward and that LGA operates in a “Type 1” configuration for the entire day. Subsequently, the simplified control is restricted to the selection of arrival and departure service rates at the beginning of each 15-minute period, under capacity constraints defined by the Operational Throughput Envelope of the runway configuration in use. This simplification captures the trade-off between arrival and departure service rates and approximates the selection of runway configurations.

The resulting control is formulated as follows. At each period $t = 1, \dots, T$, the decision-maker observes (i) the arrival queue length at the end of period $t - 1$ (i.e., $a_{t-1} \in \{0, \dots, N\}$), (ii) the departure queue length at the end of period $t - 1$ (i.e., $d_{t-1} \in \{0, \dots, N\}$), and (iii) the weather state (i.e., $w_t \in \{VMC, IMC\}$). The runway configuration cluster for period t , which we denote by RC_t , is given. The decision-maker selects the arrival service rate for period t , i.e., $\mu_t^a \in \{0, 1, \dots, A_{RC_t, w_t}\}$. Once the arrival rate is selected, the departure

service rate μ_t^d is determined by the Operational Throughput Envelope. As in the full control, the objective function is expressed as $\alpha \sum_{t=1}^T a_t^2 + \sum_{t=1}^T d_t^2$. We use a value of $\alpha = 1$. The Bellman equation is then formulated as follows:

$$J_t(a_{t-1}, d_{t-1}, w_t) = \min_{\mu_t^a \in [0, A_{RC_t, w_t}]} \left(E[a_t^2] + E[d_t^2] + E[J_{t+1}(a_t, d_t, w_{t+1})] \right), \forall t = 1, \dots, T_0 \quad (4.1)$$

The simplification of the control reduces the dimensionality of the model considerably and therefore accelerates its solution. Table 4.2 shows the size of the original model (Chapter 3) and this simplified model. In the original control, the state space had 5 dimensions: the arrival and departure queue lengths (which can each take $N + 1$ values), the runway configuration in use (we denote the number of runway configurations by $|\mathcal{RC}|$), the weather state (VMC or IMC) and the wind state (we denote the number of wind states $|\mathcal{WS}|$). In the simplified control, the size of the state space is reduced by a factor of $|\mathcal{RC}| \times |\mathcal{WS}|$, (i.e., by a factor of 100, approximately). The decision state had 2 dimensions in the original model: the selected runway configuration (which could take $|\mathcal{RC}|$ values) and the arrival service rate. In the simplified model, the decision is restricted to the selection of the arrival service rate, which can take $A_{RC_t, w_{c_t}} + 1$ values at each period t (for notational simplicity, we denote in Table 4.2 the largest of the values $A_{RC_t, w_{c_t}}$ by \bar{A}). In turn, the simplification reduces the size of the decision space by a factor $|\mathcal{RC}|$ (i.e., by a factor of 10, approximately). These simplifications reduce the computational requirements of the control by several orders of magnitude.

Table 4.2: Size of the original and simplified controls

Model	Original Model	Simplified Model
Number of states	$(N + 1)^2 \times \mathcal{RC} \times 2 \times \mathcal{WS} $	$2 \times (N + 1)^2$
Number of decisions (upper bound)	$(\bar{A} + 1) \times \mathcal{RC} $	$\bar{A} + 1$
Number of periods	T	T

This simplification enables us to consider a realistically large value of the practical queue capacity N . In the remainder of this chapter, we use a value of $N = 50$. With this value, delay estimates for the whole day are obtained in about one minute. At the same time, the

probability of the modeled queues to be full is very small, which minimizes the downward bias introduced by the practical queue capacity. Note that the value of N considered here $N = 50$ is larger than that considered in Chapter 3. This is because the underestimation of queue lengths at peak hours affects only marginally the optimal policies, but can bias significantly the comparison of modeled queue lengths to actual queue lengths (which have practically infinite queue capacities).

4.2 Model Implementation

We apply our integrated queuing model of airport congestion with the endogenous control of service rates developed at JFK, EWR and LGA (Section 2.5) to the 4-month period from June to September 2007 (henceforth “Summer 2007”). Specifically, we apply the model to each one of the days of Summer 2007 and quantify the resulting probability distributions of queue lengths. This is motivated by potential nonlinear variations of delays with flight schedules, as suggested by steady-state queuing theory [34], so the use of the average demand profile might not provide an accurate description of the distribution of delays during the period of interest. We then compare the model’s outputs to on-time performance observed in practice during the considered period.

Before proceeding further, we underline the objectives of such comparisons. As previously mentioned, our model is strategic in scope. It is based on information that is available before the day of operations and does not consider real-time operational information, such as the runway configuration actually in use, wind conditions, the occurrence of delays at other airports, the pushback times of departing flights, etc. Therefore, the delay estimates that it provides are expected to be less accurate than estimates based on such real-time information at the tactical level (see, e.g., [44, 106]). The objective of the model is to capture, at a macroscopic level, the dynamics of formation and propagation of delays over the course of the day and the magnitude of flight delays at peak morning and afternoon hours.

We first develop on-time performance metrics obtained from available operational data in Section 4.2.1, and we then compare the queue lengths and delays predicted by the model to

those observed in practice. Specifically, we show that our model provides good estimates of the expected value and the variability of the departure queue lengths (Section 4.2.2) and of the expected value of arrival and departure delays (Section 4.2.3) at JFK, EWR and LGA. Finally, we show that our model performs significantly better than a baseline model where the control of service rates is ignored (Section 4.2.4).

4.2.1 Measures of Airport On-Time Performance

In order to compare our model’s predictions to actual operations, we develop estimates of airport on-time performance from available records of operations. Note, at the outset, that establishing apples-to-apples comparisons is not straightforward. We are interested here solely in the effects of local runway congestion, thus in identifying and estimating delays that can be attributed to it. But the available data are contaminated by delays due to an array of other factors, including the propagation of upstream delays, constraints at other airports, delays in passenger buildings (e.g., check-in delays, security screening delays, boarding delays), mechanical failures and safety incidents. The on-time performance metrics presented below aim to eliminate such external delays. It should be noted, however, that the values of these metrics are subject to uncertainty and must be treated only as approximate.

Figure 4-2 shows the different phases of a flight and defines the main quantities of interest. We denote the set of all flights on a given day by \mathcal{F} . For any flight $i \in \mathcal{F}$, we denote its gate-out time by OUT_i , its wheels-off time by OFF_i , its wheels-on time by ON_i and its gate-in time by IN_i . We also consider its taxi-out time $TO_i = OFF_i - OUT_i$, its airborne time $AIR_i = ON_i - OFF_i$ and its taxi-in time $TI_i = IN_i - ON_i$. For each of these seven quantities, we consider both the planned time, which we denote by a superscript P , and the actual time, which we denote by a superscript A . For instance, OUT_i^P (resp. OUT_i^A) denotes the planned (resp. actual) gate-out time of flight i . Finally, we define, for any flight i , its gate-out delay δ_i^{OUT} , its wheels-off delay δ_i^{OFF} , its wheels-on delay δ_i^{ON} and its gate-in

delay δ_i^{IN} as follows:

$$\begin{aligned}\delta_i^{OUT} &= \max(OUT_i^A - OUT_i^P, 0) = (OUT_i^A - OUT_i^P)^+ \\ \delta_i^{OFF} &= \max(OFF_i^A - OFF_i^P, 0) = (OFF_i^A - OFF_i^P)^+ \\ \delta_i^{ON} &= \max(ON_i^A - ON_i^P, 0) = (ON_i^A - ON_i^P)^+ \\ \delta_i^{IN} &= \max(IN_i^A - IN_i^P, 0) = (IN_i^A - IN_i^P)^+\end{aligned}$$

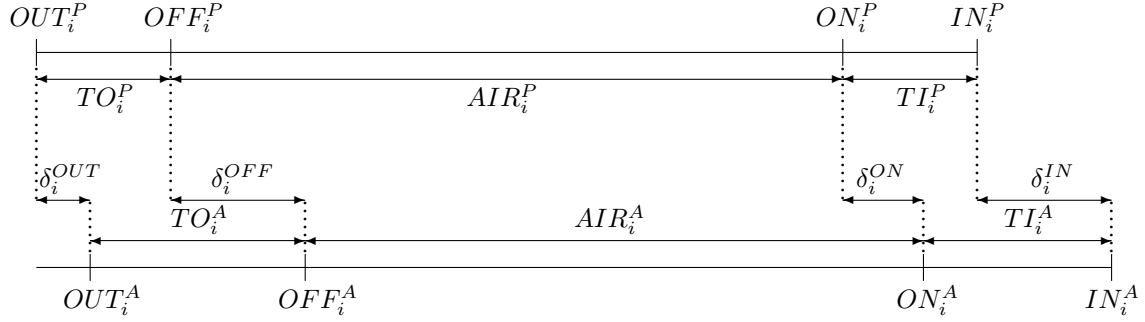


Figure 4-2: Breakdown of scheduled and actual flight times

We now derive estimates of the departure queue length at the end of each period $t = 1, \dots, T$, i.e., at time tS , where S denotes the length of each period (in this case, 15 minutes). To do so, we assume that no departure metering procedure is in place (a reasonable assumption in 2007). In other words, we assume that aircraft leave the gate as soon as they are ready to depart. In the absence of runway congestion, any departing aircraft i would take off at time $OUT_i^A + TO_i^P$, i.e., a duration equal to its unimpeded taxi-out time TO_i^P after they leave the gate. The departure queue consists of all aircraft that would have taken off at time tS in the absence of congestion, but that have not taken off yet at that time. Specifically, our estimate of the actual departure queue length at the end of period t , denoted by \hat{d}_t , is obtained as follows:

$$\hat{d}_t = \left| \{i \in \mathcal{F} \mid OUT_i^A + TO_i^P \leq tS \quad \& \quad OFF_i^A > tS\} \right|$$

Unfortunately, estimating the arrival queue length from available records of operations is almost impossible. Indeed, there is no record of when arriving aircraft demand the usage of

the runway. Physically, aircraft can be “queuing” in the terminal airspace, in the en-route airspace or at the origin airport, which complicates the estimation of the number of queuing aircraft at any time.

Next, we develop estimates of average arrival and departure delays per period $t = 1, \dots, T$. We first define the set \mathcal{A}_t (resp. \mathcal{D}_t) as the set of flights that are scheduled to arrive (resp. to depart) during period t :

$$\begin{aligned}\mathcal{A}_t &= \{i \in \mathcal{F} \mid (t-1)S \leq ON_i^P < tS\} \\ \mathcal{D}_t &= \{i \in \mathcal{F} \mid (t-1)S \leq OFF_i^P < tS\}\end{aligned}$$

The estimation of arrival delays (due to local runway congestion) requires the elimination of upstream delays incurred by the aircraft. For instance, if an aircraft incurs a departure delay at its origin airport (resulting from delays in previous flight legs, congestion at the departure airport or any other reason), then this departure delay must be removed from the estimation of the arrival delay of the flight. However, this is complicated by the practice of schedule padding by the airlines. Indeed, airlines might plan airborne times that are *longer* than the unimpeded airborne time to improve their on-time performance [109]. We therefore correct our estimates of arrival delays accordingly. For each route terminating at the considered airport, we define the unimpeded airborne time, denoted by AIR^{MIN} , as the minimum airborne time that can be flown on the route and we measure it as the lowest 10th percentile of all actual airborne times on the considered route. We then remove any surplus in the scheduled airborne time, i.e.: $(AIR^P - AIR^{MIN})^+$. In turn, we obtain the average arrival delay during period t , denoted by $\widehat{\delta}_t^A$, as follows:

$$\begin{aligned}\widehat{\delta}_t^A &= \frac{1}{|\mathcal{A}_t|} \sum_{i \in \mathcal{A}_t} \left((IN_i^A - IN_i^P)^+ - \left[(OFF_i^A - OFF_i^P)^+ - (AIR_i^P - AIR_i^{MIN})^+ \right]^+ \right)^+ \\ \widehat{\delta}_t^A &= \frac{1}{|\mathcal{A}_t|} \sum_{i \in \mathcal{A}_t} \left(\delta_i^{IN} - \left[\delta_i^{OFF} - (AIR_i^P - AIR_i^{MIN})^+ \right]^+ \right)^+\end{aligned}$$

Finally, we estimate the average departure delay during period t , denoted by $\widehat{\delta}_t^D$, as the

average taxi-out delay of all aircraft scheduled to take off during period t . In other words, we remove the gate delay from the recorded wheels-off delay, as most gate delays are not due to runway congestion (e.g., propagation from upstream disturbances, mechanical failures, delays in passenger buildings).

$$\widehat{\delta}_t^D = \frac{1}{|\mathcal{D}_t|} \sum_{i \in \mathcal{D}_t} \left((OFF_i^A - OFF_i^P)^+ - (OUT_i^A - OUT_i^P)^+ \right)^+$$

$$\widehat{\delta}_t^D = \frac{1}{|\mathcal{D}_t|} \sum_{i \in \mathcal{D}_t} (\delta_i^{OFF} - \delta_i^{OUT})^+$$

As previously mentioned, these estimates of departure queue lengths and arrival and departure delays are only approximate. Although the procedures outlined above do eliminate to a large extent delays that cannot be attributed to local runway congestion, they are still imperfect. On the arrival side, our estimation of arrival delays does not account for the possibility of en-route congestion. Even though most of the delays occur at the airports, this might create some measurement errors. More broadly, we do not consider Air Traffic Flow Management procedures in these macroscopic estimates. On the departure side, our restriction to taxi-out delays does not allow us to consider potential gate delays that are due to local runway congestion (e.g., if gate pushbacks are slowed as a result of ground congestion). Conversely, some of the taxi-out delays that we attribute to local runway congestion may in truth be due to other causes (e.g., safety incidents or pilot decisions). Note, finally, that delay estimates are generally less robust than departure queue length estimates, as a few heavily delayed flights might drive average delays up significantly.

4.2.2 Model of Departure Queue Lengths

In this section, we compare the departure queue lengths predicted by the model d_t to those observed in practice \widehat{d}_t at JFK, EWR and LGA in Summer 2007. Figure 4-3 shows the expected departure queue length for each of the 72 periods of the day as predicted by the model and the average departure queue length observed at the airports, estimated according to the procedures outlined in Section 4.2.1. Note, first and foremost, that the model evaluates

quite accurately both the dynamics of departure queues over the course of the day and their magnitude at peak hours. At JFK and EWR, departure queues tend to form in the morning (between 7 a.m. and 10 a.m.), dissipate around noon and return again in the afternoon (from 4 pm onward). In contrast, LGA seems to be congested almost continuously from 8 a.m. to 9 p.m., which is due to its almost “flat” daily demand profile. These patterns are predicted well by our model. Moreover, the model predicts peak-hour queues of a similar magnitude to those observed in practice. Specifically, modeled queue lengths are within 2-3 aircraft from those observed at the airports at peak morning and afternoon hours, which corresponds to a relative difference within 20% (see Table 4.3). Given the strategic nature of the model and the uncertainty associated with our queue length estimates, this level of accuracy is adequate.

The most significant errors of the model involve underestimating average departure queue lengths in the late afternoon and, to a lesser extent, in the late morning at JFK. We believe that this stems from the fact that a larger share of flights get delayed over the course of the day, resulting in later shifts in demand. As a result, many aircraft demand the usage of the runway system later than originally planned. In turn, the airports seem to recover from congestion less quickly than predicted by the model. Note that these effects are observed at all three airports, even though the underestimation of late-afternoon departure queues seems to be the highest at EWR.

Table 4.3: Statistics on modeled and actual departure queue lengths

Airport	Peak morning queue			Peak afternoon queue		
	Model	Actual	Error	Model	Actual	Error
JFK	17.7	15.3	+15.7%	26.6	25.2	+5.6 %
EWR	21.8	20.4	+6.7%	16.4	19.5	-16.1%
LGA	13.2	13.0	+1.2%	14.3	13.6	+5.3%

In addition to providing estimates of the expected queue lengths, our model also provides estimates of the variability of the queue lengths over the course of the day. This topic has attracted only limited attention to date despite the fact that delay variability is a very

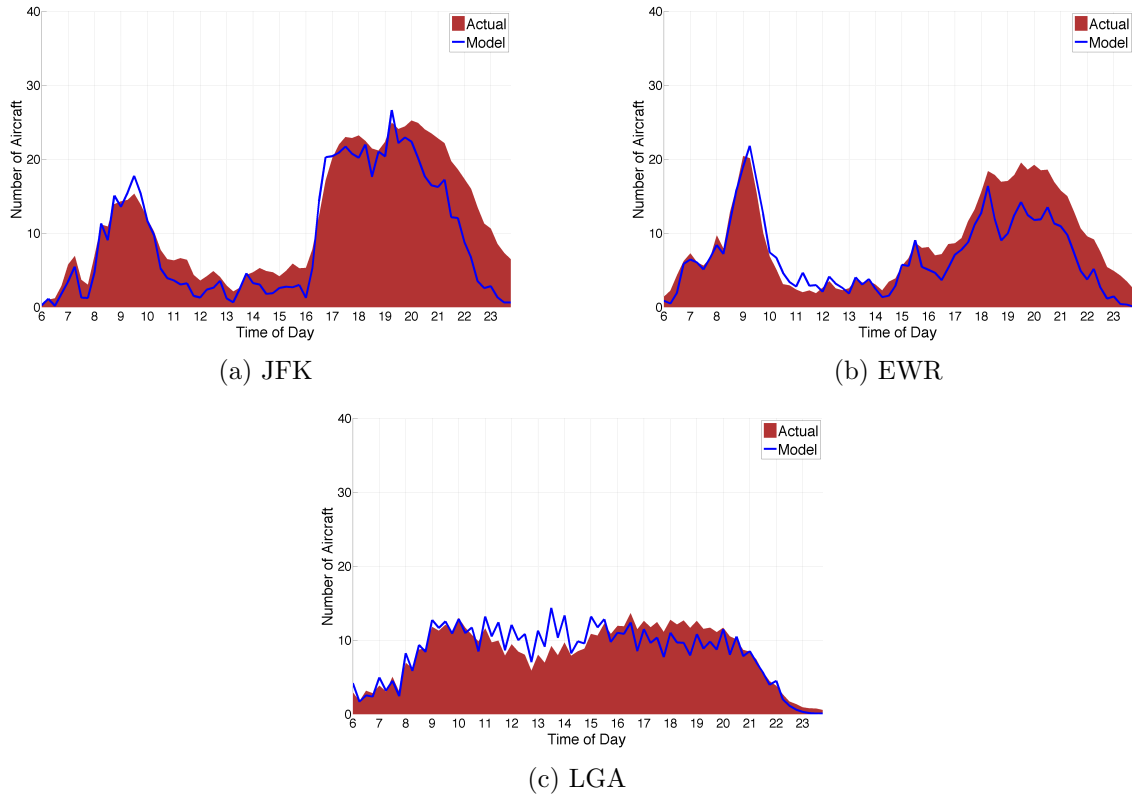


Figure 4-3: Average departure queue lengths in Summer 2007

important measure of reliability and performance. Indeed, the system-wide impacts of flight delays depend, to no small extent, on their variations from day to day. If delays are relatively similar from day to day, then it might be possible for the airlines and airport operators to anticipate them and limit their impacts. If, on the other hand, delays are extremely variable from day to day, then the knowledge of expected delays is only of limited value for planning schedules and operations.

Figure 4-4 compares the range of departure queue lengths observed in practice to the model's predictions. It shows the scatter plot of the departure queue lengths \hat{d}_t observed in Summer 2007 at the airports as a function of the time of the day, as well as the 5th and 95th percentiles of the distribution of departure queue lengths d_t obtained from our model. The comparison shows that at all three airports considered, the model estimates well the observed range of delays. First, observed departure queue lengths lie outside the predicted

range only on a very limited number of days. Second, the bounds defined by the 5th and 95th percentiles of the distribution of d_t are relatively close to the smallest and largest departure queue lengths observed in practice. Finally, note that the variability of delays is the largest at times when the average delays are also the largest. This suggests that runway congestion not only creates large delays on average, but also increases their variability, thus making these delays even more onerous to airlines and their passengers.

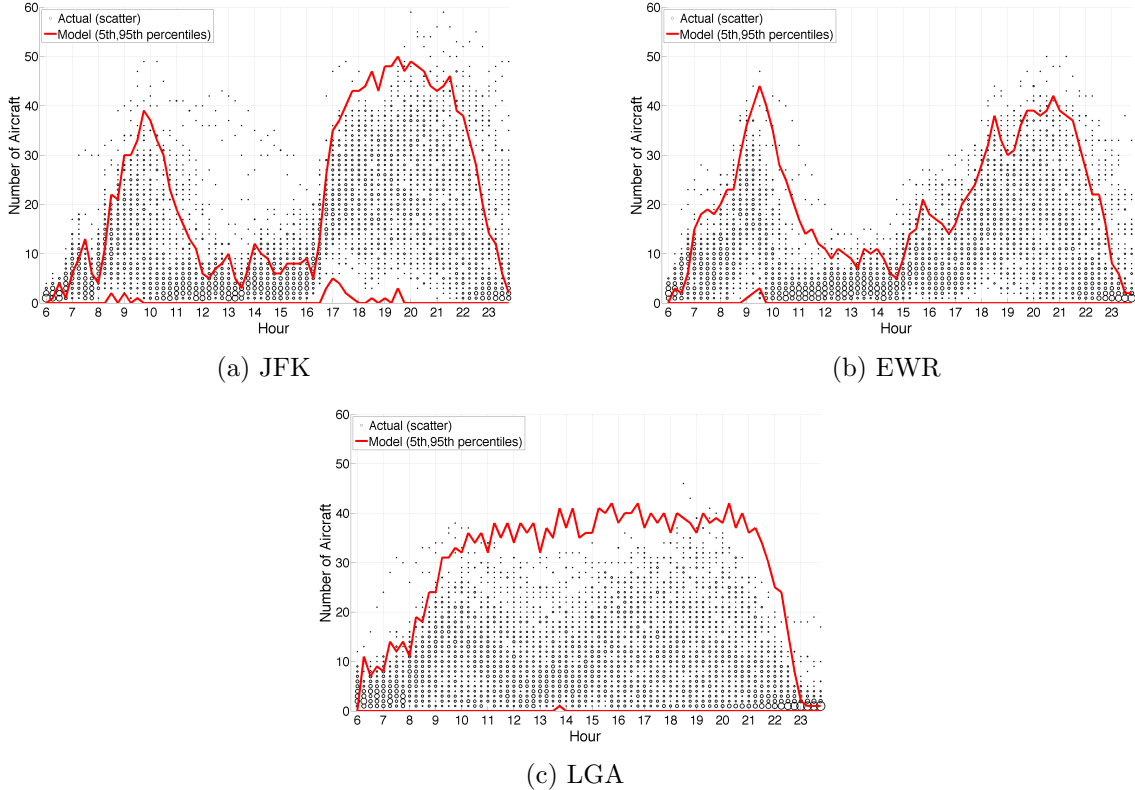


Figure 4-4: Variability of departure queue lengths in Summer 2007

4.2.3 Model of Arrival and Departure Delays

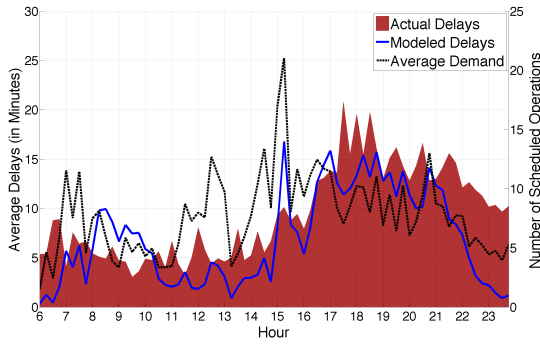
In this section, we compare the expected arrival and departure delays predicted by our model to the average delays $\widehat{\delta}_t^A$ and $\widehat{\delta}_t^D$ experienced at the airports. The purpose of this comparison is twofold. First, we aim to confirm our findings from the previous section to increase our confidence in our model’s ability to estimate departure delays at the airports.

Second, we investigate our model’s ability to estimate arrival delays. Whereas no estimate of arrival queue lengths could be obtained from available data, we developed estimates of arrival delays in Section 4.2.1 that enable such comparison. Note, however, that these comparisons are subject to more uncertainty than those from the previous section, as delay estimates are more noisy than queue length estimates.

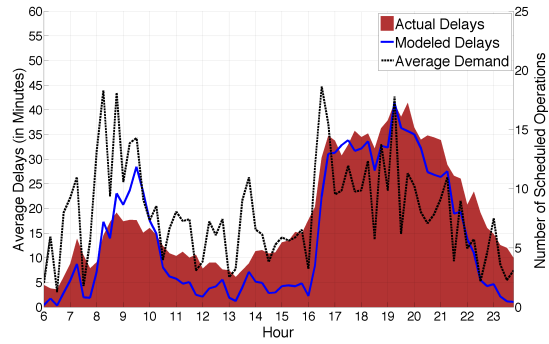
Figure 4-5 compares the expected value of the delays predicted by the model to the average delays observed in practice at JFK (Figures 4-5a and 4-5b), EWR (Figures 4-5c and 4-5d) and LGA (Figures 4-5e and 4-5f) in Summer 2007. Expected modeled delays are obtained by dividing expected queue lengths by the average service rates. We also plot the average number of flights scheduled. These results suggest that the model approximates very well the timing, dynamics and magnitude of the delays at the three airports and confirm the insights obtained in the previous section. First, the model predicts well the evolution and fluctuations of delays over the course of the day. Second, the most significant errors, percent-wise, occur at the least congested periods, such as late evening hours, when traffic is low. Third, the model approximates quite accurately the magnitude of flight delays, especially at peak hours. Table 4.4 reports the expected arrival and departure delays predicted by the model (a) over an entire day of operations and (b) during the peak delay period between 5 p.m. and 9 p.m., and the corresponding average delays observed in practice. The error of our model lies within 20% at JFK, within 30% at EWR and within 15% at LGA. With the exception of peak departure delays at EWR, this corresponds to an error in average delays of only 2-3 minutes.

Table 4.4: Statistics on modeled and actual delays

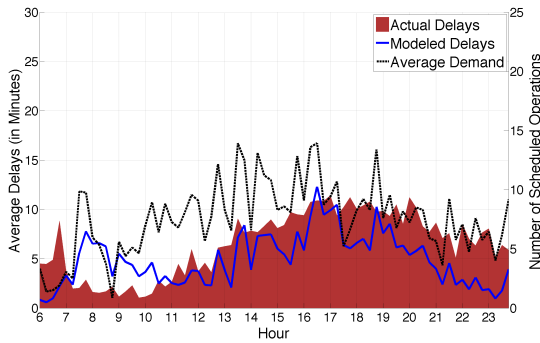
Airport	Type	Average delays			Average peak delays		
		Model	Actual	Error	Model	Actual	Error
JFK	Arrivals	8.1 min	9.9 min	-18.2%	13.2 min	15.6 min	-15.2%
	Departures	18.3 min	21.6 min	-15.0%	33.3 min	36.0 min	-7.2%
EWR	Arrivals	5.7 min	7.1 min	-19.8%	7.4 min	9.9 min	-25.5%
	Departures	13.1 min	16.1 min	-18.8%	18.8 min	26.8 min	-29.7%
LGA	Arrivals	6.9 min	6.4 min	+7.3%	7.1 min	8.4 min	-15.7%
	Departures	17.1 min	16.2 min	+5.5%	19.0 min	20.9 min	-9.1%



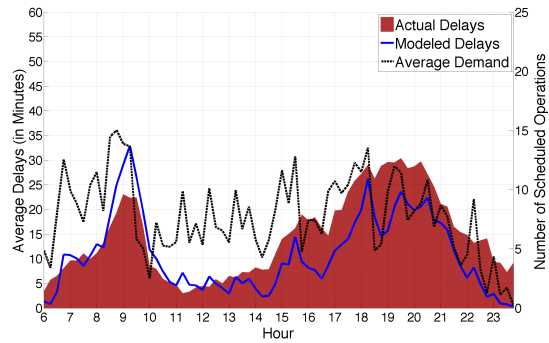
(a) JFK – Arrival Delay



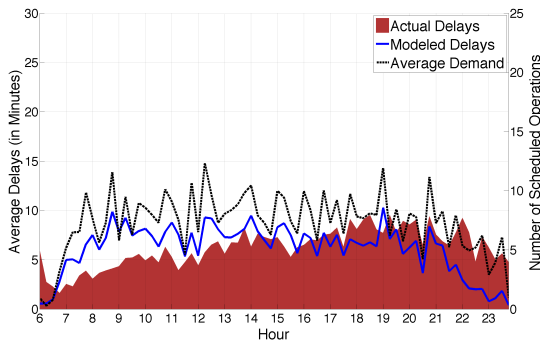
(b) JFK – Departure Delay



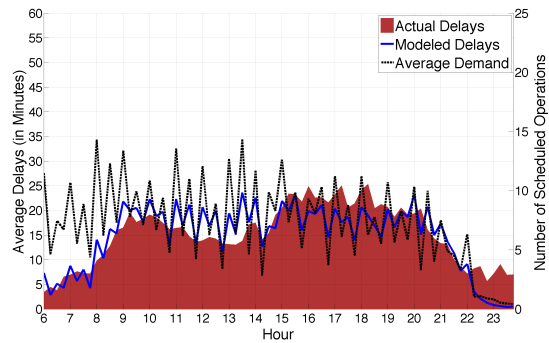
(c) EWR – Arrival Delay



(d) EWR – Departure Delay



(e) LGA – Arrival Delay



(f) LGA – Departure Delay

Figure 4-5: Predicted and actual delays at JFK, EWR and LGA in Summer 2007

As in the case of the departure queue lengths, the model also underestimates delays at off-peak hours. This results in modeled average arrival and departure delays that are slightly smaller than the average delays observed in practice, in most cases (Table 4.4). These model errors may stem from several factors, which represent important opportunities for future research. First, our estimates of actual delays are only approximate (see Section 4.2.1).

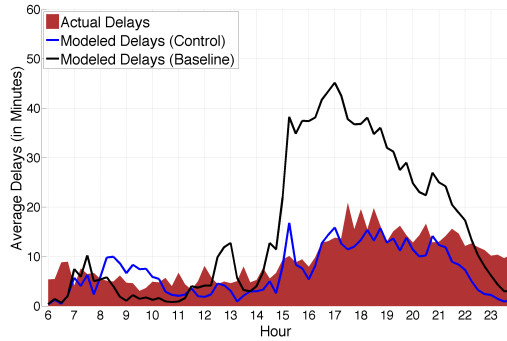
Specifically, our estimation procedures may not fully eliminate external delays that are not due to local runway congestion and are thus not accounted for in our model (e.g., inefficiencies in airline operations, propagation of delays incurred in previous flight legs, etc.). Off-peak delay estimates are the most sensitive to the effects of these external delays, as averages are computed for a much smaller sample of flights. Second (as discussed in Section 4.2.2), the exclusive reliance of our model on the *a priori* schedule of flights ignores potential disturbances to that schedule and schedule updates that might occur over the day. Such disturbances could result in the shifting of some flights to later times, and thus in continued congestion after peak hours.

Nonetheless, our model estimates the relative magnitude of arrival and departure delays accurately. Indeed, it predicts arrival delays of about half the magnitude of departure delays at JFK, EWR and LGA, that are observed in practice. (Please note the different scales of the y-axes in Figures 4-5a, 4-5c and 4-5e, on the one hand, and Figures 4-5b, 4-5d and 4-5f, on the other.) This is an important result that stems from the integration of our control of arrival and departure service rates into our queuing model of airport congestion. We further investigate the benefits of our integrated approach in the next section.

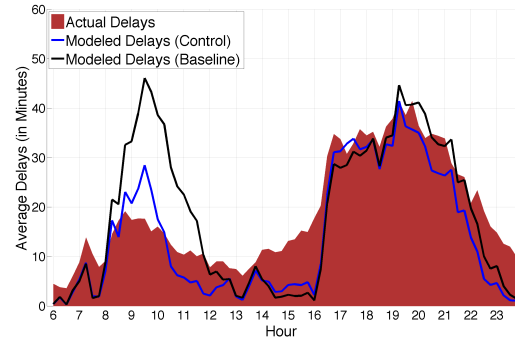
4.2.4 Benefits of the Integrated Approach

We now compare our integrated queuing model of airport congestion with endogenous control of arrival and departure service rates to a baseline model where no control is exercised and the service rates are kept constant throughout the day of operations (see Figure 4-1). The baseline model considers equal values for the arrival service rate and the departure service rate, equal to the capacity of the airport when it serves arrivals and departures in equal numbers. This is the modeling assumption that is most widely used in macroscopic models of airport congestion. Specifically, we set the values of the arrival and departure service rates equal to 10 movements per 15 minutes at JFK and EWR and to 9 movements per 15 minutes at LGA [104]. Figure 4-6 shows the expected arrival and departure delays resulting from the application of the two models as well as the average delays incurred in practice at JFK

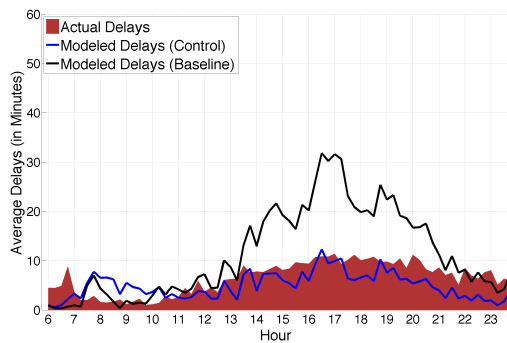
(Figures 4-6a and 4-6b), EWR (Figures 4-6c and 4-6d) and LGA (Figures 4-6e and 4-6f).



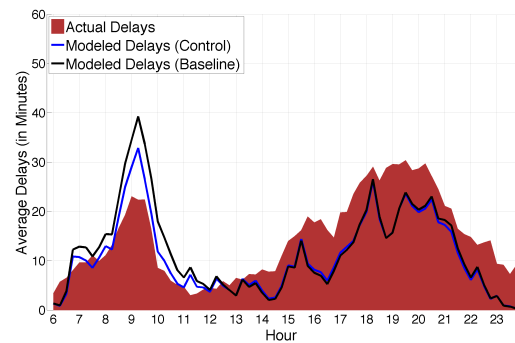
(a) JFK – Arrival Delay



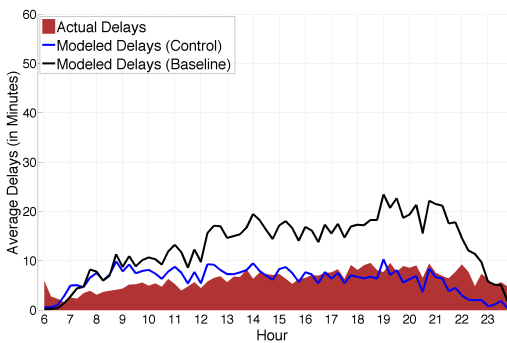
(b) JFK – Departure Delay



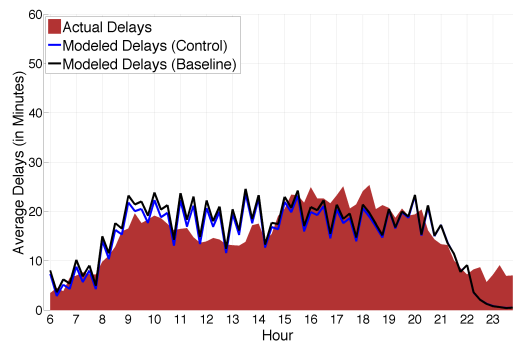
(c) EWR – Arrival Delay



(d) EWR – Departure Delay



(e) LGA – Arrival Delay



(f) LGA – Departure Delay

Figure 4-6: Modeled delays with and without the control of service rates in Summer 2007

Note that the control of service rates has a significant effect on the dynamics and magnitude of arrival and departure delays. The baseline model greatly overestimates arrival delays in the afternoon at all three airports, by an estimated 98% at JFK, 94% at EWR and 124% at LGA. The corresponding (absolute) error is smaller than 20% at all three airports with

our model (see Table 4.4). The baseline model also overestimates departure delays in the morning at JFK and, to a lesser extent, at EWR. The effects of the control of service rates on departure delays are smaller at LGA because the departure throughput at LGA is less sensitive to the arrival throughput (see Figure 2-5) and the balance of scheduled arrivals and departures is more steady over the day.

The results shown in Figure 4-6 stem to a large extent from the fact that JFK and EWR face a departure peak in the morning and an arrival peak in early afternoon. In turn, most of available capacity at these airports is used to operate departures in the morning and arrivals in early afternoon (around 3 pm). The baseline model, which considers fixed values of arrival and departure service rates, fails to capture how service rates are varied with the schedules of flights. Note, also, that the baseline model cannot be fixed by simply changing the estimates of the arrival and departure service rates. Indeed, departure delays are overestimated in the morning but slightly underestimated in the afternoon. In turn, improving the model's fit in the morning would require increasing the departure rate. But this would also lead to a decrease in estimated departure delays in the afternoon, and thus to a significant deterioration of the predictive ability of the model. In contrast, our control of arrival and departure service rates replicates the dynamic allocation of airport capacity to operate landings and takeoffs efficiently, for any given schedule of flights. In turn, the integration of this tactical control improves significantly the predictive power of our macroscopic queuing model of airport congestion at the strategic level.

4.3 Scheduling and On-time Performance Trends

Airlines' schedules at JFK, EWR and LGA have undergone important changes since 2007. Figure 4-7 shows the average schedules at JFK (Figure 4-7a), EWR (Figure 4-7b) and LGA (Figure 4-7c) in Summer 2007, 2008 and 2010. Note, first, that the schedules of flights at JFK were smoother in 2008 than in 2007. Approximately the same total number of flights was scheduled at JFK in Summer 2008 as in Summer 2007 but peak-hour scheduling levels were significantly lower in 2008 than in 2007. This is mostly due to the introduction of

schedule limits in May 2008 at JFK and EWR. In contrast, flight caps were already in place at LGA in 2007, and hourly schedules at EWR were less peaked in 2007 than at JFK, so smaller changes were experienced between 2007 and 2008 at these two airports. Second, the three airports experienced a small decline in demand between 2008 and 2010, estimated at 8.6% at JFK, 6.6% at EWR and 3.3% at LGA. This is mostly due to the economic downturn and, perhaps, to a recent profitability-based trend of “capacity discipline” in the US airline industry [121].

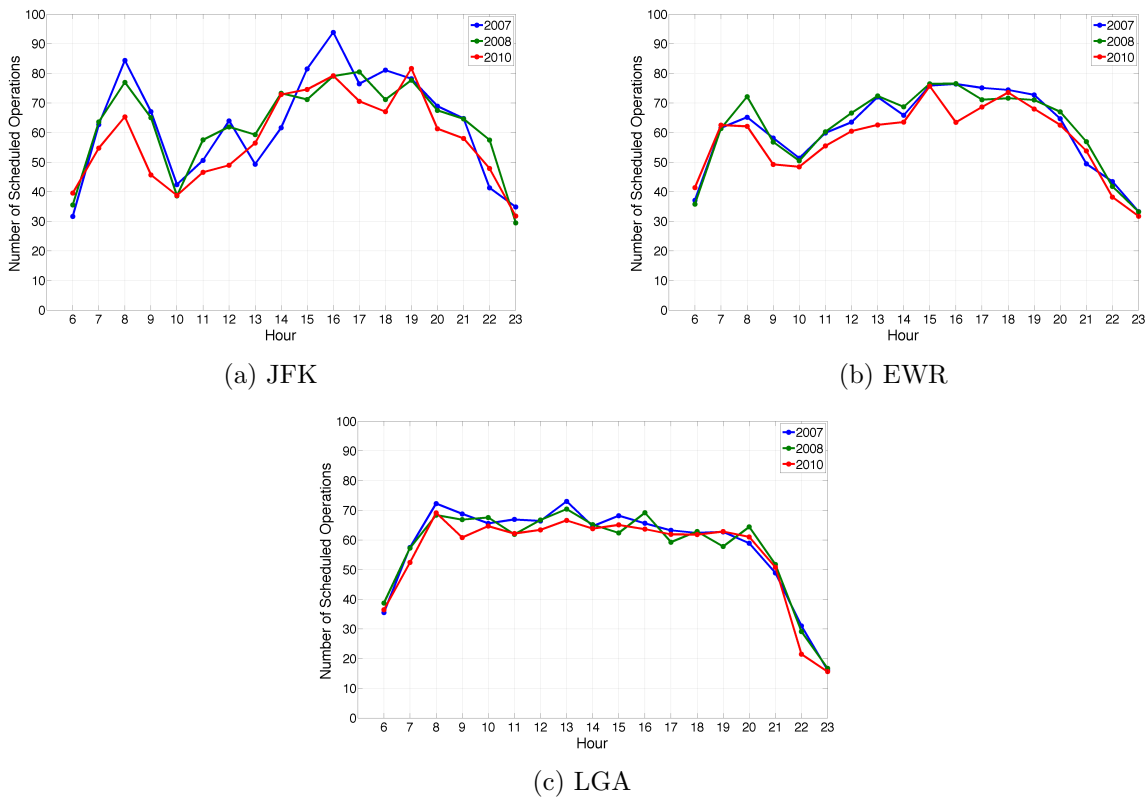


Figure 4-7: Average number of flights scheduled per hour in Summer 2007, 2008 and 2010

At the same time, substantial delay reductions have been observed in practice since 2007. At the three airports, the local departure delays were 25% to 50% lower in Summer 2011 than in Summer 2007. In order to understand the extent to which this large decrease in delays can be attributed to the comparatively small changes in scheduling described in Figure 4-7, we apply our model of airport congestion with the schedules of flights from the 5 Summer

periods between 2007 and 2011. The other input parameters, including the runway system capacity estimates (Figure 2-5), are left unchanged. Any change in modeled delays is thus due to changes in flight schedules. We then compare the evolution in average arrival and departure delays obtained with the model to that observed in practice. If the model predicts an evolution of delays similar to the one observed in practice, then the delay decrease can be attributed primarily to the changes in demand during the considered period. If, on the other hand, the model predicts a delay reduction that is significantly smaller than the one observed in practice, then the delay reduction observed in practice is likely to be due to other factors than the changes in demand, including an improvement of airport traffic-handling performance. The results are summarized in Figure 4-8 and Table 4.5.

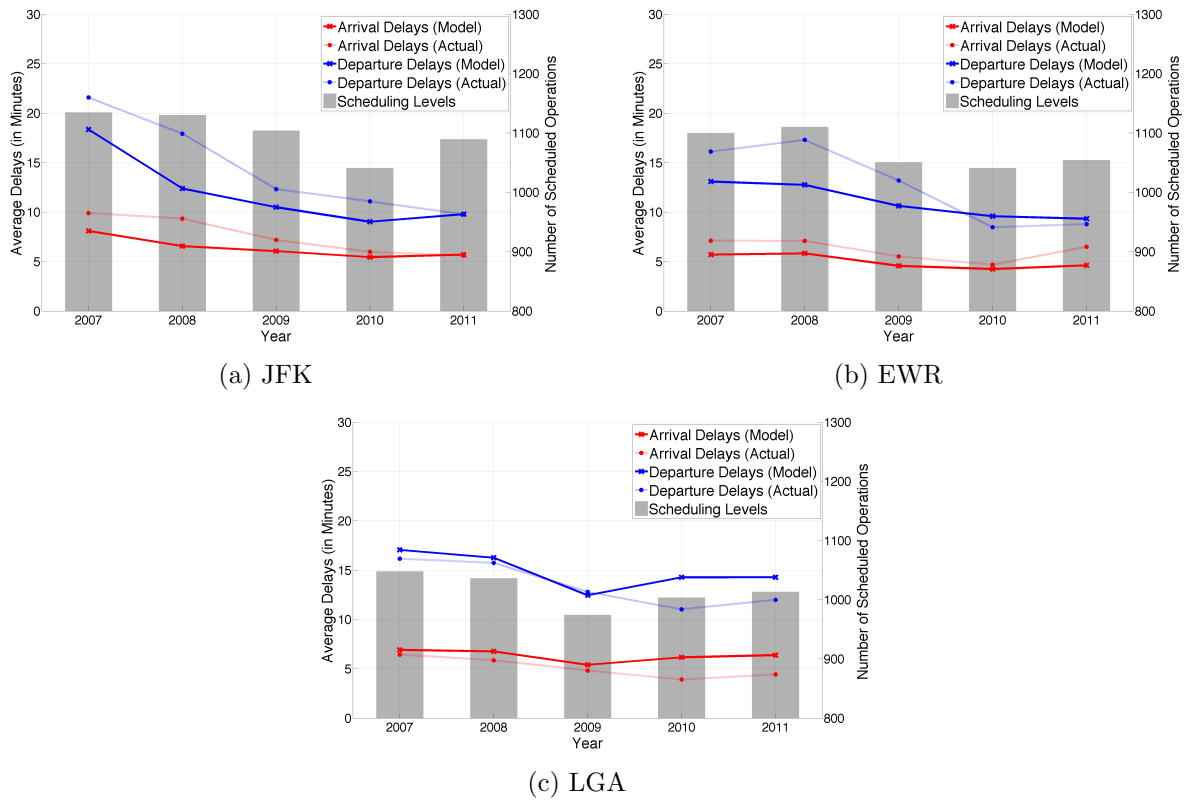


Figure 4-8: Predicted and actual average delays from Summer 2007 to Summer 2011

The main observations are twofold. First, the model does predict large delay reductions between 2007 and 2011 at the three airports that are, in most cases, of an order similar to

Table 4.5: Percent of change in demand and delays between 2007 and 2011

Airport	Year	Demand	Arrival delays		Departure delays	
			Model	Actual	Model	Actual
JFK	2007	Baseline	Baseline	Baseline	Baseline	Baseline
	2008	-0.4%	-19.0%	-5.7%	-32.5%	-17.0%
	2009	-2.7%	-25.2%	-27.4%	-42.8%	-42.9%
	2010	-8.3%	-32.8%	-39.6%	-50.8%	-48.7%
	2011	-4.0%	-29.5%	-43.2%	-46.6%	-54.7%
EWR	2007	Baseline	Baseline	Baseline	Baseline	Baseline
	2008	+0.9%	+2.1%	-0.1%	-2.6%	+7.3%
	2009	-4.5%	-20.0%	-22.5%	-18.9%	-18.2%
	2010	-5.3%	-25.7%	-34.1%	-26.8%	-47.5%
	2011	-4.1%	-19.0%	-8.5%	-28.7%	-45.5%
LGA	2007	Baseline	Baseline	Baseline	Baseline	Baseline
	2008	-1.1%	-2.1%	-9.2%	-4.7%	-2.6%
	2009	-7.0%	-21.9%	-25.2%	-27.0%	-20.8%
	2010	-4.2%	-10.9%	-39.3%	-16.3%	-31.8%
	2011	-3.3%	-7.7%	-31.2%	-16.3%	-25.9%

those observed in practice. Second, the delay reductions observed in practice tend to be larger than those predicted by the model, especially at EWR and LGA.

Note, first, that the model predicts reductions in arrival and departure delays of 30% to 50% at JFK, of 25% at EWR and of 10% to 15% at LGA between Summer 2007 and Summer 2010, when the demand was the lowest. This suggests that the small decline in demand observed during the period might explain very large declines in average delays. Queuing theory has established that delays vary non-linearly with demand when an airport operates close to capacity in steady-state conditions [34]. Our results suggest that the nonlinear relationship between scheduling levels and delays is also very much valid for the stochastic and dynamic queuing model with an endogenous control of service rates considered in this chapter. In turn, airport on-time performance is expected to be extremely sensitive to even small changes in demand.

It is noteworthy that departure delays declined significantly at JFK between Summer 2007 and Summer 2008, according to both the model and actual data. As previously discussed, approximately as many flights were operated during these two periods (see also Ta-

ble 4.5). The improvements in on-time performance were therefore primarily due to changes in the *distribution* of the flights over the day. By comparison, delays were of similar magnitude in Summer 2007 and Summer 2008 at EWR and LGA, where smaller scheduling changes were observed over this time period. These results indicate that on-time performance is sensitive not only to the total number of flights scheduled, but also to their distribution over the day. Specifically, for any given number of flights scheduled in a day, the more evenly they are distributed over its course, the smaller the average delays will be.

Finally, the actual delay reductions between 2007 and 2011 have been larger than the model's predictions at the three airports. For instance, the model slightly underestimates arrival and departure delays in Summer 2007 at EWR but slightly overestimates them in Summer 2011. In particular, significant delay reductions have been observed at EWR between Summer 2009 and Summer 2010, while flight schedules were similar. As well, on-time performance seems to have improved at LGA between 2009 and 2010, even though demand was larger in 2010. This may suggest that the efficiency of operations has improved over the time period considered. Examples of such improvements might include increases in airport capacity resulting from improved air traffic management procedures, improved surface congestion management, etc. This is a topic that certainly deserves further research.

4.4 Conclusion

We presented an original approach to airport congestion modeling that integrates an endogenous control of arrival and departure service rates into a stochastic and dynamic queuing model. The control simulates the tactical utilization of available capacity and provides a systematic means of selecting arrival and departure service rates in any queuing model of airport congestion, as a function of the schedule of flights and estimates of airport capacity. The resulting integrated model of airport congestion provides a fast and flexible tool to quantify on-time performance at the strategic level at different airports under different demand and capacity scenarios, while accounting for the impact of flight schedules and airport capacity on the utilization of airport capacity.

We implemented the model at JFK, EWR and LGA. We developed empirical estimates of airport on-time performance and compared our model’s outputs to historical records of operations. We showed, first, that the model predicts accurately the expected value and the variability of departure queue lengths over the day at the three airports. Second, we showed that the model estimates well the timing, dynamics and magnitude of expected arrival and departure delays, although off-peak delays are slightly underestimated. In particular, the model achieves a significantly better predictive performance than a baseline model that does not consider the endogenous control of service rates. Third, the application of the model has indicated that flight delays are very sensitive to the number of daily flights and to the distribution of flights over the day, when an airport operates close to capacity. Results suggest that the delay reductions that have been observed at JFK, EWR and LGA over the past few years can be primarily attributed to the comparatively small changes in scheduling and, to a lesser extent, to improvements in air traffic handling procedures.

The sensitivity of airport on-time performance to the schedule of flights motivates the investigation of the potential of scheduling adjustments to mitigate air traffic congestion. In the remainder of this thesis, we integrate the model of airport congestion developed in this chapter into models of airport scheduling interventions.

Chapter 5

Integrated Capacity Utilization and Scheduling

This chapter develops our integrated approach to airport congestion mitigation. It is motivated by results shown in Chapter 3 and 4 that underscore the interdependencies between the schedule of flights and airport capacity utilization procedures. This approach jointly optimizes scheduling interventions at the *strategic* level (i.e., months in advance) and the utilization of airport capacity at the *tactical* level (i.e., over each day of operations), subject to scheduling, capacity, and delay reduction constraints. This is accomplished through the development of an *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that combines three different models. First, the Stochastic Queuing Model of airport congestion quantifies flight delays as a function of flight schedules and arrival and departure service rates. Second, the Dynamic Programming model of capacity utilization optimizes the sequential control of runway configurations and of arrival and departure service rates to minimize congestion costs, for any schedule of flights. Third, a model of scheduling interventions uses a schedule of flights determined by airline competitive scheduling as its starting point and modifies that schedule in a way that (i) reduces the imbalances between airport demand and capacity; (ii) preserves the connectivity of aircraft and passenger itineraries; and (iii) minimizes the changes to flight schedules. It is formulated as an Integer Programming model.

The contributions of this chapter fall into four categories:

- We develop an *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that integrates the Stochastic Queuing Model of airport congestion and the Dynamic Programming model of capacity utilization into an Integer Programming model of scheduling interventions.
- We introduce an iterative solution algorithm that solves simultaneously the schedule optimization problem and the capacity utilization problem.
- We show from computational results at JFK Airport that substantial delay reductions can be achieved through small scheduling changes.
- We show that the integrated approach developed in this chapter provides significant benefits, as compared to a sequential approach under which flight schedules and operating procedures are optimized consecutively.

Papers based on this chapter can be found in [60, 61].

In the remainder of this paper, we first develop the ICUSM in Section 5.1 and our iterative solution algorithm in Section 5.2. We then present computational results at JFK in Section 5.3 and quantify the benefits of our integrated approach in Section 5.4. Section 5.5 concludes.

5.1 Integrated Capacity Utilization and Scheduling Model

5.1.1 Model Presentation

As detailed in Chapter 2, the ICUSM takes as inputs: (i) estimates of airport capacity (Operational Throughput Envelopes); and (ii) the original schedule of flights created by the airlines for any given day of operations in the absence of demand management, as well as planned aircraft and passenger connections (see Figure 2-1a). In this chapter, we assume

that the connections data are either provided by the airlines or determined, in the statistical sense, using historical information regarding aircraft and passenger itineraries. The ICUSM optimizes jointly: (i) scheduling interventions and (ii) capacity utilization policies (see Figure 2-1b).

A schematic representation of the model’s formulation is presented below. The ICUSM aims to find a schedule of flights that minimizes the *displacement* (defined in Section 5.1.2) from the original schedule, subject to scheduling, network connectivity, capacity, and queue length reduction constraints. Scheduling and network connectivity constraints ensure that no flight is eliminated, and that scheduled block times, aircraft connections and passenger itineraries are left unchanged. In combination with the objective function (“minimize schedule displacement”), they force the model to search for scheduling interventions that minimize interference with airline competitive scheduling. Capacity constraints ensure that arrivals and departures are serviced at rates defined by the Operational Throughput Envelope corresponding to the runway configuration in use and to observed weather conditions. Finally, queue length reduction constraints ensure that, at the end of each 15-minute period, the peak expected arrival and departure queue lengths do not exceed prespecified limits, denoted by A_{MAX} and D_{MAX} , respectively.

minimize Schedule Displacement
 subject to Scheduling and network connectivity constraints
 Operating capacity constraints
 Peak expected arrival queue length lower than A_{MAX}
 Peak expected departure queue length lower than D_{MAX}

The main novelties in the model lie in the form of the capacity constraints and the queue length reduction constraints. First, airport capacity is characterized as the *expected* number of movements that can be operated per unit of time, thus enabling the incorporation of a stochastic model of airport congestion into the scheduling model. Second, instead of

applying predetermined schedule limits that are externally given and kept roughly constant throughout the day, the capacity constraints applies constraints on the number of flights *operated* based on engineering limitations, and the queue length reduction constraints integrate on-time performance targets *into* the scheduling model.

The resulting problem cannot be formulated as a single Mixed Integer Program, because (a) the probabilistic evolution of arrival and departure queues depends *endogenously* on the schedule of flights, and (b) the stochastic relationship between flight schedules and delays is *nonlinear*. In other words, changes in flight schedules, i.e., in the decision variables, induce nonlinear changes in the probabilistic dynamics of arrival and departure queue lengths.

The ICUSM therefore relies on an original solution architecture to formulate and solve the problem described above. Before describing it in detail, we introduce the following notations. We divide a day of operations into periods of 15 minutes each. We denote by A_t and D_t the random variables describing arrival and departure queue lengths at the end of period t , respectively. Observed arrival and departure queue lengths are realizations of these random variables and are denoted by a_t and d_t , respectively (consistently with the notations in Chapter 3). Scheduling decisions aim to reduce *expected* queue lengths, i.e., $E(A_t)$ and $E(D_t)$, while capacity utilization policies are based on *observed* queue lengths a_t and d_t . We also denote by λ_t^X and λ_t^Y the aggregate schedule of flights, i.e., the number of scheduled arrivals and departures per period t across all airlines.

The ICUSM consists of three main modules, shown in Figure 5-1:

- An Integer Programming model of scheduling interventions: It takes as input the original schedule of flights and connections data. For any value of the schedule displacement, it generates feasible schedules of flights, subject to the scheduling and network connectivity constraints. This, in turn, determines the aggregate schedule (i.e., λ_t^X and λ_t^Y). We use a formulation similar to the one from Pyrgiotis and Odoni [96].
- The Dynamic Programming model of capacity utilization (Chapters 3 and 4): It takes as inputs the modified schedule of flights and estimates of airport capacity, and it optimizes capacity utilization policies to minimize congestion costs, subject to the

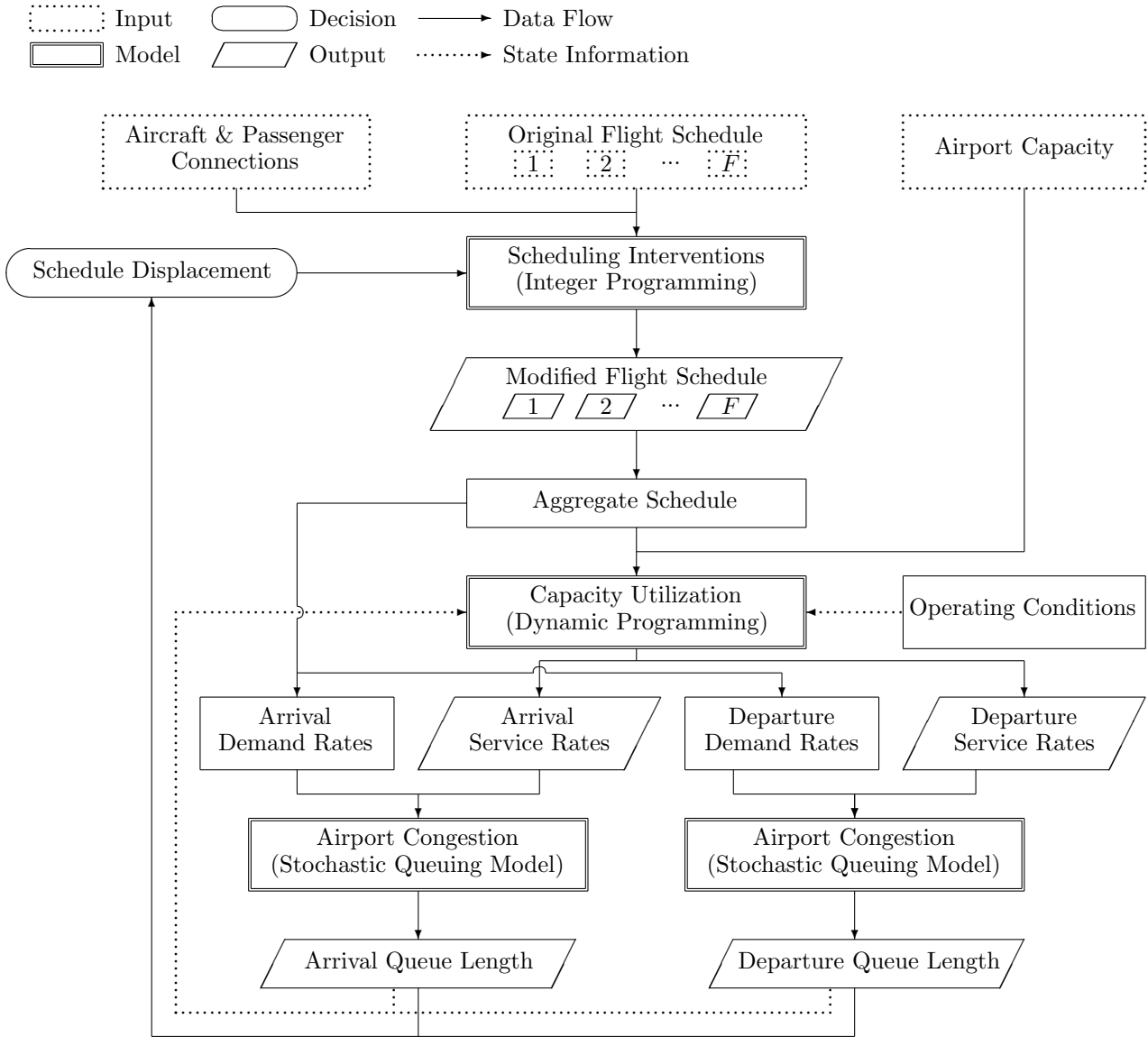


Figure 5-1: Modeling Architecture

capacity constraints. It returns the selected runway configuration and service rates for each period, as a function of a state variable (dashed lines) that includes observed queue lengths (i.e., a_t and d_t) and operating conditions (i.e., weather and winds).

- The Stochastic Queuing Model of airport congestion: It takes as inputs the demand rates (determined by the modified schedule of flights) and the service rates (determined by capacity utilization policies) for the arrival and departure queuing systems, and

returns the probabilistic evolution of arrival and departure queue lengths over the day (i.e., A_t and D_t).

In summary, the model of scheduling interventions modifies the schedule of flights, and, in combination, the model of capacity utilization and the model of airport congestion return the probability distribution of the arrival and departure queue lengths that minimize congestion costs under the proposed schedule. The comparison of the expected queue lengths with the queue length limits (i.e., whether or not $E(A_t) \leq A_{\text{MAX}}$ and $E(D_t) \leq D_{\text{MAX}}$ for each period t) then informs on whether the considered schedule displacement is too large, or too small. Iterating between these three modules determines, in turn, the optimal schedule displacement.

We formulate the ICUSM in the remainder of this section, and we describe in Section 5.2 our solution algorithm that integrates the different modules and solves the ICUSM.

5.1.2 Model of Scheduling Interventions

We formulate in this section the framework for scheduling interventions as an Integer Programming model. It is similar to the one in [96], but a new formulation is introduced here.

We denote by K the airport where the scheduling interventions are applied. We consider in the model all the flights that arrive at K or that leave from K , as well as all flights that are flown by an aircraft that visits K during the day. This results in flights that do not operate at K being included in the model. For instance, if an aircraft operates the itinerary $K \rightarrow L \rightarrow M \rightarrow K$, then the flight $L \rightarrow M$ is included in the model. (Note that each “flight leg” in an aircraft’s itinerary is defined as a “flight”.) This is because the rescheduling of flight $K \rightarrow L$ or $M \rightarrow K$ might require a change in the scheduled time of flight leg $L \rightarrow M$ to maintain feasible connections. The total number of flights considered is denoted by F . We include in the model all 15-minute periods from the earliest departure time to the latest arrival time of the F flights. We denote by T the corresponding number of periods.

Sets

\mathcal{T} = the set of time periods, $\{1, \dots, T\}$

\mathcal{F} = the set of flights, $\{1, \dots, F\}$

$\mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}$ = the set of flights $i \in \mathcal{F}$ scheduled to land/take off at airport K

$\mathcal{C} \subset \mathcal{F} \times \mathcal{F}$ = the set of flight pairs (i, j) such that there is a connection between i and j

The set \mathcal{C} includes all pairs of flights (i, j) between which there is an aircraft connection (i.e., flights i and j are flown by the same aircraft and j is the immediate successor of i) or a passenger connection (i.e., there is at least one passenger connecting from flight i to flight j). The mathematical treatment of aircraft and passenger connections is similar.

Parameters

$$S_{it}^{\text{arr}}/S_{it}^{\text{dep}} = \begin{cases} 1 & \text{if flight } i \text{ is originally scheduled to land/take off no earlier than period } t \\ 0 & \text{otherwise} \end{cases}$$

$t_{ij}^{\text{min}}/t_{ij}^{\text{max}}$ = the minimum/maximum connection time between flights i and j , for $(i, j) \in \mathcal{C}$

The main novelty of our formulation, as compared to the one in [96], lies in the form of the scheduling parameters S_{it}^{arr} and S_{it}^{dep} . Note that S_{it}^{arr} (resp. S_{it}^{dep}) is equal to 1 if flight i is originally scheduled to land (resp. to take off) *no earlier than* period t (instead of *in* period t), i.e., S_{it}^{arr} and S_{it}^{dep} are of the form $(1, \dots, 1, 0, \dots, 0)$ (instead of $(0, \dots, 0, 1, 0, \dots, 0)$). This is similar to the formulation of Bertsimas and Stock Patterson [20] for a network ATFM problem and affects the definition of the decision variables below.

The parameters t_{ij}^{min} and t_{ij}^{max} are expressed as numbers of 15-minute periods. If there is an aircraft connection *and* a passenger connection between flights i and j , then we maintain both by setting the corresponding value of t_{ij}^{min} (resp. of t_{ij}^{max}) to the largest of the minimum (resp. to the smallest of the maximum) aircraft and passenger connection times.

Decision Variables

$$w_{it}^{\text{arr}}/w_{it}^{\text{dep}} = \begin{cases} 1 & \text{if flight } i \text{ is rescheduled to land/take off no earlier than period } t \\ 0 & \text{otherwise} \end{cases}$$

u_i = the displacement of flight $i \in \mathcal{F}$

λ_t^X/λ_t^Y = the number of scheduled arrivals/departures in period $t \in \mathcal{T}$

δ/Δ = the maximal flight displacement/the total schedule displacement

The variables w_{it}^{arr} and w_{it}^{dep} are the decision counterparts of the input parameters S_{it}^{arr} and S_{it}^{dep} . By convention, we assume that $w_{i,T+1}^{\text{arr}} = w_{i,T+1}^{\text{dep}} = 0, \forall i \in \mathcal{F}$. The displacement variables u_i are expressed as numbers of 15-minute periods. The value of u_i can be positive or negative, thus allowing any flight to be rescheduled later or earlier in the day.

Objective We consider the following two-stage lexicographic objective:

$$\text{lex min } (\delta, \Delta) \tag{5.1}$$

In other words, we first minimize the *maximal* displacement that any flight will sustain, denoted by δ . Among all feasible schedules that can be obtained under this objective, we select one that minimizes the *total* displacement, denoted by Δ . This choice is motivated by inter-flight equity concerns, as it ensures that no flight will incur a disproportionately large displacement [96]. We denote by δ^* and Δ^* the optimal value of δ and Δ , respectively.

Scheduling Constraints

$$|u_i| \leq \delta \quad \forall i \in \mathcal{F} \quad (5.2)$$

$$\sum_{i \in \mathcal{F}} |u_i| = \Delta \quad (5.3)$$

$$w_{i1}^{\text{arr}} = 1 \quad \forall i \in \mathcal{F} \quad (5.4)$$

$$w_{i1}^{\text{dep}} = 1 \quad \forall i \in \mathcal{F} \quad (5.5)$$

$$\sum_{t \in \mathcal{T}} (w_{it}^{\text{arr}} - S_{it}^{\text{arr}}) = u_i \quad \forall i \in \mathcal{F} \quad (5.6)$$

$$\sum_{t \in \mathcal{T}} (w_{it}^{\text{dep}} - S_{it}^{\text{dep}}) = u_i \quad \forall i \in \mathcal{F} \quad (5.7)$$

$$\sum_{t \in \mathcal{T}} (w_{jt}^{\text{dep}} - w_{it}^{\text{arr}}) \geq t_{ij}^{\text{min}} \quad \forall (i, j) \in \mathcal{C} \quad (5.8)$$

$$\sum_{t \in \mathcal{T}} (w_{jt}^{\text{dep}} - w_{it}^{\text{arr}}) \leq t_{ij}^{\text{max}} \quad \forall (i, j) \in \mathcal{C} \quad (5.9)$$

$$w_{it}^{\text{arr}} \geq w_{i,t+1}^{\text{arr}} \quad \forall i \in \mathcal{F}, \forall t \in \mathcal{T} \quad (5.10)$$

$$w_{it}^{\text{dep}} \geq w_{i,t+1}^{\text{dep}} \quad \forall i \in \mathcal{F}, \forall t \in \mathcal{T} \quad (5.11)$$

$$\sum_{i \in \mathcal{F}^{\text{arr}}} (w_{it}^{\text{arr}} - w_{i,t+1}^{\text{arr}}) = \lambda_t^X \quad \forall t \in \mathcal{T} \quad (5.12)$$

$$\sum_{i \in \mathcal{F}^{\text{dep}}} (w_{it}^{\text{dep}} - w_{i,t+1}^{\text{dep}}) = \lambda_t^Y \quad \forall t \in \mathcal{T} \quad (5.13)$$

Constraints (5.2) and (5.3) define the maximal displacement of flights and the total displacement of flights. Constraints (5.4) and (5.5) ensure that no flight is eliminated. Constraint (5.6) (resp. Constraint (5.7)) defines the displacement of any flight as the difference between its rescheduled arrival (resp. departure) time and its original arrival (resp. departure) time. The combination of constraints (5.6) and (5.7) ensures that scheduled block times are left unchanged. Constraints (5.8) and (5.9) force connection times to be larger than the minimum connection times and smaller than the maximum connection times that are imposed. Constraints (5.10) and (5.11) ensure that $w_{i.}^{\text{arr}}$ and $w_{i.}^{\text{dep}}$ are non-increasing for each flight i , consistently with their definition. Last, Constraints (5.12) and (5.13) define

the *aggregate* number of scheduled arrivals and departures per 15-minute period t , i.e., λ_t^X and λ_t^Y . This aggregate schedule will determine the arrival and departure demand rates into the Stochastic Queuing Model of congestion, and thus impact on-time performance at the airport. The queue length reduction constraints will thus aim to limit λ_t^X and λ_t^Y .

5.1.3 Queue Length Reduction Constraints

The combination of the Dynamic Programming model of capacity utilization and of the Stochastic Queuing Model of airport congestion (Chapters 3 and 4) determines the relationship between the number of scheduled arrivals and departures per period (λ_t^X and λ_t^Y , $t \in \mathcal{T}$) and the probability distribution of the arrival and departure queue lengths that minimize congestion costs, i.e., A_t and D_t for $t \in \mathcal{T}$. We can represent this relationship as the following function q :

$$q : (\lambda_1^X, \dots, \lambda_T^X, \lambda_1^Y, \dots, \lambda_T^Y) \mapsto (A_1, \dots, A_T, D_1, \dots, D_T) \quad (5.14)$$

We then need to define the queue length reduction constraints that ensure that the expected arrival and departure queue lengths do not exceed the prespecified limits A_{MAX} and D_{MAX} , respectively. They can be represented as Constraints (5.15) and (5.16). Obviously, A_t and D_t depend on flight schedules according to the relationship (5.14).

$$E(A_t) \leq A_{\text{MAX}} \quad \forall t \in \mathcal{T} \quad (5.15)$$

$$E(D_t) \leq D_{\text{MAX}} \quad \forall t \in \mathcal{T} \quad (5.16)$$

However, Constraints (5.15) and (5.16) cannot be expressed as linear constraints to the Integer Programming scheduling model. As mentioned earlier, this is because arrival and departure queue lengths depend nonlinearly on the schedule of flights. In the next section, we develop a solution algorithm that integrates these expected queue length reduction constraints into the scheduling model presented in this section.

5.2 Iterative Solution Algorithm

In this section, we present an original solution algorithm that integrates the models of scheduling interventions, of capacity utilization, and of airport congestion to solve the ICUSM. The algorithm finds the minimal schedule displacement (Equation (5.1)) to meet the queue length reduction constraints (5.15) and (5.16). It relies on the observation that, in contrast to stochastic queue dynamics, *deterministic* queue dynamics can easily be integrated into the Integer Programming scheduling framework. We first describe the corresponding parameters, variables and constraints. We then use this formulation to solve the ICUSM with *stochastic* queue dynamics.

5.2.1 Deterministic Queue Dynamics

We consider in this section $D(t)/D(t)/1$ queuing models of the arrival and departure queues. In order to simplify the model, we do not integrate the Markovian model of weather conditions (described in Section 2.3.2) but, instead, we consider queue dynamics during “all-VMC” and “all-IMC” days. For simplicity, we present below a generic case and we omit the “VMC” and “IMC” indices, but all sets, parameters and variables are defined under both VMC and IMC.

We first define the following set:

\mathcal{S}_t = the set of linear segments of the Operational Throughput Envelope
of the runway configuration in use during period t

Any Operational Throughput Envelope is then defined as a set of affine equations of the form $\alpha_s x + \beta_s y \leq \gamma_s, \forall s \in \mathcal{S}_t$, where the parameters $\alpha_s > 0$, $\beta_s \geq 0$ and $\gamma_s \geq 0$ define each segment of the envelope.

We add two pairs of variables:

μ_t^a / μ_t^d = the arrival / departure service rate selected during period t

A_t / D_t = the arrival / departure queue length at the end of period t

As in the Dynamic Programming model of capacity utilization, μ_t^a and A_t are integer variables and μ_t^d and D_t are continuous variables. The introduction of these continuous variables makes the model a Mixed Integer Program.

Finally, we add three constraints. Constraint (5.17) ensures that the service rates lie within the bounds defined by the Operational Throughput Envelope. Constraints (5.18) and (5.19) define the deterministic queue dynamics. In any period t , the arrival (resp. departure) queue length grows by the number of landings (resp. takeoffs) scheduled, minus the number of landings (resp. takeoffs) operated. These deterministic dynamics, in contrast to stochastic dynamics, depend linearly on flight schedules and service rates.

$$\alpha_s \mu_t^a + \beta_s \mu_t^d \leq \gamma_s \quad \forall t \in \mathcal{T}, \forall s \in \mathcal{S}_t \quad (5.17)$$

$$A_t = A_{t-1} + \lambda_t^X - \mu_t^a \quad \forall t \in \mathcal{T} \quad (5.18)$$

$$D_t = D_{t-1} + \lambda_t^Y - \mu_t^d \quad \forall t \in \mathcal{T} \quad (5.19)$$

As previously mentioned, we consider VMC and IMC queue dynamics. We denote by A_t^{VMC} (resp. D_t^{VMC}) the arrival (resp. departure) queue length at the end of period t during an “all-VMC day”. We denote by A_t^{IMC} and D_t^{IMC} the corresponding queue lengths during an “all-IMC” day.

5.2.2 Collinearity Assumption

At this point, we have integrated *deterministic* queue dynamics into the Integer Programming model of scheduling. However, we aim to find the optimal schedule that meets delay reduction targets under *stochastic* queue dynamics. Previous research has shown that deterministic

queuing models lead to significantly smaller delay estimates than stochastic queuing models [54]. Therefore, solving the ICUSM with deterministic queue dynamics would lead to an overly optimistic schedule, as a schedule of flights might meet the delay reduction targets with deterministic queue dynamics, but *not* with stochastic queue dynamics.

Nonetheless, it has also been found that delays estimated with stochastic and deterministic queuing models exhibit strong collinearity [54, 84]. Whereas stochastic delays are larger than deterministic delays, the dynamics of formation and propagation of delays through the course of the day are similar under deterministic and stochastic queue dynamics. We therefore make the following collinearity assumption: given two distinct schedules of flights, the one that leads to the smallest peak delays under a deterministic model will also lead to the smallest peak expected delays under a stochastic model. Under this assumption, for any given schedule displacement, the schedule that minimizes peak deterministic delays will also be the schedule that minimizes peak expected stochastic delays.

5.2.3 A Bi-level Iterative Solution Algorithm

Using the collinearity assumption, we develop an iterative bi-level solution algorithm to the ICUSM. At the flight level, for any given value of the maximal displacement δ and/or of the total displacement Δ , we determine a schedule that minimizes a measure q_{MAX} of *deterministic* delays, defined in Equations (5.20) to (5.22) (where $M_1, M_2 \gg 1$ are very large scalars):

$$q_{\text{MAX}}^{\text{VMC}} := M_1 \times \max \left(\frac{\max_{t \in \mathcal{T}} A_t^{\text{VMC}}}{A_{\text{MAX}}}, \frac{\max_{t \in \mathcal{T}} D_t^{\text{VMC}}}{D_{\text{MAX}}} \right) + \sum_{t \in \mathcal{T}} \left(\frac{A_t^{\text{VMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{VMC}}}{D_{\text{MAX}}} \right) \quad (5.20)$$

$$q_{\text{MAX}}^{\text{IMC}} := M_1 \times \max \left(\frac{\max_{t \in \mathcal{T}} A_t^{\text{IMC}}}{A_{\text{MAX}}}, \frac{\max_{t \in \mathcal{T}} D_t^{\text{IMC}}}{D_{\text{MAX}}} \right) + \sum_{t \in \mathcal{T}} \left(\frac{A_t^{\text{IMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{IMC}}}{D_{\text{MAX}}} \right) \quad (5.21)$$

$$q_{\text{MAX}} := M_2 \times q_{\text{MAX}}^{\text{VMC}} + q_{\text{MAX}}^{\text{IMC}} \quad (5.22)$$

Stated differently, we consider here a multi-stage, sequential, lexicographic formulation, captured by the very large scalars M_1 and M_2 . We first minimize a measure of peak deter-

ministic delays that are incurred during an “all-VMC” day. We “normalize” the arrival (resp. departure) queue length by a factor A_{MAX} (resp. D_{MAX}). The purpose of the normalization is to capture the relative “cost” of increasing the expected arrival (resp. departure) queue length vis-à-vis the target levels A_{MAX} (resp. D_{MAX}). Among all schedules that minimize the largest normalized queue lengths, we select the one that minimizes the “total” normalized queue length, expressed as $\sum_{t \in \mathcal{T}} \left(\frac{A_t^{\text{VMC}}}{A_{\text{MAX}}} + \frac{D_t^{\text{VMC}}}{D_{\text{MAX}}} \right)$ (see Equation (5.20)).

For a large schedule displacement, it might be possible to keep scheduling levels within the bounds defined by the VMC Operational Throughput Envelopes and thus to obtain null VMC queue lengths at any time, i.e., $a_t^{\text{VMC}} = d_t^{\text{VMC}} = 0, \forall t \in \mathcal{T}$. If this occurs, we minimize a measure of deterministic delays during “all-IMC” days, i.e., $q_{\text{MAX}}^{\text{IMC}}$ (Equation (5.21)), subject to the additional constraints:

$$A_t^{\text{VMC}} = D_t^{\text{VMC}} = 0 \quad \forall t \in \mathcal{T} \quad (5.23)$$

At the aggregate level, we use the resulting schedule of flights to determine the optimal control of arrival and departure service rates and we compute the resulting queue lengths under *stochastic* queue dynamics. This determines whether the optimal displacement is larger or smaller than the considered displacement δ/Δ (see details in the next two paragraphs). If the peak expected delays are *larger* than A_{MAX} and D_{MAX} , then the optimal displacement is larger than the considered displacement, i.e., the scheduling interventions have to be more aggressive in order to meet the targets. If, however, the delays are *smaller* than the queue length targets, then a feasible solution has been found and the optimal displacement is smaller than the considered displacement. This approach is based on the non-increasing relationship between schedule displacement and deterministic delays, hence between schedule displacement and expected stochastic delays under the collinearity assumption (Section 5.2.2).

Algorithm 3 presents the algorithm that determines the optimal maximal displacement δ^* by iteratively updating a lower bound of δ^* , denoted by $\underline{\delta}$. We initialize the algorithm with a value of $\underline{\delta} = 0$, i.e., no flight is displaced. We apply the Dynamic Programming model

of capacity utilization and we estimate stochastic delays with the original schedule of flights. If the queue length targets are met, then the optimal maximal displacement δ^* is equal to 0. Otherwise, we increase the value of $\underline{\delta}$ to 1. We obtain the schedule that minimizes peak delays for a value of $\delta = 1$. We do not impose any restriction on the total displacement at this point. Using the modified schedule, we re-optimize the control of service rates and we compute stochastic delays. If the queue length targets are met, then the optimal maximal displacement δ^* is equal to 1. Otherwise, we increase the value of $\underline{\delta}$ to 2, and repeat the process until the targets are met. We denote by Δ_0 the total displacement that minimizes q_{MAX} for a maximal displacement equal to δ^* .

Algorithm 3 Determination of δ^*

Initialization: $\underline{\delta} = 0, z_{\text{end}} = 0$
while $z_{\text{end}} = 0$ **do**
Solve the model of scheduling interventions, under *deterministic* queue dynamics $\rightarrow \lambda_t^X / \lambda_t^Y, \forall t$
 minimize q_{MAX} (Equation 5.22)
 subject to Scheduling constraints: Equations (5.4) to (5.3)
 Deterministic queuing constraints: Equations (5.17) to (5.19)
 Maximal displacement: $\delta = \underline{\delta}$
 $\Delta_0 \leftarrow \sum_{i \in \mathcal{F}} |u_i|$
Solve the model of capacity utilization (Equation (3.3)) $\rightarrow \mu_t^a(a_{t-1}, d_{t-1}, w_t) / \mu_t^d(a_{t-1}, d_{t-1}, w_t), \forall t$
Solve the model of airport congestion (Equation (2.1)) $\rightarrow A_t, D_t, \forall t$
 if $A_t \leq A_{\text{MAX}}, \forall t$ and $D_t \leq D_{\text{MAX}}, \forall t$ **then** $z_{\text{end}} \leftarrow 1$
 else $\underline{\delta} \leftarrow \underline{\delta} + 1$
 end if
end while
 $\delta^* \leftarrow \underline{\delta}$
Return δ^*, Δ_0

Algorithm 4 shows the iterative algorithm that determines the optimal total displacement Δ^* , given the optimal maximal displacement δ^* . We denote by $\overline{\Delta}$ and $\underline{\Delta}$ an upper bound and a lower bound of Δ^* , respectively. We initialize the algorithm by setting $\underline{\Delta}$ to 0, which corresponds to the situation where no flight is displaced, and $\overline{\Delta}$ to Δ_0 , which provides the smallest peak queue length attainable. We proceed by dichotomy. At each iteration, we consider a tentative value of the total displacement, denoted by Δ^{try} , at the midpoint of $\overline{\Delta}$ and $\underline{\Delta}$. We find the schedule that minimizes peak deterministic delays for the value of $\Delta = \Delta^{\text{try}}$. Using this schedule, we optimize the control of service rates and we compute

stochastic delays. If the queue length targets are met, then the optimal total displacement is at most equal to Δ^{try} , so we set $\bar{\Delta}$ to Δ^{try} . Otherwise, the optimal total displacement is larger than Δ^{try} , so we set $\underline{\Delta}$ to Δ^{try} . We repeat this process until $\bar{\Delta}$ and $\underline{\Delta}$ have converged to the same value, which is then equal to the optimal total displacement Δ^* . This dichotomic algorithm converges in $\mathcal{O}(\log_2(F))$ iterations.

Algorithm 4 Determination of Δ^* , given δ^*

```

Initialization:  $\underline{\Delta} = 0, \bar{\Delta} = \Delta_0, z_{\text{end}} = 0$ 
while  $z_{\text{end}} = 0$  do
   $\Delta^{\text{try}} \leftarrow \frac{\bar{\Delta} + \underline{\Delta}}{2}$ 
  Solve the model of scheduling interventions, under deterministic queue dynamics  $\rightarrow \lambda_t^X / \lambda_t^Y, \forall t$ 
    minimize  $q_{\text{MAX}}$  (Equation 5.22)
    subject to Scheduling constraints: Equations (5.4) to (5.3)
               Deterministic queuing constraints: Equations (5.17) to (5.19)
               Maximal displacement:  $\delta = \delta^*$ 
               Total displacement:  $\Delta = \Delta^{\text{try}}$ 
  Solve the model of capacity utilization (Equation (3.3))  $\rightarrow \mu_t^a(a_{t-1}, d_{t-1}, w_t) / \mu_t^d(a_{t-1}, d_{t-1}, w_t), \forall t$ 
  Solve the model of airport congestion (Equation (2.1))  $\rightarrow A_t, D_t, \forall t$ 
    if  $A_t \leq A_{\text{MAX}}, \forall t$  and  $D_t \leq D_{\text{MAX}}, \forall t$  then  $\bar{\Delta} \leftarrow \Delta^{\text{try}}$ 
    else  $\underline{\Delta} \leftarrow \Delta^{\text{try}}$ 
    end if
    if  $\underline{\Delta} = \bar{\Delta}$  then  $z_{\text{end}} \leftarrow 1$ 
    else  $z_{\text{end}} \leftarrow 0$ 
    end if
end while
Return  $\Delta^*$ 

```

The iterative algorithm relies on our collinearity assumption. In fact, this assumption may introduce an error in some cases. There may exist, for a given displacement, a schedule of flights that reduces peak expected stochastic delays to a greater extent than the schedule minimizing peak deterministic delays. In such instances, the algorithm may not be able to find the exact optimal solution. Nonetheless, we expect such errors to be of second order. In general, the modified schedule of flights that we obtain is expected to be very close to the optimal schedule. The iterative algorithm thus determines a good approximation of the optimal displacement and of the optimal schedule meeting the expected queue length targets.

5.2.4 Size of the Formulation

The solution algorithm presented in this section iterates between three phases: a Mixed Integer Program (MIP), a Dynamic Program (DP), and a Stochastic Queuing Model (SQM). The efficiency of the algorithm depends on the size of each of these sub-problems. We report in Table 5.1 the number of variables and constraints of the MIP, the number of states, actions and time periods of the DP and the number of states and time periods of the SQM. Note that the complexity of the MIP scales quadratically with the number of flights F included in the model. It is thus sensitive to variations in the number of flights scheduled. In contrast, the size of the DP and of the SQM only depends on fixed variables and thus does not change from one computational setting to another.

Table 5.1: Size of the model

Model	Size
MIP	Number of binary variables: $2FT + F$ Number of integer variables: $2F + 4T$ Number of continuous variables: $2T$ Number of constraints (upper bound): $2F^2 + 2FT + 6F + 6T + \bar{\Gamma}T + 1$
DP	Number of states: $2(N + 1)^2$ Number of actions: $\bar{\Gamma} + 1$ Number of periods: T_0
SQM	Number of states: $2(N + 1)^2$ Number of periods: T_0

5.3 Computational Results

In this section, we show results of the ICUSM at JFK. Similar results obtained at LGA are provided in Appendix C. Scheduling and capacity data were obtained using the procedures described in Section 2.5. Aircraft connection data were obtained using information on aircraft tail numbers reported in the ASPM database. In cases where this information is missing (which is the case for approximately 40% of the flights, including all international flights), we applied a simple fleet assignment model that reconstructs aircraft itineraries, described

in Appendix B. We use the *minimum* aircraft turnaround time between any pair of flights estimated by Pyrgiotis [94] as a function of the aircraft type, of the airline and of whether the airport is a hub airport for the airline or not. We define the *maximum* turnaround time of any pair of successive flights as the turnaround time that was actually planned by the airline plus 15 minutes, so aircraft utilization remains very close to what was originally planned. We obtained passenger connection data from a database developed by Barnhart et al. [11], based on a discrete choice model for estimating historical passenger flows. We estimate the minimum passenger connection time at JFK as the 5th percentile of the distribution of all planned passenger connection times. These aircraft and passenger connection data were used to construct the set \mathcal{C} and the parameters t^{\min} and t^{\max} .

We describe the convergence of the iterative algorithm (Section 5.3.1), and the effects of scheduling interventions on flight schedules and airport on-time performance (Sections 5.3.2 and 5.3.3). We discuss the benefits of our integrated approach in the next section. Section 5.5 concludes.

5.3.1 Convergence of the Iterative Algorithm

In this section, we describe the convergence of the iterative algorithm developed in Section 5.2. Figure 5-2a shows the maximal displacement δ and the upper and lower bounds of the total displacement, $\bar{\Delta}$ and $\underline{\Delta}$, after each iteration of the algorithm, obtained with the schedule of 05/25/2007. The queue length reduction targets are set to $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$, i.e., we require that the expected arrival and departure queue lengths do not exceed 11 aircraft and 15 aircraft, respectively. Figure 5-2b shows the peak expected arrival and departure queue lengths after each iteration.

During the first three iterations, the value of the maximal displacement δ is updated (see Algorithm 3). After the first iteration, i.e. with the original schedule of flights, peak expected arrival and departure queue lengths are respectively equal to 13.7 and 32.3 aircraft. After the second iteration, i.e. with a maximal displacement of one 15-minute period, the queue length targets are still not met as the expected departure queue length peaks at 17.0 aircraft.

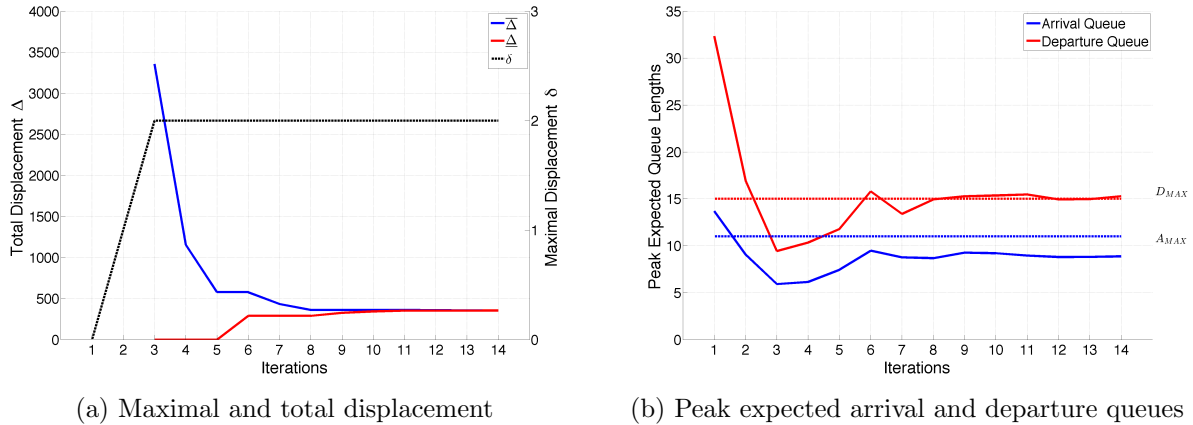


Figure 5-2: Convergence of the algorithm: displacement and queue lengths per iteration

At the third iteration, a schedule that meets the queue length targets is found with a value of the maximal displacement $\delta = 2$. Therefore, the optimal maximal displacement δ^* is equal to 2. In the remaining iterations, we apply the dichotomic algorithm (see Algorithm 4). It essentially finds the schedule that leads to expected delays as close as possible to the queue length targets by adjusting the upper and lower bounds of the total displacement until they converge to the same value.

Table 5.2 reports computational results for three distinct days in 2007, for different values of the queue length reduction targets A_{MAX} and D_{MAX} . As expected, the optimal solution is obtained after 11-14 iterations in each case (i.e., in $\mathcal{O}(\log_2(F))$). The computational time of the MIP exhibits high variability from one day to another, and from one value of the total displacement to another. Even though no clear trend appears, computational times seem to be smallest for small values of the number of flights F and for either very large or moderate values of the total displacement Δ . For instance, each iteration terminates quickly with $A_{MAX} = 15$ and $D_{MAX} = 20$, which induce a relatively small schedule displacement. In contrast, some iterations with more aggressive queue length reduction targets are more time-consuming. In some cases, no optimal solution is found and the solution obtained after 1,800 seconds is returned.

The overall iterative algorithm terminates after a small number of iterations and its computational times are reasonable. In practical instances, the running time of the whole

algorithm varies from 90 minutes to several hours. Note that running time is a secondary consideration in this case, as long as it is not prohibitively long. In view of its strategic nature, the model is run “off-line” several months in advance to determine the recommended scheduling interventions. Thus, running times of even several hours are perfectly acceptable. Moreover, a close-to-optimal solution is found after only 6-9 iterations, when the range between the upper bound and the lower bound of the total displacement becomes of the order of 10%. This typically reduces computational times by one third. In summary, a modified schedule that meets the queue length targets while minimizing the changes from the original schedule of flights can be obtained quickly after a small number of iterations.

Table 5.2: Computational results

		Day: 02/04/2007 $F = 1,465, T = 170$					Day: 09/18/2007 $F = 1,592, T = 185$					Day: 05/25/2007 $F = 1,678, T = 184$				
		Time (s)					Time (s)					Time (s)				
Iter.	δ	Δ	MIP	DP	SQM	δ	Δ	MIP	DP	SQM	δ	Δ	MIP	DP	SQM	
$A_{\text{MAX}} = 15, D_{\text{MAX}} = 20$	1	0	0	9	64	49	0	0	14	64	50	0	0	11	65	50
	2	1	1,465	19	64	50	1	1,592	23	64	50	1	1,678	17	65	50
	3	1	279	23	63	49	1	284	21	64	49	1	371	18	64	50
	4	1	139	35	64	50	1	142	1,528	65	51	1	186	84	64	50
	5	1	70	18	64	49	1	71	25	64	50	1	93	138	64	50
	6	1	35	16	64	50	1	36	21	65	50	1	47	28	65	50
	7	1	18	13	63	50	1	18	22	65	51	1	70	23	64	50
	8	1	27	13	64	50	1	27	21	65	51	1	82	34	64	50
	9	1	31	16	64	50	1	32	22	65	51	1	88	19	64	50
	10	1	33	16	64	49	1	34	21	65	51	1	91	60	65	50
	11	1	32	16	64	49	1	33	20	65	51	1	90	24	64	50
$A_{\text{MAX}} = 11, D_{\text{MAX}} = 15$	1	0	0	9	62	49	0	0	16	64	49	0	0	11	64	50
	2	1	1,465	10	64	50	1	1,592	22	64	50	1	1,678	17	64	50
	3	1	279	23	64	49	1	303	20	64	50	2	3,356	137	64	50
	4	1	139	20	63	49	1	152	576	64	49	2	1,155	153	64	50
	5	1	70	18	64	50	1	228	485	63	49	2	578	176	64	50
	6	1	105	21	64	50	1	190	1,800	63	49	2	289	1,630	64	50
	7	1	122	15	63	50	1	209	25	65	51	2	434	1,800	64	50
	8	1	114	20	63	50	1	200	1,800	64	50	2	362	1,800	64	50
	9	1	118	18	65	49	1	205	1,800	64	50	2	326	1,800	64	50
	10	1	116	20	63	49	1	207	28	64	51	2	344	1,278	64	50
	11	1	115	18	64	50	1	206	31	64	51	2	353	1,800	64	50
	12	-	-	-	-	-	-	-	-	-	-	2	358	1,800	64	50
	13	-	-	-	-	-	-	-	-	-	-	2	356	1,800	65	50
	14	-	-	-	-	-	-	-	-	-	-	2	355	1,800	64	51

5.3.2 Optimal Schedules and Delays

This section shows the effects of the ICUSM on flight schedules and delays. Table 5.3 reports the optimal displacement and the optimal peak expected queue lengths for the 9 days considered and for different values of A_{MAX} and D_{MAX} . The second column shows the number of flights F included in the model (the number in parentheses is the number of flights to or from JFK). In most cases, the first queue length reduction targets ($A_{\text{MAX}} = 15$ and $D_{\text{MAX}} = 20$) induce a very small displacement: no flight is displaced by more than 15 minutes and only 2% to 4% of the flights scheduled at JFK are displaced. At the same time, peak expected departure delays are reduced by an estimated 25% to 30%. For the two busiest days, the optimal displacement is slightly larger, but peak expected departure delays are reduced by an estimated 40%. Even though the constraint on peak arrival queue lengths is non-binding (the peak expected arrival queue lengths under the *original* schedules are all smaller than 15 aircraft) the rescheduling of flights reduces, in most cases, peak arrival queue lengths by 3% to 10%. In other words, de-peaking the departure schedule also de-peaks the arrival schedule, an effect primarily due to the aircraft and passenger connectivity constraints.

Flight delays can be even more significantly mitigated by applying more stringent queue length reduction targets. In this series of tests, setting $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$ reduces peak expected arrival (resp. departure) delays by 5% to 40% (resp. 45% to 60%). Although larger than previously, the schedule displacement remains moderate. The maximal displacement is equal to 1 15-minute period for 7 days and to 2 periods for the two busiest days. Substantial delay reductions can therefore be achieved without eliminating any flight, any aircraft or passenger connection and without displacing any flight by more than 30 minutes or, in most cases, by 15 minutes. Moreover, 75% to 95% of the flights scheduled at JFK are *not* displaced at all.

Note, as well, that the optimal maximal and total displacements do not increase monotonically with the number of flights F considered in the model. This is a surprising observation as it might be expected that the more flights are scheduled in a day, the more costly the queue

Table 5.3: Optimal displacement and peak queue lengths for the 9 days considered

Day	F	$A_{MAX} = \infty$ $D_{MAX} = \infty$		$A_{MAX} = 15$ $D_{MAX} = 20$		$A_{MAX} = 11$ $D_{MAX} = 15$		δ^*	Δ^*	Arr. Queue	Dep. Queue
		Arr. Queue	Dep. Queue	δ^*	Δ^*	Arr. Queue	Dep. Queue				
02/04	1,465 (1,070)	13.6	28.0	1	32	13.0 (-4.3%)	19.6 (-30.1%)	1	116	10.3 (-24.1%)	14.9 (-44.9%)
01/10	1,550 (1,112)	10.7	27.0	1	28	10.7 (+0.1%)	19.5 (-27.5%)	1	72	10.4 (-2.7%)	15.0 (-44.4%)
02/08	1,601 (1,137)	13.3	27.1	1	26	12.9 (-2.9%)	19.8 (-27.1%)	1	74	10.9 (-17.9%)	14.9 (-45.0%)
01/25	1,582 (1,153)	13.7	27.2	1	33	13.4 (-2.7%)	19.7 (-27.8%)	1	105	10.7 (-22.3%)	14.7 (-46.2%)
09/18	1,592 (1,162)	14.6	28.1	1	34	13.7 (-6.2%)	19.9 (-29.4%)	1	206	8.4 (-42.7%)	15.0 (-46.9%)
10/15	1,598 (1,177)	13.2	28.1	1	41	12.5 (-5.3%)	19.7 (-30.0%)	1	194	8.9 (-32.5%)	15.0 (-46.8%)
06/01	1,608 (1,192)	12.5	26.7	1	43	12.2 (-2.5%)	19.6 (-26.6%)	1	140	9.8 (-21.5%)	15.0 (-43.9%)
07/07	1,553 (1,212)	12.2	34.7	2	129	10.6 (-13.1%)	20.0 (-42.4%)	2	305	9.4 (-22.4%)	14.9 (-57.0%)
05/25	1,678 (1,229)	13.7	32.3	1	91	12.3 (-10.3%)	19.8 (-38.8%)	2	356	8.8 (-35.6%)	15.0 (-53.7%)

length reduction constraints would be. However, the optimal displacement depends not only on F , but also on the way these flights are distributed over the day. The more peaked the original schedule, the larger the peak expected arrival and departure delays and, in turn, the larger the optimal displacement is likely to be. For instance, even though more flights were scheduled at JFK on 01/10 than on 02/04, peak expected departure queue lengths are larger on 02/04 than on 01/10 under the original schedules and the optimal total displacement is larger for the schedule of 02/04 than for that of 01/10.

We now explore the effects of increasingly stringent queue length reduction targets imposed on the schedule of 05/25/2007. Table 5.4 defines 5 different tests and reports, for each one, the optimal displacement, the peak expected queue lengths and the average delays over the whole day. In the first test, we impose no queue length reduction constraint. The original schedule is therefore left unchanged and the expected arrival and departure

queue lengths peak at 13.7 aircraft and 32.3 aircraft, respectively. In the four remaining tests, we progressively reduce A_{MAX} and D_{MAX} , so the schedule displacement increases and flight delays decrease. Again, substantial delay reductions can be achieved through limited changes in flight schedules. The peak expected arrival and departure delays can be reduced by approximately 15% and 40%, respectively, without displacing any flight by more than 15 minutes. This corresponds to respective declines in the *average* arrival and departure delays during the whole day of 5% and 20%. Further delay reductions can be achieved by displacing some flights by 30 minutes. Test 4 indicates that the peak expected arrival and departure delays can be reduced by as much as 35% and 55%, respectively, which corresponds to respective reductions in the average arrival and departure delays of nearly 20% and 40%. Test 5 suggests that even more substantial delay reductions can be achieved with a maximal displacement of $\delta^* = 2$ 15-minute periods, but this comes at the price of a much larger total displacement Δ^* . We further investigate the sensitivity of the optimal displacement to the queue length reduction targets in Section 5.3.3.

Table 5.4: Displacement and delays for different expected queue length targets

Test	Targets		Optimal Disp.		Peak Queue Lengths		Average Delay	
	A_{MAX}	D_{MAX}	δ^* (periods)	Δ^* (periods)	Arrival (A/C)	Departure (A/C)	Arrival (mins.)	Departure (mins.)
1	∞	∞	0	0	13.7	32.3	9.1	21.5
2	13	25	1	37	12.9	23.9	8.9	19.0
					(-5.6%)	(-26.1%)	(-2.2%)	(-11.3%)
3	12	20	1	105	11.6	19.6	8.7	16.9
					(-15.6%)	(-39.3%)	(-5.1%)	(-21.2%)
4	11	15	2	356	8.8	15.0	7.5	12.3
					(-35.6%)	(-53.7%)	(-17.9%)	(-42.9%)
5	10	10	2	1,129	6.5	9.9	5.6	8.2
					(-52.7%)	(-69.3%)	(-38.9%)	(-61.8%)

Figure 5-3 depicts the original schedule of 05/25/2007 (Figure 5-3a) and the optimal modified schedule obtained with $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$ (Figure 5-3b). Since no flight is eliminated, the scheduling interventions reduce peak scheduling levels by rescheduling flights more evenly through the course of the day. Whereas over 30 flights were originally

scheduled during some periods of the day, no more than 22 flights are scheduled at any period after schedule de-peakng. Moreover, the arrival and departure schedules may be affected differently. For instance, a large number of departures are scheduled at JFK in the morning while arrival demand lies below capacity. As a result, the ICUSM “smoothens” the morning schedule of takeoffs but leaves the arrival schedule almost unchanged.

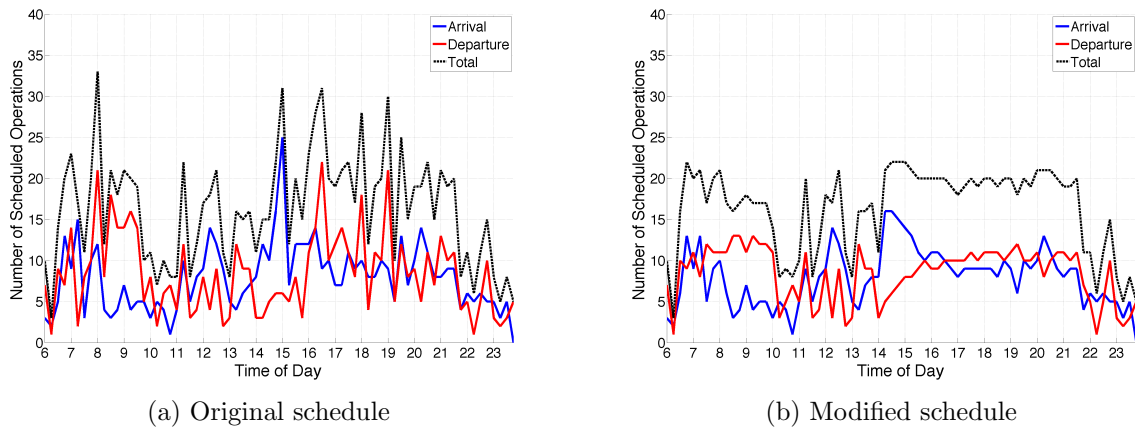


Figure 5-3: Original and modified schedules of flights ($A_{MAX} = 11$ and $D_{MAX} = 15$)

Note that the modified schedule is *not* distributed evenly over the day. This is an important observation as it is well known that, for any given number of flights, delays will be the smallest when the flights are evenly distributed over the day. But a completely “flat” schedule would generally induce a larger displacement than the solution produced by the ICUSM. Instead, the model maintains some schedule peaks and valleys, albeit of smaller magnitude than the corresponding variations in the original schedule. For instance, the modified schedule exhibits above-capacity scheduling levels at some peak morning and afternoon hours and a schedule slack at off-peak hours. This is far more realistic and consistent with airline preferences and underlying passenger demand than a completely flat schedule. Therefore, the solution from the ICUSM induces smaller changes to the original schedule of flights than a schedule obtained through the imposition of schedule limits. Put another way, for a given schedule displacement, applying flight caps may not result in a delay-minimizing schedule of flights. We further investigate the benefits of our integrated approach in Section 5.4.

Figure 5-4 shows the evolution of the expected arrival (Figure 5-4a) and departure (Fig-

ure 5-4b) queue lengths over the day under the original schedule of 05/25/2007 and the modified schedules from Tests 3, 4 and 5 (see Table 5.4). As detailed previously, the scheduling interventions reduce, in these cases, peak expected arrival and departure queue lengths by an estimated 15% to 55% and 40% to 60%, respectively. The smoothing of flight schedules slightly extends peak scheduling periods, so queues may form earlier with the modified schedules than with the original schedule. For instance, afternoon departure queue lengths start to build up around 3 pm under the modified schedule, while they remain very low until 4 pm under the original schedule. However, the magnitude of these delays remains much more manageable under the modified schedule. Instead of increasing almost instantaneously to over 30 aircraft, the expected departure queue length increases at a lower rate up to 10 to 20 aircraft, depending on the scenario considered. The queue lengths then become stable until the end of the evening peak under the modified schedules.

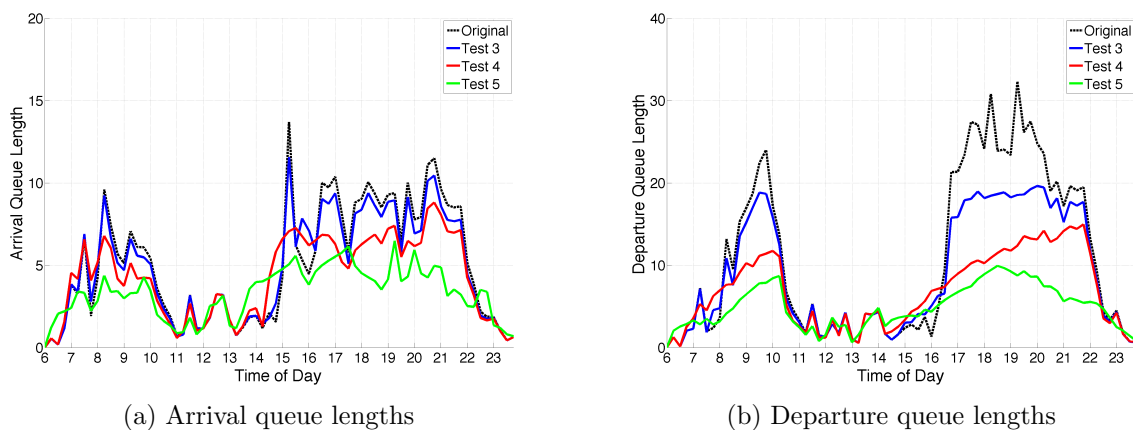


Figure 5-4: Expected queue lengths before and after the scheduling interventions

5.3.3 Sensitivity of Displacement to Queue Length Targets

In order to guide the selection of the queue length reduction targets, we plot in Figure 5-5 the sensitivity of the optimal maximal displacement δ^* and of the optimal total displacement Δ^* to A_{MAX} and D_{MAX} . In this figure, we impose the same arrival and departure queue length targets, i.e., $A_{MAX} = D_{MAX}$. Under this assumption, the departure queue length

constraint is generally binding as the departure queue was more peaked than the arrival queue under the original schedule (see Table 5.3). Note that the schedule displacement increases super-linearly as the queue length reduction targets become more stringent. In other words, significant delay reductions can be achieved through limited interference with the original schedule of flights, while the most stringent delay reduction objectives may require much larger schedule changes.

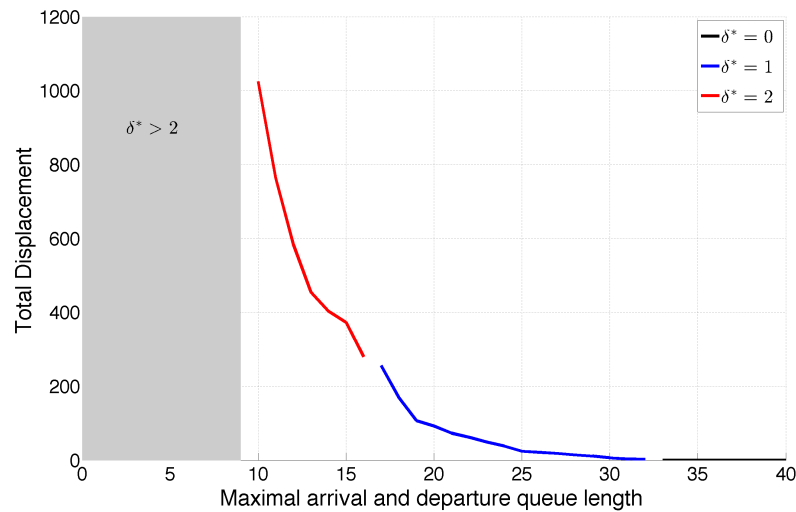


Figure 5-5: Sensitivity of the optimal displacement as a function of $A_{\text{MAX}} = D_{\text{MAX}}$

First, peak expected departure delays can be reduced by nearly 50% with a small schedule displacement. Expected queue length targets as low as $A_{\text{MAX}} = D_{\text{MAX}} = 17$ aircraft can be met without displacing any flight by more than 15 minutes ($\delta^* = 1$). Further delay reductions can be achieved by displacing some flights by 30 minutes, while keeping the total displacement Δ^* relatively low. For instance, targets equal to $A_{\text{MAX}} = D_{\text{MAX}} = 15$ aircraft reduce peak expected departure delays by 55% with a total displacement equal to 373 15-minute periods.

In contrast, the most stringent delay reduction targets induce significantly larger schedule displacements. For instance, reducing the peak expected departure queue from 20 to 10 aircraft increases the optimal total displacement from 92 to 1,025 15-minute periods. In addition, the most aggressive delay reduction objectives cannot be achieved without substantially

interfering with airline schedules. In the case presented here, the expected queue lengths cannot be kept below 9 aircraft throughout the day without displacing some flights by more than 30 minutes. Further reducing the queue length targets might require even more substantial changes in the planned network of flights, including the elimination of some flights. In other words, the most stringent delay reduction targets cannot be met through the type of intervention modeled in the ICUSM and would require more aggressive demand management strategies. Nonetheless, the results from this section suggest that the combination of limited changes in airline schedules and the optimization of operating procedures can result in substantial mitigation of airport congestion.

5.4 Benefits of Integration

In this chapter, we have developed an original approach that integrates flight scheduling and airport capacity utilization. However, in practice, scheduling and operational decisions are typically made sequentially and independently. Flight schedules are produced months in advance by the airlines and consider only in rough terms (or, often not at all) the endogeneity of airport operations, i.e., the impact of scheduling decisions on airport capacity utilization and on flight delays. Conversely, tactical models of airport operations consider the flight schedule as given and are concerned with minimizing congestion costs for that specific schedule. In this section, we compare our integrated approach to congestion mitigation with a sequential approach that solves the scheduling and operational problems separately, in consecutive steps. Figure 5-6 shows a schematic representation of the integrated approach (Figure 5-6a) and of the sequential approach (Figure 5-6b).

To develop a model of the sequential approach, we smoothen the original schedule of flights by applying scheduling limits [96]. The queue length reduction constraints (see Section 5.1.1) are replaced by scheduling limit constraints defined below (Constraints (5.24), (5.25) and (5.26)), which state that the number of scheduled flights, landings and takeoffs at any period t must not exceed limits denoted by C_t , C_t^{arr} and C_t^{dep} , respectively. For each value of C_t , we select values of C_t^{arr} and C_t^{dep} slightly larger (by 10%) than $\frac{C_t}{2}$. The parameters

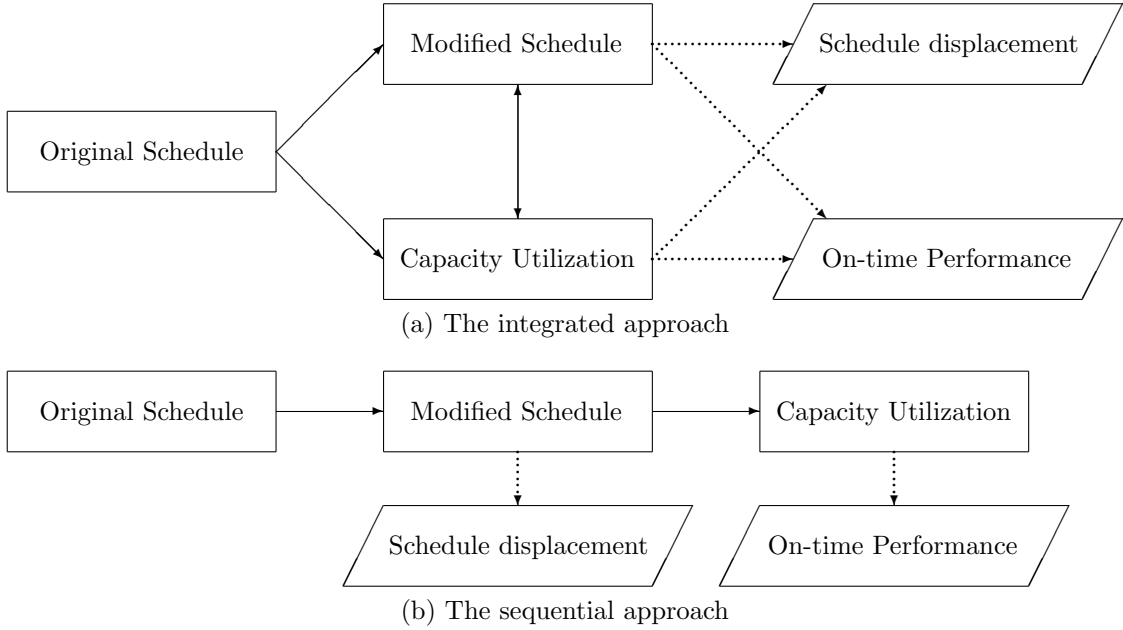


Figure 5-6: Integrated and sequential approaches to airport congestion mitigation

C_t , C_t^{arr} and C_t^{dep} can be chosen as vectors of integers, but they need not be constant over the T periods. We denote by \widehat{C} the average value of the series of limits C_t , e.g., $\widehat{C} = 20.50$ if the number of scheduled flights per 15-minute period is capped by 20, 21, 20, 21, etc. Consistently with practice, we assume that these limits exhibit hourly periodicity.

$$\lambda_t^X + \lambda_t^Y \leq C_t \quad \forall t \in \mathcal{T} \quad (5.24)$$

$$\lambda_t^X \leq C_t^{\text{arr}} \quad \forall t \in \mathcal{T} \quad (5.25)$$

$$\lambda_t^Y \leq C_t^{\text{dep}} \quad \forall t \in \mathcal{T} \quad (5.26)$$

Following the rescheduling of flights, we optimize the control of runway configurations and of arrival and departure service rates under the resulting schedule by subsequently applying the Dynamic Programming model of capacity utilization. Finally, we compute expected arrival and departure queue lengths with the stochastic queuing model of congestion. For any pair of queue length reduction targets A_{MAX} and D_{MAX} , we denote by \widehat{C}^* the optimal level of schedule limits, i.e., the largest limits that satisfy the queue length reduction constraints.

Table 5.5 reports the optimal value of the maximal displacement δ^* and of the total displacement Δ^* under the integrated and sequential approaches and the optimal schedule limits \widehat{C}^* under the sequential approach. The results show that the integrated approach induces, in our 9 computational tests, a substantially smaller schedule displacement than the sequential approach. In other words, the same delay reductions can be achieved under the integrated approach through much smaller changes in flight schedules than under the sequential approach. This, in turn, demonstrates the importance of integrating queue length reduction constraints into the scheduling framework, instead of applying predetermined limits. This is consistent with our previous observation that the optimal schedule of flights obtained with the ICUSM is not flat but, instead, exhibits peaks and valleys.

Table 5.5: Optimal displacement under the integrated and sequential approaches

Day	$A_{\text{MAX}} = 15$ $D_{\text{MAX}} = 20$					$A_{\text{MAX}} = 11$ $D_{\text{MAX}} = 15$				
	Integrated		Sequential			Integrated		Sequential		
	δ^*	Δ^*	\widehat{C}^*	δ^*	Δ^*	δ^*	Δ^*	\widehat{C}^*	δ^*	Δ^*
02/04	1	32	22.25	1	135	1	116	20.00	2	276
01/10	1	28	22.25	1	144	1	72	20.75	1	268
02/08	1	26	22.50	1	119	1	74	20.25	2	250
01/25	1	33	22.50	1	143	1	105	20.75	2	252
09/18	1	34	20.50	1	199	1	206	19.75	2	303
10/15	1	41	20.75	1	269	1	194	19.50	2	440
06/01	1	43	21.75	1	226	1	140	20.25	2	331
07/07	2	129	20.75	2	325	2	305	19.50	2	645
05/25	1	91	20.25	2	353	2	356	19.50	2	591

Another interesting observation is that, for the sequential approach, the optimal scheduling limits \widehat{C}^* vary from day to day, for given queue length reduction targets. In general, \widehat{C}^* tends to be smallest on the busiest days. This is explained by the fact that the duration of peak scheduling periods is the longest on the busiest days and therefore schedule limits have to be smaller to prevent queues from growing above the specified queue length limits. This result underscores that the “optimal” stringency of demand management measures at any given airport depends strongly on the total number of flights in a day and on how these

flights are distributed over the day. All else being equal, the busier the day, the more aggressive demand management should be. This suggests that current approaches to “Level 3” schedule coordination worldwide, which impose scheduling limits based solely on single-value estimates of airport capacity, could be significantly improved.

Finally, we compare the optimal expected queue lengths that can be obtained under the integrated and sequential approaches, for a given schedule displacement. Figure 5-7 shows the expected arrival (Figure 5-7a) and departure (Figure 5-7b) queue lengths obtained with the 09/18/2007 schedule, with $\delta = 1$ (i.e., no flight is displaced by more than 15 minutes) and $\Delta = 0.2 \times F$ (i.e., no more than 20% of the flights are displaced). A similar picture emerges with other values of displacement. The black, dashed lines show the expected arrival and departure queue lengths under the integrated approach developed in this chapter. The red, solid lines show the corresponding queue lengths obtained under the sequential approach, after determining the “optimal” value of C , i.e., the largest value of C that leads to a schedule displacement smaller than (δ, Δ) . The figure shows that the sequential approach results in a substantially larger peak expected departure queue length than the integrated approach, by an estimated 26%. This pattern was obtained consistently in all our computational tests. Note, also, that the departure queue length is greater under the integrated approach than under the sequential approach in the morning. This is because the original schedule of flights on 09/18/2007 exhibits a larger number of departures in the morning than at peak afternoon hours. As a result, the ICUSM assigns a larger number of departures per period in the morning than in the afternoon, while the sequential approach applies the same scheduling limits throughout the day regardless of the scheduling patterns at the airport. In other words, for a given “budget” of schedule displacement, the integrated approach does much better in selecting *which* flights to reschedule to achieve the greatest delay reductions. Overall, we estimate that the congestion costs (which depend quadratically on arrival and departure queue lengths) are 25% to 50% *larger* under the sequential approach than under the integrated approach.

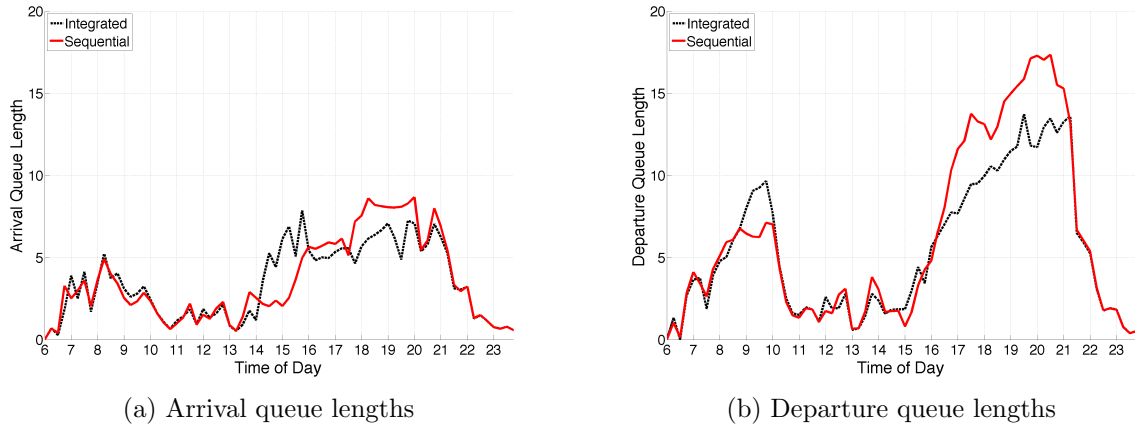


Figure 5-7: Expected queue lengths under the integrated and sequential approaches

5.5 Conclusion

We have developed an original integrated approach to airport congestion mitigation that jointly optimizes the rescheduling of flights through scheduling interventions at the strategic level and the utilization of airport capacity at the tactical level. We have introduced and implemented an *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that integrates (i) a Stochastic Queuing Model of airport congestion, (ii) a Dynamic Programming model of capacity utilization, and (iii) an Integer Programming model of scheduling interventions. We have developed an iterative solution algorithm to this model that was shown to converge after a modest number of iterations and, given the model’s strategic nature, in reasonable computational times.

The application of the model to JFK Airport has yielded two main observations. First, our computational tests suggest that even a moderate level of rescheduling, combined with optimal operating procedures, can mitigate substantially congestion at busy US airports. Peak arrival and departure delays at JFK could be reduced by 20%-40% and by 40%-60%, respectively, without eliminating any flight, without modifying the scheduled time of 75%-90% of the flights scheduled at that airport and without shifting the scheduled time of any flight by more than 15 or 30 minutes. In addition, the proposed schedule maintains all aircraft and passenger connections. Second, our integrated approach has been shown to

provide significant benefits, compared to a typical sequential approach where scheduling and operational decisions are made successively.

These results suggest that the integrated approach presented in this thesis can provide better solutions to the trade-off between schedule displacement and flight delays and might thus improve significantly airport demand management practices worldwide. At airports operating without scheduling limits, such as the great majority of US airports, the approach can be used to estimate the potential delay reductions that could be achieved through limited interference with airline competitive scheduling. At schedule-facilitated, “Level 2” airports, it optimizes the rescheduling of flights, taking into account the impact of such rescheduling on airport operations. It also quantifies the delay reductions that can be achieved through such rescheduling. And, if applied at schedule-coordinated, “Level 3” airports, the approach provides a flexible and technically advanced way for allocating scarce airport capacity and meeting on-time performance objectives without relying on the notion of a flat or nearly-flat, administratively determined “declared capacity”.

Based on these positive results, we extend this modeling approach in the next chapter to account for two additional objectives in the design of scheduling interventions: inter-airline equity and airline collaboration.

Chapter 6

Inter-airline Equity and Collaboration in Scheduling Interventions

This chapter proposes, implements, and assesses mechanisms for allocating airport capacity to the airlines to limit the extent of overscheduling at peak hours, while minimizing interference with airline competitive scheduling. These mechanisms involve the airlines providing scheduling inputs to a central decision-maker (e.g., administratively appointed schedule coordinators at slot-controlled airports, the Federal Aviation Administration in the United States), who then produces a modified schedule of flights to reduce anticipated delays at a considered airports. The approach developed here builds upon the ICUSM developed in Chapter 5, but it adds two important considerations. First, it incorporates inter-airline equity objectives in the optimization of scheduling interventions in order to balance scheduling adjustments equitably among the airlines. Second, it designs non-monetary mechanisms that enable the airlines to provide their scheduling preferences and priorities, and it integrates them into the optimization of scheduling interventions.

The contribution of this chapter falls into four categories:

- We define efficiency, inter-airline equity and on-time performance to characterize the performance of scheduling interventions mechanisms and we develop a tractable lexicographic architecture to characterize the trade space between these three objectives.

- We provide a theoretical discussion of the conditions under efficiency and equity can be jointly optimized, or not.
- We propose non-monetary, credit-based mechanisms for airport scheduling interventions enabling the airlines to provide their scheduling preferences.
- We show from computational results that inter-airline equity can be achieved in airport scheduling interventions at no (or minimal) efficiency losses and that enabling airline collaboration improves the outcome of scheduling interventions.

A paper based on this chapter can be found in [64].

The remainder of this chapter is organized as follows. In Section 6.1, we motivate the research reported here by discussing the limitations of the ICUSM and reviewing the related literature. In Section 6.2, we formulate a model of airport scheduling interventions that builds upon the ICUSM to account for inter-airline equity and airline scheduling preferences. Section 6.3 discusses the conditions under which efficiency and inter-airline equity can be jointly maximized. In Section 6.4, we propose non-monetary, credit-based mechanisms that define the inputs provided by the airlines and the subsequent scheduling interventions. In Section 6.5, we apply the models and mechanisms to JFK. Section 6.6 concludes.

6.1 Motivation

The ICUSM developed in Chapter 5 provides a modeling framework for optimizing congestion-mitigating scheduling interventions, while accounting for the way airport capacity utilization procedures can be modified as a result of the scheduling changes. Results suggested that significant delay reductions could be achieved through limited scheduling interventions. However, it suffers from the following three limitations:

1. The ICUSM does not account for inter-airline equity considerations. Its two-stage lexicographic formulation (Equation (5.1)) ensures equity at the flight level, i.e., no flight is disproportionately displaced. However, it does *not* ensure equity at the airline level. In turn, its solution may penalize some airlines disproportionately.

2. The ICUSM assumes perfect information about the connections that need to be maintained. However, aircraft, passenger, and crew connections are not known with perfect accuracy by the central decision-maker (or even, in some cases, by the airlines themselves) at the time of the scheduling interventions.
3. The ICUSM assumes that all flights are equally inconvenient to reschedule. Whereas this “a flight is a flight” paradigm is standard in the airline industry, it may not be consistent with airlines’ scheduling preferences and with underlying passenger demand.

To address these limitations, we develop in this chapter a modeling approach that builds upon the ICUSM to optimize scheduling interventions, but accounts for inter-airline equity and airlines’ preferences regarding *which* flights to reschedule (in order to maintain network connections or to reschedule the more flexible flights). These topics have been the subject of recent research. First, the trade-off between efficiency (i.e., maximizing the sum of agents’ utilities) and equity (i.e., balancing utilities fairly among the agents) in resource allocation has been formalized by Bertsimas et al. [16, 17], who obtained theoretical bounds on the “price of fairness” and the “price of efficiency”, i.e., the relative loss in efficiency if fairness is maximized, and conversely. Their results suggest that, under certain conditions, significant equity gains can be achieved at limited efficiency losses, which motivates the consideration of inter-airline equity in airport scheduling interventions. Second, market-based mechanisms based on congestion pricing [28, 32, 24] or slot auctions [99, 4, 55] have been proposed for airport demand management to address the limitations of the administrative slot control policies by enabling the airlines to provide their scheduling preferences. However, they have not been successfully implemented in the current institutional environment. An important challenge thus lies in the design of non-monetary mechanisms enabling the airlines to reveal their scheduling preferences.

In a related area, non-monetary, distributed, and equitable mechanisms have been developed for allocating air transportation capacity at the tactical level through Air Traffic Flow Management (ATFM) initiatives. Whereas early ATFM developments were centralized, and based on efficiency objectives exclusively (minimizing total congestion costs), their successful

implementation was made possible by involving the airlines in the decision-making process. The Collaborative Decision Making (CDM) paradigm decentralized operating decisions to the airlines whenever possible, through airline collaboration and slot trading, to determine which flights to delay and which flights to prioritize [118]. Moreover, recent studies incorporated inter-airline equity objectives into ATFM models to make the outcome of centralized decision-making more acceptable to each individual airline [9, 19]. This chapter integrates these two objectives into the design and optimization of scheduling interventions.

The problem of airport scheduling interventions exhibits several differences from the ATFM problem. First, the information available at the time of flight scheduling is more limited than that available on the day of operations. Second, unlike in ATFM, no standard of equity has been accepted in the industry with respect to scheduling interventions. Third, scheduling interventions may result in flights being rescheduled later or earlier than their preferred scheduled times requested by the airlines. This contrasts with the situation in ATFM where the flights can only be delayed but cannot be moved earlier than their scheduled time. Thus, the ATFM schemes of ration-by-schedule and schedule compression do not have any direct analogs in the context of scheduling interventions. It is thus necessary to propose new metrics of inter-airline equity, to design new mechanisms for revealing airline scheduling preferences, and to develop new modeling frameworks for scheduling interventions. This is the focus of this chapter.

6.2 Multi-criteria Modeling Architecture

We present first our modeling approach to airport scheduling interventions that incorporates inter-airline equity considerations and enables airline collaboration. The modeling structure, the types of decisions that are made, and the scheduling and network connectivity constraints are identical to those in the ICUSM, but the main difference lies in the objectives of scheduling interventions.

In addition to the notations from Chapter 5, we partition the set of flights scheduled at the airport, i.e., $\mathcal{F}^{\text{arr}} \cup \mathcal{F}^{\text{dep}}$, into subsets of flights scheduled by the different airlines, and

the set of connections \mathcal{C} into sets of flight pairs from different airlines. With the notations introduced below, we have: $\mathcal{F}_{a_1} \cap \mathcal{F}_{a_2} = \emptyset, \forall a_1 \neq a_2 \in \mathcal{A}$ and $\cup_{a \in \mathcal{A}} \mathcal{F}_a = \mathcal{F}^{\text{arr}} \cup \mathcal{F}^{\text{dep}}$, and $\mathcal{C}_{a_1} \cap \mathcal{C}_{a_2} = \emptyset, \forall a_1 \neq a_2 \in \mathcal{A}$ and $\cup_{a \in \mathcal{A}} \mathcal{C}_a = \mathcal{C}$. Note that this assumes that all connections are intra-airline connections. We also introduce a parameter v to characterize “flight valuations”, reflecting airlines’ preferences. Flights with lower values can be thought of as less “costly” flights to reschedule, or as the flights that exhibit larger timetabling flexibility.

- \mathcal{A} = the set of airlines, $\{1, \dots, A\}$
- \mathcal{F}_a = the set of flights scheduled by airline a at the considered airport
- \mathcal{C}_a = the set of connections between flights scheduled by airline a
- v_i = the valuation of flight i

In this section, we assume that connections data (\mathcal{C} , t^{\min} , and t^{\max}) and flight valuations data (v) are known by the central decision-maker. We develop performance attributes for scheduling interventions and a modeling architecture to characterize the associated trade space. In Section 6.4, we relax the assumption of the full knowledge of connections and flight valuations data by the central decision-maker. Instead, we design non-monetary mechanisms that formalize the process through which the airlines can provide their scheduling inputs.

6.2.1 Performance Attributes

We consider the following three performance attributes of scheduling interventions: efficiency, inter-airline equity, and on-time performance. Efficiency and on-time performance extend the notions of schedule displacement and expected queue lengths limits, respectively, that are considered in the ICUSM, while the notion of equity is added to this framework.

Efficiency This refers to the mechanism’s ability to meet airline scheduling preferences. Similarly to Chapter 5, efficiency is measured by the displacement from the schedule of flights provided by the airlines. We consider the following two efficiency objectives. First, we define *min-max efficiency* as the largest flight displacement δ . Second, we define *weighted*

efficiency as the weighted total schedule displacement, denoted by Δ (which generalizes the total displacement considered in the ICUSM in a way that accounts for flight valuations). Directionally, maximizing efficiency involves minimizing δ and/or Δ .

$$\delta = \max_{i \in \mathcal{F}} |u_i| \implies \min \delta \quad (6.1)$$

$$\Delta = \sum_{i \in \mathcal{F}} v_i |u_i| \implies \min \Delta \quad (6.2)$$

Inter-airline Equity This refers to the mechanism’s ability to balance schedule displacement fairly among the airlines. We describe each airline’s disutility by its average weighted per-flight displacement, denoted by σ_a . Perfect equity is achieved when the weighted per-flight displacement borne by any airline is proportional to its number of flights scheduled. In order to maximize inter-airline equity, we minimize airline disutilities lexicographically, i.e., we first minimize the largest per-flight displacement borne by any airline, then the second-largest, etc. This extends the min-max equity formulation [16, 17], and has been applied in other resource allocation problems [68, 88, 112].

$$\sigma_a = \frac{1}{|\mathcal{F}_a|} \sum_{i \in \mathcal{F}_a} v_i |u_i|, \forall a \in \mathcal{A} \implies \text{lex min } \sigma \quad (6.3)$$

We quantify inequity by the largest airline disutility, denoted by Φ :

$$\Phi = \max_{a \in \mathcal{A}} \sigma_a \quad (6.4)$$

On-time performance This refers to the mechanism’s ability to mitigate airport congestion. We quantify on-time performance by a non-decreasing function of the arrival and departure queue lengths A_t and D_t (which depend on the schedule of flights according to Equation (5.14)), denoted by $g(A_1, \dots, A_T, D_1, \dots, D_T)$. Examples of such functions could quantify the peak expected arrival and departure queue lengths, the total delay experienced over a day of operations, the 95th percentile of the peak expected arrival and departure queue

lengths, etc. Maximizing on-time performance involves minimizing the function g .

$$\min \{g(A_1, \dots, A_T, D_1, \dots, D_T)\} \quad (6.5)$$

The optimization of scheduling interventions is a multi-objective optimization problem. First, each of these three performance attributes comprises several dimensions (e.g., minimizing min-max efficiency vs. weighted efficiency; minimizing the largest airline disutility vs. variations in airlines' utilities for equity; minimizing arrival vs. departure delays for on-time performance). Moreover, there exists a trade-off between efficiency and on-time performance, quantified by the ICUSM: the larger the schedule displacement, the larger the potential delay reductions (up to a limit). Finally, there may be, for given on-time performance objectives, a trade-off between efficiency and equity [16, 17].

6.2.2 Lexicographic Modeling Approach

We characterize the trade space between efficiency, equity, and on-time performance in airport scheduling interventions. In order to provide a transparent and optimal characterization of this trade space, we aim to find its Pareto frontier, i.e., the set of solutions such that no other solution could simultaneously improve the three objectives. This representation of the trade space is flexible enough to be used by system managers and policy makers to select the most appropriate level of compromise between these conflicting objectives. To this end, we develop a lexicographic optimization approach that (i) fixes on-time performance targets; (ii) maximizes efficiency under on-time performance targets; and (iii) maximizes equity under on-time performance and efficiency targets.

First, we quantify on-time performance by the peak expected arrival and departure delays, i.e. $g(A_1, \dots, A_T, D_1, \dots, D_T) = (\max_{t \in \mathcal{T}} E(A_t), \max_{t \in \mathcal{T}} E(D_t))$. This function is motivated by the objective of controlling the largest delays experienced over the day of operations, and is consistent with the on-time performance constraints from the ICUSM (Constraints (5.15) and (5.16)). We still denote peak expected arrival and departure queue lengths targets by

A_{MAX} and D_{MAX} , respectively. We then aim to find the “best” schedule that meets these on-time performance targets.

Second, we determine the schedule of flights that maximizes efficiency, subject to scheduling constraints, network connectivity constraints, and on-time performance constraints. As in Chapter 5, we formulate the efficiency-maximizing problem by lexicographically maximizing, first, min-max efficiency δ , and, second, weighted efficiency Δ . This is expressed in Problems **P1** and **P2** described below:

P1 We minimize min-max efficiency δ , subject to scheduling, network connectivity and on-time performance constraints. We denote by δ^* its optimal value.

$$\begin{aligned} \min \quad & \delta \text{ (Equation (6.1))} \\ \text{s.t.} \quad & \text{Scheduling and network connectivity constraints: (5.4) to (5.13)} \\ & \text{On-time performance constraints: (5.14) to (5.16)} \end{aligned}$$

P2 We minimize weighted efficiency Δ , subject to scheduling, network connectivity and on-time performance constraints, and subject to the constraint that no flight may be displaced by more than δ^* . We denote by Δ^* its optimal value.

$$\begin{aligned} \min \quad & \Delta \text{ (Equation (6.2))} \\ \text{s.t.} \quad & \text{Scheduling and network connectivity constraints: (5.4) to (5.13)} \\ & \text{On-time performance constraints: (5.14) to (5.16)} \\ & \text{Min-max efficiency objectives: } |u_i| \leq \delta^*, \forall i \in \mathcal{F} \end{aligned}$$

Third, we maximize inter-airline equity, subject to scheduling constraints, network connectivity constraints, on-time performance constraints, and efficiency targets. This is formulated in the class of problems **P3**(ρ) described below:

P3(ρ) We fix efficiency targets, and we lexicographically minimize airline disutilities, subject to scheduling, network connectivity, on-time performance, and efficiency constraints.

We characterize the trade space between efficiency and equity by varying the efficiency target. Specifically, we impose that min-max efficiency must be optimal (i.e., no flight may be rescheduled by more than δ^*) and we denote by ρ the relative loss in weighted efficiency that is allowed (i.e., we constrain the weighted displacement to be at most equal to $(1 + \rho)\Delta^*$). When $\rho = \infty$, we only maximize equity (without any weighted efficiency consideration). When $\rho = 0$, we maximize equity, under optimal efficiency.

$$\begin{aligned}
& \text{lex min } \sigma \text{ (Equation (6.3))} \\
& \text{s.t. Scheduling and network connectivity constraints: (5.4) to (5.13)} \\
& \text{On-time performance constraints: (5.14) to (5.16)} \\
& \text{Min-max efficiency objectives: } \delta \leq \delta^*, \forall i \in \mathcal{F} \\
& \text{Weighted efficiency objectives: } \Delta \leq (1 + \rho) \Delta^*
\end{aligned}$$

Problems **P1**, **P2**, and **P3**(ρ) combined determine the Pareto frontier of the trade space between efficiency, equity, and on-time performance in scheduling interventions. First, variations in the on-time performance targets A_{MAX} and D_{MAX} quantify the trade-off between the costs of scheduling interventions (in terms of inefficiency and inequity) and delay reductions. Second, for any on-time performance targets A_{MAX} and D_{MAX} , varying the parameter ρ quantifies the potential trade-off between weighted efficiency and inter-airline equity.

Figure 6-1 illustrates our approach to maximizing weighted efficiency and inter-airline equity, for given on-time performance targets A_{MAX} and D_{MAX} , and the optimal value of min-max efficiency δ^* . Specifically, it shows hypothetical variations in three airlines' disutilities (σ_1 , σ_2 , and σ_3) as a function of the weighted efficiency target $\Delta = (1 + \rho)\Delta^*$. By construction, the region on the left side of Δ^* is infeasible, i.e., the schedule displacement has to be at least equal to Δ^* to ensure that the problem is feasible. Moreover, the largest airline disutility Φ is a non-increasing function of the value of weighted efficiency Δ (i.e., of ρ). Note that the other airlines' utilities (here, σ_2 and σ_3) may increase or decrease as Φ is reduced. As the largest airline disutility Φ attains its optimal value, the second-largest disutility may still be larger than its optimal value. In this case, further increases in ρ may yield further improvements

in the lexicographic minimization of airline disutilities. Optimal equity is attained when the largest, second largest, third largest, etc. airline disutilities have all reached their optimal values (i.e., the values that would be obtained without any efficiency consideration, or with $\rho = \infty$). Note that Figure 6-1 shows an instance where the order of airline disutilities remains identical for all values of ρ (i.e., in this case, $\sigma_1(\rho) > \sigma_2(\rho) > \sigma_3(\rho), \forall \rho \geq 0$), but this need not be the case (i.e., the curves may intersect). We denote by $\sigma^*(\rho)$ the equity-maximizing vector of airline per-flight displacement, as a function of ρ , and by $\Phi^*(\rho) = \max_{a \in \mathcal{A}} \sigma_a^*(\rho)$. We denote by Δ^{eq} the smallest equity-maximizing value of Δ , and by ρ^* the minimum loss in weighted efficiency required to attain optimal equity (i.e., $\Delta^{\text{eq}} = (1 + \rho^*)\Delta^*$). With these notations, the “price of efficiency” and the “price of equity” will be characterized by $P_{\text{eff}} = \frac{\Phi^*(0) - \Phi^*(\infty)}{\Phi^*(\infty)}$, and by $P_{\text{eq}} = \frac{\Delta^{\text{eq}} - \Delta^*}{\Delta^*} = \rho^*$, respectively.

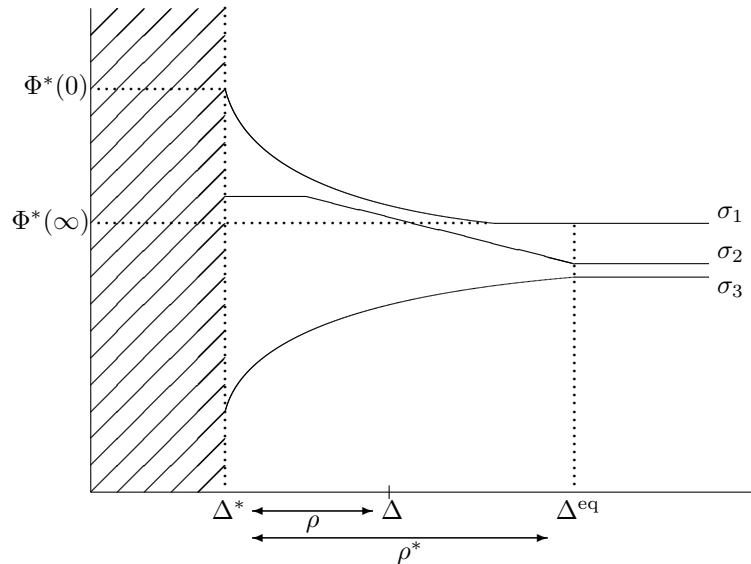


Figure 6-1: A schematic trade space between weighted efficiency and equity

6.2.3 Solution Architecture

As the on-time performance constraints are not linear, solving Problems **P1**, **P2**, and **P3**(ρ) requires a solution algorithm such as the ones developed in Chapter 5. To solve Problem **P1**, we apply directly Algorithm 3. To solve Problem **P2**, we use an algorithm very similar to Algorithm 4, i.e., we iteratively update an upper bound $\bar{\Delta}$ and a lower bound $\underline{\Delta}$ of

the total displacement Δ . The only difference is the stopping criteria. Since the weighted displacement Δ is continuous (while the total displacement considered in Chapter 5 was integral), the algorithm terminates when the following condition is satisfied: $\frac{\bar{\Delta} + \Delta}{2} \leq \varepsilon_{\Delta}$. This ensures that the weighted displacement obtained is within ε_{Δ} of its optimal value. We use a value of ε_{Δ} of 1%.

When solving the equity-maximizing problem (Problem **P3**(ρ)), one candidate approach is to design a similar iterative algorithm. This would consist of iteratively updating upper and lower bounds of the optimal value of airlines' disutilities σ_a until convergence. However, the computational requirements of the solution algorithms developed in Chapter 5 prevent them from being applied repeatedly for each airline, for several values of the parameter ρ , and with different sets of inputs. For this reason, we develop here an alternative approach that approximates Problem **P3**(ρ) while ensuring computational tractability.

Specifically, we replace the on-time performance constraints (Constraints (5.15) and (5.16)) by scheduling limit constraints (Constraints (6.6) and (6.7), defined below). These constraints ensure that, for any period t , the number of scheduled arrivals and departures does not exceed limits denoted by $\widehat{\lambda}_t^{\text{arr}}$ and $\widehat{\lambda}_t^{\text{dep}}$, respectively. We refer to these constraints as "simplified on-time performance constraints".

$$\lambda_t^{\text{arr}} \leq \widehat{\lambda}_t^{\text{arr}}, \forall t \in \mathcal{T} \quad (6.6)$$

$$\lambda_t^{\text{dep}} \leq \widehat{\lambda}_t^{\text{dep}}, \forall t \in \mathcal{T} \quad (6.7)$$

The resulting model is formulated below, and we refer to it as $\widehat{\mathbf{P3}}(\rho)$:

$$\begin{aligned} \text{lex min} \quad & \sigma \text{ (Equation (6.3))} \\ \text{s.t.} \quad & \text{Scheduling and network connectivity constraints: (5.4) to (5.13)} \\ & \text{Simplified on-time performance constraints: (6.6) and (6.7)} \\ & \text{Min-max efficiency objectives: } \delta \leq \delta^*, \forall i \in \mathcal{F} \\ & \text{Weighted efficiency objectives: } \Delta \leq (1 + \rho) \Delta^* \end{aligned}$$

Unlike Problem $\mathbf{P3}(\rho)$, Problem $\widehat{\mathbf{P3}}(\rho)$ is an Integer Programming model and can be solved directly using a commercial solver. Instead of iterating 10-15 times between an Integer Program, a Dynamic Program, and a Stochastic Queuing Model, it involves solving only one Integer Program. Therefore, its solution is substantially faster than that of Problem $\mathbf{P3}(\rho)$. The main challenge lies in setting appropriate values of the scheduling limits $\widehat{\lambda}_t^{\text{arr}}$ and $\widehat{\lambda}_t^{\text{dep}}$. If set too high, they would not enable the resulting arrival and departure queue lengths to meet their targets A_{MAX} and D_{MAX} . If set too low, they may not minimize the impact on airline schedules of flights.

We select scheduling limits $\widehat{\lambda}_t^{\text{arr}}$ and $\widehat{\lambda}_t^{\text{dep}}$ equal to the aggregate number of scheduled arrivals and departures obtained by solving Problems $\mathbf{P1}$ and $\mathbf{P2}$. In other words, we first determine the efficiency-maximizing schedule of flights. We then look for flight schedules that achieve the same aggregate schedule (but not necessarily the same schedule for each individual flight), while yielding a Pareto-optimal solution to the trade-off between weighted efficiency and equity. By construction, the schedule thus obtained meets the delay reduction constraints (5.15) and (5.16). Moreover, the aggregate schedule is determined through a procedure that starts with the scheduling inputs provided by the airlines, and can thus exhibit some peaks and valleys in accordance with airline scheduling preferences and passenger demand. On the other hand, since Constraints (6.6) and (6.7) could be more restrictive than Constraints (5.15) and (5.16), this solution approach may yield a sub-optimal solution. We expect nonetheless these errors to be of second order. In any case, this approach yields improvements in inter-airline equity, as compared to the outcome of the ICUSM.

Our full solution architecture is shown in Figure 6-2. It takes as inputs scheduling data, connections data, and flight valuation data, as well as on-time performance targets A_{MAX} and D_{MAX} set by the central decision-maker. First, we solve successively Problems $\mathbf{P1}$ and $\mathbf{P2}$ using the iterative algorithm described in Figure 5-1, and we store the optimal efficiency values (δ^* and Δ^*) and the aggregate schedule ($\widehat{\lambda}_t^{\text{arr}}$ and $\widehat{\lambda}_t^{\text{dep}}$). Second, we solve Problems $\widehat{\mathbf{P3}}(\rho)$ to determine the Pareto frontier of the trade space between weighted efficiency and equity to achieve this aggregate schedule (i.e., $\widehat{\lambda}_t^{\text{arr}}$ and $\widehat{\lambda}_t^{\text{dep}}$). We start by maximizing equity

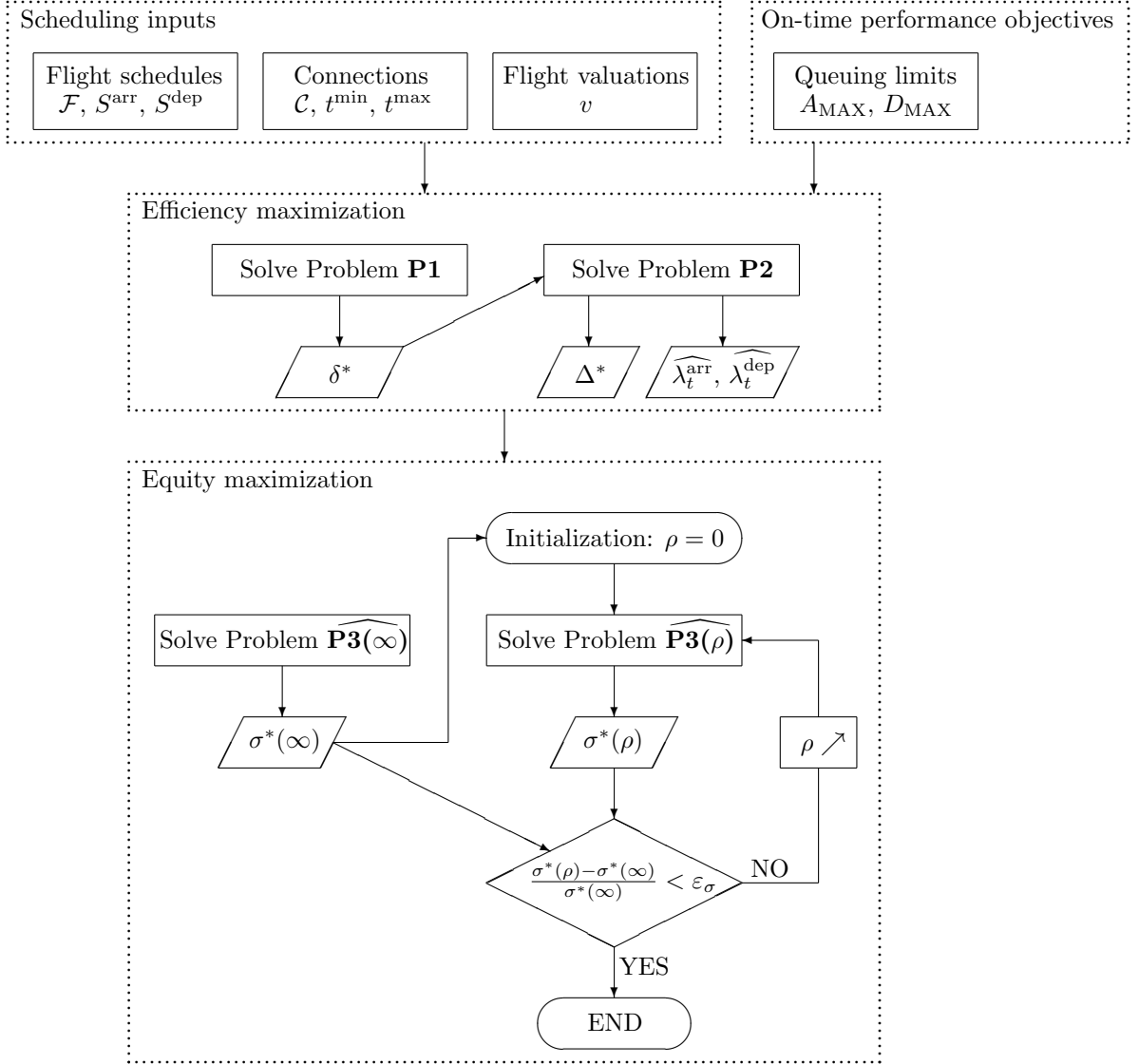


Figure 6-2: Solution architecture

with no efficiency constraint ($\rho = \infty$). We then maximize equity under optimal efficiency ($\rho = 0$), and we relax progressively the weighted efficiency requirements by increasing ρ by increments of 0.001, until optimal equity is reached. We use the following stopping criteria: $\frac{\sigma_a^*(\rho) - \sigma_a^*(\infty)}{\sigma_a^*(\infty)} \leq \varepsilon_\sigma, \forall a \in \mathcal{A}$, i.e., the algorithm terminates when all airlines' disutilities are within ε_σ of their equity-maximizing value. We use here a value of ε_σ of 1%. This algorithm characterizes, for any pair of on-time performance targets A_{MAX} and D_{MAX} , (i) the efficiency-maximizing schedule of flights, and (ii) feasible flight schedules that achieve the

same aggregate airport schedule and yield different Pareto-optimal solutions to the trade-off between weighted efficiency and equity.

As mentioned before, we assumed thus far that the connections and flight valuations data are known to the central decision maker with accuracy. In reality, that is not the case. We thus need to design mechanisms to enable the airlines to provide the scheduling inputs (i.e., scheduling, connections, and flight valuations data). We do this in Section 6.4. In the next section, we discuss the trade-off between efficiency and equity.

6.3 A Theoretical Discussion on Inter-airline Equity

In this section, we first show that efficiency and equity can be jointly optimized in a simple case where (i) no network connections are requested, (ii) all flights are equally valued, and (iii) airline schedules of flights exhibit some regularity. We then discuss the factors that may violate these conditions, and thus create a trade-off between efficiency and equity in scheduling interventions.

6.3.1 Cases of Joint Maximization of Efficiency and Equity

We first identify instances where efficiency and equity can be jointly maximized. To this end, we focus on cases where there is no connection between any pair of flights (i.e., $\mathcal{C} = \emptyset$) and where all flights are equally valued (i.e., $v_i = 1, \forall i \in \mathcal{F}$). In the case of $\mathcal{C} = \emptyset$, the problems of rescheduling arriving flights and departing flights are decoupled, and can thus be treated separately. For simplicity of the exposition, we focus on the case of $\delta^* = 1$ period, which is also the most common case encountered with real-world data (see Section 6.5), but the insights can be extended to the case where $\delta^* > 1$. Finally, we restrict the equity problem to the problem of minimizing the largest airline disutility, but the discussion can be easily extended to the problem of lexicographically minimizing airline disutilities (by considering each airline disutility one at a time). We consider the following efficiency-maximizing (EFF) and equity-maximizing (EQ) problems defined below.

$$\begin{array}{ll}
\min & \sum_{i \in \mathcal{F}} |u_i| \quad (\text{EFF}) \\
\text{s.t.} & w_{it} \geq w_{i,t+1}, \forall i \in \mathcal{F}, \forall t \in \mathcal{T} \\
& w_{i1} = 1, \forall i \in \mathcal{F} \\
& \sum_{t \in \mathcal{T}} (w_{it} - S_{it}) = u_i, \forall i \in \mathcal{F} \\
& \sum_{i \in \mathcal{F}} (w_{it} - w_{i,t+1}) \leq \widehat{\lambda}_t, \forall t \in \mathcal{T} \\
& |u_i| \leq 1, \forall i \in \mathcal{F}
\end{array}
\quad
\begin{array}{ll}
\min & \max_{a \in \mathcal{A}} \left(\frac{1}{|\mathcal{F}_a|} \sum_{i \in \mathcal{F}_a} |u_i| \right) \quad (\text{EQ}) \\
\text{s.t.} & w_{it} \geq w_{i,t+1}, \forall i \in \mathcal{F}, \forall t \in \mathcal{T} \\
& w_{i1} = 1, \forall i \in \mathcal{F} \\
& \sum_{t \in \mathcal{T}} (w_{it} - S_{it}) = u_i, \forall i \in \mathcal{F} \\
& \sum_{i \in \mathcal{F}} (w_{it} - w_{i,t+1}) \leq \widehat{\lambda}_t, \forall t \in \mathcal{T} \\
& |u_i| \leq 1, \forall i \in \mathcal{F}
\end{array}$$

We identify two conditions under which efficiency and equity can be jointly maximized: (i) if the scheduling interventions that take place during different time periods are independent (Proposition 1), or (ii) if each airline's share of flights is identical across all periods (Proposition 2), or both. We denote by \mathcal{D}_t the set of flights scheduled during period t before the scheduling interventions, i.e., $\mathcal{D}_t = \{i \in \mathcal{F} \mid S_{it} = 1 \& S_{i,t+1} = 0\}$. By convention, we have $\mathcal{D}_0 = \mathcal{D}_{T+1} = \emptyset$ and $\widehat{\lambda}_0 = \widehat{\lambda}_{T+1} = 0$.

In Proposition 1, we study the case where the scheduling interventions that take place during different time periods are independent. Specifically, we assume that the number of flights to reschedule at any period t is lower than the number of flights that can be added to the schedule at periods $t-1$ and $t+1$. We show that efficiency and equity can be jointly maximized in that case.

Proposition 1. *If $(|\mathcal{D}_t| - \widehat{\lambda}_t)^+ \leq (\widehat{\lambda}_{t-1} - |\mathcal{D}_{t-1}|)^+ + (\widehat{\lambda}_{t+1} - |\mathcal{D}_{t+1}|)^+, \forall t \in \mathcal{T}$, then there exists a solution that solves (EFF) and (EQ).*

Proof: Any feasible solution has to displace at least $(|\mathcal{D}_t| - \widehat{\lambda}_t)^+$ flights per period, so $\Delta^* \geq \sum_{t \in \mathcal{T}} (|\mathcal{D}_t| - \widehat{\lambda}_t)^+$. We first construct a feasible solution that reschedules exactly $\sum_{t \in \mathcal{T}} (|\mathcal{D}_t| - \widehat{\lambda}_t)^+$ flights. To do so, for any period t such that $|\mathcal{D}_t| > \widehat{\lambda}_t$, we reschedule flights first to the preceding period, up to capacity, and then to the following period. Specifically, we select a subset $\mathcal{K}_t^- \subset \mathcal{D}_t$ such that $|\mathcal{K}_t^-| = \min \left\{ (\widehat{\lambda}_{t-1} - |\mathcal{D}_{t-1}|)^+, (|\mathcal{D}_t| - \widehat{\lambda}_t)^+ \right\}$. Note that \mathcal{K}_t^- is not uniquely determined, but we can choose any subset of \mathcal{D}_t that verifies this

property. We then select a subset $\mathcal{K}_t^+ \subset \mathcal{D}_t \setminus \mathcal{K}_t^-$ such that $|\mathcal{K}_t^+| = \left(|\mathcal{D}_t| - \widehat{\lambda}_t - |\mathcal{K}_t^-|\right)^+$. As $\mathcal{K}_t^-, \mathcal{K}_t^+$ is not uniquely determined. Under the assumption of the proposition, we have: $|\mathcal{K}_t^+| \leq \left(\widehat{\lambda}_{t+1} - |\mathcal{D}_{t+1}|\right)^+$. We can therefore define u^{eff} as follows: $u_i^{\text{eff}} = -1, \forall i \in \mathcal{K}_t^-$ and $u_i^{\text{eff}} = +1, \forall i \in \mathcal{K}_t^+$. We define w^{eff} accordingly (based on the constraints of (EFF)). The solution $(w^{\text{eff}}, u^{\text{eff}})$ is a feasible solution that verifies $\sum_{i \in \mathcal{F}} |u_i^{\text{eff}}| = \sum_{t \in \mathcal{T}} \left(|\mathcal{D}_t| - \widehat{\lambda}_t\right)^+$. Therefore, it solves (EFF) and $\Delta^* = \sum_{t \in \mathcal{T}} \left(|\mathcal{D}_t| - \widehat{\lambda}_t\right)^+$.

Let's now denote by $(w^{\text{eq}}, u^{\text{eq}})$ an optimal solution of (EQ). If $\sum_{i \in \mathcal{F}} |u_i^{\text{eq}}| = \Delta^*$, then u^{eq} also solves (EFF). We now assume that $\sum_{i \in \mathcal{F}} |u_i^{\text{eq}}| > \Delta^*$. For each $t \in \mathcal{T}$, we define the following set: $\mathcal{I}_t = \{i \in \mathcal{D}_t \mid |u_i^{\text{eq}}| = 1\}$. We have $|\mathcal{I}_t| \geq \left(|\mathcal{D}_t| - \widehat{\lambda}_t\right)^+, \forall t \in \mathcal{T}$ (otherwise, u^{eq} would not be a feasible solution of (EQ)). We can construct a set $\mathcal{J}_t \subseteq \mathcal{I}_t$ such that $|\mathcal{J}_t| = \left(|\mathcal{D}_t| - \widehat{\lambda}_t\right)^+$ for all $t \in \mathcal{T}$. As \mathcal{K}_t^- and \mathcal{K}_t^+ earlier, \mathcal{J}_t is not uniquely determined, but we can choose any subset of \mathcal{I}_t that verifies this property. Let \mathcal{J} be defined by $\mathcal{J} = \cup_{t \in \mathcal{T}} \mathcal{J}_t$. We construct a solution u^* as follows: $u_i^* = 0, \forall i \notin \mathcal{J}$ and $u_i^* = u_i^{\text{eq}}, \forall i \in \mathcal{J}$. We define w^* accordingly. By construction, for each period t such that $|\mathcal{D}_t| > \widehat{\lambda}_t$, this solution displaces exactly $|\mathcal{D}_t| - \widehat{\lambda}_t$ flights, so $\sum_{i \in \mathcal{F}} (w_{it}^* - w_{i,t+1}^*) = \widehat{\lambda}_t$. Moreover, the number of flights rescheduled to the preceding and following time periods is smaller than under solution $(w^{\text{eq}}, u^{\text{eq}})$ (i.e., $\sum_{i \in \mathcal{F}} (w_{i,t-1}^* - w_{i,t}^*) \leq \sum_{i \in \mathcal{F}} (w_{i,t-1}^{\text{eq}} - w_{i,t}^{\text{eq}})$ and $\sum_{i \in \mathcal{F}} (w_{i,t+1}^* - w_{i,t}^*) \leq \sum_{i \in \mathcal{F}} (w_{i,t+2}^{\text{eq}} - w_{i,t}^{\text{eq}})$). Therefore, (w^*, u^*) is a feasible solution of (EFF) and (EQ). Moreover, it verifies: $\sum_{i \in \mathcal{F}} |u_i^*| = \Delta^*$, and $|u_i^*| \leq |u_i^{\text{eq}}|, \forall i \in \mathcal{F}$. Therefore, u^* solves (EFF) and (EQ). \square

Before moving to the second instance where efficiency and equity can be jointly maximized, we characterize the optimal solution of the following problem. We refer to this problem as Problem $\mathcal{P}(\Delta)$, where Δ designates any non-negative integer. It involves allocating a “budget” of Δ items across “groups” (or airlines, in our case) indexed by $a \in \mathcal{A}$ in

a way that minimizes the largest weighted cost borne by any group.

$$\begin{aligned}
\min \quad & \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|} \\
\text{s.t.} \quad & \sum_{a \in \mathcal{A}} U_a \geq \Delta \\
& U_a \geq 0, U_a \text{ integer}
\end{aligned}$$

Lemma 1 shows that there is a solution of $\mathcal{P}(\Delta)$ that binds the constraint $\sum_{a \in \mathcal{A}} U_a \geq \Delta$.

Lemma 1. *There exists an optimal solution $(U_a)_{a \in \mathcal{A}}$ of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a = \Delta$*

Proof: Let $(U_a^0)_{a \in \mathcal{A}}$ denote an optimal solution of $\mathcal{P}(\Delta)$ and let us assume that $\sum_{a \in \mathcal{A}} U_a^0 > \Delta$. We denote by ε the integer defined by $\varepsilon = \sum_{a \in \mathcal{A}} U_a^0 - \Delta > 0$. We introduce a vector $(\eta_a)_{a \in \mathcal{A}}$ such that η_a is non-negative and integer for all $a \in \mathcal{A}$ and that verifies $\sum_{a \in \mathcal{A}} \eta_a = \varepsilon$ (Note that the vector $(\eta_a)_{a \in \mathcal{A}}$ thus defined is not unique, but we can choose any vector that verifies these properties). We then define $U_a = U_a^0 - \eta_a, \forall a \in \mathcal{A}$. We have $\sum_{a \in \mathcal{A}} U_a = \Delta$ and $U_a \leq U_a^0, \forall a \in \mathcal{A}$, so U solves $\mathcal{P}(\Delta)$. \square

In Lemma 2, we prove an intermediary result that shows that, if we start with an optimal solution of Problem $\mathcal{P}(\Delta)$ and we add one item to any group, then the resulting weighted cost borne by that group is at least equal to the optimal value of Problem $\mathcal{P}(\Delta)$.

Lemma 2. *If $(U_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta)$, then $\frac{U_c+1}{|\mathcal{F}_c|} \geq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}, \forall c \in \mathcal{A}$*

Proof: Let us consider $c \in \mathcal{A}$. If $U_b = 0, \forall b \neq c$, then $\frac{U_c}{|\mathcal{F}_c|} = \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$, and thus: $\frac{U_c+1}{|\mathcal{F}_c|} > \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. If $\exists b \neq c : U_b > 0$, then we define \bar{U} as follows: $\bar{U}_c = U_c + 1, \bar{U}_b = U_b - 1$ and $\bar{U}_a = U_a, \forall a \neq b, c$. \bar{U} is a feasible solution of $\mathcal{P}(\Delta)$, so $\max_{a \in \mathcal{A}} \frac{\bar{U}_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. Moreover, $\bar{U}_a \leq U_a, \forall a \neq c$, so $\max_{a \neq c} \frac{\bar{U}_a}{|\mathcal{F}_a|} \leq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. In turn, the maximum of $\frac{\bar{U}_a}{|\mathcal{F}_a|}$ is attained in c , that is $\max_{a \in \mathcal{A}} \frac{\bar{U}_a}{|\mathcal{F}_a|} = \frac{\bar{U}_c}{|\mathcal{F}_c|}$. It then comes: $\frac{\bar{U}_c}{|\mathcal{F}_c|} \geq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$, i.e., $\frac{U_c+1}{|\mathcal{F}_c|} \geq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. \square

Lemma 3 then shows that if we start with a solution of Problem $\mathcal{P}(\Delta)$ that binds the constraint $\sum_{a \in \mathcal{A}} U_a \geq \Delta$, then we can construct a solution of Problem $\mathcal{P}(\Delta - 1)$ by removing one item from the group (or one of the groups) that bears the largest weighted cost.

Lemma 3. *If $\Delta \geq 1$ and $(U_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a = \Delta$, then there exists an optimal solution $(U_a^0)_{a \in \mathcal{A}}$ of $\mathcal{P}(\Delta - 1)$ and $a_0 \in \mathcal{A}$ such that: $U_{a_0}^0 = U_{a_0} - 1$ and $U_a^0 = U_a, \forall a \neq a_0$.*

Proof: Let $(U_a)_{a \in \mathcal{A}}$ be an optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a = \Delta$. We choose $a_0 \in \arg \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$, and we define U^0 such that: $U_{a_0}^0 = U_{a_0} - 1$ and $U_a^0 = U_a, \forall a \neq a_0$. We have $\sum_{a \in \mathcal{A}} U_a^0 = \Delta - 1$, so U^0 is a feasible solution of $\mathcal{P}(\Delta - 1)$. Note also that since $U_a^0 \leq U_a, \forall a \in \mathcal{A}$, $\max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$. Let $b \in \mathcal{A}$ be such that $\frac{U_b^0}{|\mathcal{F}_b|} = \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$. Let $(V_a)_{a \in \mathcal{A}}$ be any feasible solution of $\mathcal{P}(\Delta - 1)$.

- If $V_b \geq U_b^0$, then $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \frac{V_b}{|\mathcal{F}_b|} \geq \frac{U_b^0}{|\mathcal{F}_b|} = \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$.
- If $V_b < U_b^0$, then $V_b \leq U_b^0 - 1$, since V_b and U_b^0 are integer. Thus we have: $\exists c : V_c \geq U_c^0 + 1$, as $\sum_{a \in \mathcal{A}} U_a^0 = \Delta - 1$.
 - If $c = a_0$, then $V_{a_0} \geq U_{a_0}$, and $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \frac{V_{a_0}}{|\mathcal{F}_{a_0}|} \geq \frac{U_{a_0}}{|\mathcal{F}_{a_0}|} = \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$, and thus: $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$.
 - If $c \neq a_0$, then $V_c \geq U_c + 1$. Then $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \frac{U_c + 1}{|\mathcal{F}_c|}$ and it follows from Lemma 2 that: $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. It comes: $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$.

Therefore, $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a^0}{|\mathcal{F}_a|}$ and U^0 is an optimal solution of $\mathcal{P}(\Delta - 1)$. \square

Lemma 4 extends Lemma 3 to construct, from a solution of Problem $\mathcal{P}(\Delta)$ that binds the constraint $\sum_{a \in \mathcal{A}} U_a \geq \Delta$, solutions of Problems $\mathcal{P}(0), \dots, \mathcal{P}(\Delta - 1)$ such that each one differs from the following one by only one element.

Lemma 4. *If $\Delta \geq 1$ and $(U_a^\Delta)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a^\Delta = \Delta$, then there exist $(U_a^0)_{a \in \mathcal{A}}, \dots, (U_a^{\Delta-1})_{a \in \mathcal{A}}$ that are optimal solutions of $\mathcal{P}(0), \dots, \mathcal{P}(\Delta - 1)$ and $a_1, \dots, a_\Delta \in \mathcal{A}$ such that: $U_{a_i}^{i-1} = U_{a_i}^i - 1$ and $U_a^{i-1} = U_a^i, \forall a \neq a_i, \forall i = 1, \dots, \Delta$.*

Proof: This is obtained directly by repeatedly applying Lemma 3 to $\mathcal{P}(\Delta), \mathcal{P}(\Delta - 1), \dots, \mathcal{P}(2)$ and $\mathcal{P}(1)$. \square

We now introduce the following notations. We denote by γ the greatest common divisor of $(|\mathcal{F}_a|)_{a \in \mathcal{A}}$, i.e., $\gamma = \gcd(|\mathcal{F}_a|)_{a \in \mathcal{A}}$. We also introduce $\xi_a = \frac{|\mathcal{F}_a|}{\gamma}$ and $N = \sum_{a \in \mathcal{A}} \xi_a$. We

show in Lemma 5 that if we know a solution of Problem $\mathcal{P}(\Delta)$, then we can construct easily the solution of $\mathcal{P}(\Delta + N)$ by adding ξ_a items to each group a .

Lemma 5. *If $(U_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta)$, then $(U_a + \xi_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta + N)$.*

Proof: Let $(U_a)_{a \in \mathcal{A}}$ be an optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a = \Delta$. We have: $\frac{U_a + \xi_a}{|\mathcal{F}_a|} = \frac{1}{\gamma} + \frac{U_a}{|\mathcal{F}_a|}, \forall a \in \mathcal{A}$, and thus: $\max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|} = \frac{1}{\gamma} + \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|}$. Let $b \in \mathcal{A}$ be such that $\frac{U_b + \xi_b}{|\mathcal{F}_b|} = \max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|}$. Let $(V_a)_{a \in \mathcal{A}}$ be any feasible solution of $\mathcal{P}(\Delta + N)$.

- If $V_b \geq U_b + \xi_b$, then $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \frac{U_b + \xi_b}{|\mathcal{F}_b|} = \max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|}$.
- If $V_b < U_b + \xi_b$, then $V_b \leq U_b + \xi_b - 1$, since V_b, U_b and ξ_b are integer. Thus we have: $\exists c : V_c \geq U_c + \xi_c + 1$, since $\sum_{a \in \mathcal{A}} (U_a + \xi_a) = \Delta + N$. Then $\frac{V_c}{|\mathcal{F}_c|} \geq \frac{1}{\gamma} + \frac{U_c + 1}{|\mathcal{F}_c|}$ and it follows from Lemma 2 that: $\frac{V_c}{|\mathcal{F}_c|} \geq \frac{1}{\gamma} + \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|} = \max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|}$. We thus have: $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|}$.

Thus, $\max_{a \in \mathcal{A}} \frac{V_a}{|\mathcal{F}_a|} \geq \max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|}$, and therefore $(U_a + \xi_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta + N)$.

Let now assume that $(U_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U_a > \Delta$. There exists $(U'_a)_{a \in \mathcal{A}}$ that is another optimal solution of $\mathcal{P}(\Delta)$ such that $\sum_{a \in \mathcal{A}} U'_a = \Delta$ (Lemma 1). According to the discussion above, $(U'_a + \xi_a)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(\Delta + N)$. We then have: $\max_{a \in \mathcal{A}} \frac{U_a + \xi_a}{|\mathcal{F}_a|} = \frac{1}{\gamma} + \max_{a \in \mathcal{A}} \frac{U_a}{|\mathcal{F}_a|} = \frac{1}{\gamma} + \max_{a \in \mathcal{A}} \frac{U'_a}{|\mathcal{F}_a|} = \max_{a \in \mathcal{A}} \frac{U'_a + \xi_a}{|\mathcal{F}_a|}$. Therefore, $(U_a + \xi_a)_{a \in \mathcal{A}}$ is also an optimal solution of $\mathcal{P}(\Delta + N)$. \square

Finally, Lemma 6 uses the result from Lemma 5 to construct a sequence of N elements a_1, \dots, a_N in \mathcal{A} that contains exactly ξ_a repetitions of each $a \in \mathcal{A}$ and from which we can construct the solution of Problem $\mathcal{P}(\Delta)$, for any $\Delta \geq 0$. Their order is chosen such that, for each $\Delta = 1, \dots, N$, we can construct a solution of Problem $\mathcal{P}(\Delta)$ by counting the number of times a_i is equal to a , for $i = 1, \dots, \Delta$. In other words, $\mathcal{P}(\Delta)$ is solved by the vector U defined by $U_a = \sum_{i=1}^{\Delta} \mathbb{1}(a_i = a), \forall a \in \mathcal{A}$, where $\mathbb{1}$ denotes the indicator function. If $\Delta > N$, then we use a similar process based on the Euclidean division of Δ by N in combination with Lemma 5.

Lemma 6. *There exists a sequence $(a_1, \dots, a_N) \in \mathcal{A}^N$ such that, for any $\Delta \geq 0$, if q and r denote the quotient and remainder of the Euclidean division of Δ by N , then $(U_a)_{a \in \mathcal{A}}$ defined by $U_a = q\xi_a + \sum_{i=1}^r \mathbb{1}(a_i = a), \forall a \in \mathcal{A}$ is an optimal solution of $\mathcal{P}(\Delta)$. (By convention, $\sum_{i=1}^0 \mathbb{1}(a_i = a) = 0, \forall a \in \mathcal{A}$).*

Proof: We consider an optimal solution $(U_a^N)_{a \in \mathcal{A}}$ of Problem $\mathcal{P}(N)$ such that $\sum_{a \in \mathcal{A}} U_a^N = N$ (Lemma 1). According to Lemma 4, there exist $(U_a^0)_{a \in \mathcal{A}}, \dots, (U_a^{N-1})_{a \in \mathcal{A}}$ and $a_1, \dots, a_N \in \mathcal{A}$ such that for all $p = 1, \dots, N$: $(U_a^p)_{a \in \mathcal{A}}$ is an optimal solution of $\mathcal{P}(p)$ and $U_{a_p}^{p-1} = U_{a_p}^p - 1$ and $U_a^{p-1} = U_a^p, \forall a \neq a_p$. In other words, there exists a sequence a_1, \dots, a_N such that for all $p = 1, \dots, N - 1$, the vector U defined by $U_a = \sum_{i=1}^p \mathbb{1}(a_i = a)$ solves Problem $\mathcal{P}(p)$. We apply this result to $r \in \{0, \dots, N - 1\}$, and according Lemma 5 (applied q times), the vector U defined by $U_a = q\xi_a + \sum_{i=1}^r \mathbb{1}(a_i = a), \forall a \in \mathcal{A}$ is an optimal solution of $\mathcal{P}(r + qN)$, i.e., of $\mathcal{P}(\Delta)$. \square

With these lemmas, we can now turn to Proposition 2. It studies the case where each airline's share of flights is identical across all periods. Specifically, we assume that the number of flights scheduled by each airline a during each period t is the product of an airline-related factor α_a and a period-related factor β_t . In this case, the joint optimization of efficiency and equity comes from the similarity of airline schedules of flights.

Proposition 2. *If there exist $(\alpha_a)_{a \in \mathcal{A}}$ and $(\beta_t)_{t \in \mathcal{T}}$ such that $|\mathcal{D}_t \cap \mathcal{F}_a| = \alpha_a \beta_t, \forall a \in \mathcal{A}, t \in \mathcal{T}$, then there exists a solution that solves (EFF) and (EQ).*

Proof: We consider an optimal solution of (EFF), which we denote by $(w^{\text{eff}}, u^{\text{eff}})$. We denote by X_t^+ (resp. X_t^-) the number of flights that, under solution $(w^{\text{eff}}, u^{\text{eff}})$, are displaced from period t to period $t - 1$ (resp. $t + 1$), i.e., $X_t^- = |\{i \in \mathcal{D}_t | u_i^{\text{eff}} = -1\}|$ (resp. $X_t^+ = |\{i \in \mathcal{D}_t | u_i^{\text{eff}} = +1\}|$). We also denote by X_t the total number of flights displaced from period t , i.e., $X_t = X_t^- + X_t^+, \forall t \in \mathcal{T}$. The optimal objective value function of (EFF) is $\Delta^* = \sum_{i=1}^n |u_i^{\text{eff}}| = \sum_{t \in \mathcal{T}} (X_t^- + X_t^+) = \sum_{t \in \mathcal{T}} X_t$. We aim to construct a solution (w^*, u^*) that is feasible, efficient and equitable.

A sufficient condition for (w^*, u^*) to be feasible and efficient is to ensure that, for each period t , the number of flights rescheduled to the preceding (resp. following) period under

solution (w^*, u^*) is equal to that under solution $(w^{\text{eff}}, u^{\text{eff}})$, that is $|\{i \in \mathcal{D}_t | u_i^* = -1\}| = X_t^-$ and $|\{i \in \mathcal{D}_t | u_i^* = +1\}| = X_t^+, \forall t \in \mathcal{T}$. Indeed, if this condition is verified, the aggregate schedule is identical under solutions $(w^{\text{eff}}, u^{\text{eff}})$ and (w^*, u^*) (i.e., $\sum_{i \in \mathcal{F}} (w_{it}^{\text{eff}} - w_{i,t+1}^{\text{eff}}) = \sum_{i \in \mathcal{F}} (w_{it}^* - w_{i,t+1}^*)$), so solution (w^*, u^*) is feasible. Moreover, we have, under this condition: $\sum_{i \in \mathcal{D}_t} |u_i^*| = X_t^- + X_t^+, \forall t \in \mathcal{T}$, and by summing over t we obtain: $\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{D}_t} |u_i^*| = \sum_{t \in \mathcal{T}} X_t$, i.e., $\sum_{i \in \mathcal{F}} |u_i^*| = \Delta^*$, so solution (w^*, u^*) is efficient.

A sufficient condition for (w^*, u^*) to be equitable is to ensure that the vector U defined by $U_a = \sum_{i \in \mathcal{F}_a} |u_i^*|, \forall a \in \mathcal{A}$ solves Problem $\mathcal{P}(\Delta^*)$ studied in the lemmas above.

We assume without loss of generality that $\gcd(\alpha_a)_{a \in \mathcal{A}} = 1$. As $|\mathcal{F}_a| = \alpha_a (\sum_{t \in \mathcal{T}} \beta_t), \forall a \in \mathcal{A}$, $\gcd(|\mathcal{F}_a|)_{a \in \mathcal{A}} = \sum_{t \in \mathcal{T}} \beta_t$ and $\frac{|\mathcal{F}_a|}{\gcd(|\mathcal{F}_a|)_{a \in \mathcal{A}}} = \alpha_a, \forall a \in \mathcal{A}$. We denote by $N = \sum_{a \in \mathcal{A}} \alpha_a$. According to Lemma 6, there exists a sequence $(a_1, \dots, a_N) \in \mathcal{A}^N$ such that $\sum_{i=1}^N \mathbb{1}(a_i = a) = \alpha_a, \forall a \in \mathcal{A}$ and the vector U defined by $U_a = q\alpha_a + \sum_{i=1}^r \mathbb{1}(a_i = a), \forall a \in \mathcal{A}$ is an optimal solution of $\mathcal{P}(\Delta^*)$, where q and r denote the quotient and the remainder of the Euclidean division of Δ^* by N (i.e., $\Delta^* = qN + r$). We denote by Ψ the sequence $\Psi = (a_1, \dots, a_N, \dots, a_1, \dots, a_N, a_1, \dots, a_r)$, where the full sequence (a_1, \dots, a_N) is repeated q times. By construction, $\sum_{i=1}^{\Delta^*} \mathbb{1}(\Psi_i = 1) = q\alpha_a + \sum_{i=1}^r \mathbb{1}(a_i = a), \forall a \in \mathcal{A}$ and thus the vector U defined by $U_a = \sum_{i=1}^{\Delta^*} \mathbb{1}(\Psi_i = a), \forall a \in \mathcal{A}$ is an optimal solution of $\mathcal{P}(\Delta^*)$.

We now construct a solution (w^*, u^*) that verifies (i) the sufficient conditions for feasibility and efficiency: $|\{i \in \mathcal{D}_t | u_i^* = -1\}| = X_t^-, \forall t \in \mathcal{T}$ and $|\{i \in \mathcal{D}_t | u_i^* = +1\}| = X_t^+, \forall t \in \mathcal{T}$, and (ii) the sufficient condition for equity: $\sum_{i \in \mathcal{F}_a} |u_i^*| = \sum_{i=1}^{\Delta^*} \mathbb{1}(\Psi_i = a), \forall a \in \mathcal{A}$.

To do so, we construct a solution that displaces flights from the sequence of airlines Ψ , i.e., a solution that displaces flights from airline Ψ_1, \dots , from airline Ψ_{X_1} in period 1, flights from airline Ψ_{X_1+1}, \dots , from airline $\Psi_{X_1+X_2}$ in period 2, etc. (of course, each airline may be repeated several times in each sequence). We denote by y_t the number of flights displaced between period 1 and period $t - 1$, i.e., $y_t = \sum_{s=1}^{t-1} X_s$. Note that $y_1 = 0$ and $y_{T+1} = \sum_{s \in \mathcal{T}} X_s = \Delta^*$. We denote by V_{at} the number of times airline a is repeated in the X_t indices between $y_t + 1$ and y_{t+1} , i.e.: $V_{at} = \sum_{i=y_t+1}^{y_{t+1}} \mathbb{1}(\Psi_i = a), \forall a \in \mathcal{A}$. Given the periodicity of the sequence Ψ , we have: $V_{at} \leq \alpha_a \beta_t$. We can thus define a set $\mathcal{J}_{at} \subseteq \mathcal{D}_t \cap \mathcal{F}_a$

such that $|\mathcal{J}_{at}| = V_{at}$. As in the proof of Proposition 1, \mathcal{J}_{at} is not uniquely determined, but we can choose any subset of $\mathcal{D}_t \cap \mathcal{F}_a$ that verifies this property. We construct a solution that displaces the flights in the sets \mathcal{J}_{at} such that the number of flights rescheduled to period $t - 1$ (resp. $t + 1$) is equal to X_t^- (resp. X_t^+). For each $t \in \mathcal{T}$, we partition $\cup_{a \in \mathcal{A}} \mathcal{J}_{at}$ into two subsets \mathcal{K}_t^+ and \mathcal{K}_t^- such that $|\mathcal{K}_t^+| = X_t^+$ and $|\mathcal{K}_t^-| = X_t^-$. Again, these subsets are not uniquely determined. We then define (i) $u_i^* = -1, \forall i \in \mathcal{K}_t^-$, (ii) $u_i^* = +1, \forall i \in \mathcal{K}_t^+$, (iii) $u_i^* = 0, \forall i \notin \mathcal{K}_t^- \cup \mathcal{K}_t^+$. We define w^* accordingly (based on the constraints of (EFF) and (EQ)).

By construction, the solution (w^*, u^*) satisfies the sufficient conditions for feasibility and efficiency, so it solves (EFF). Moreover, we have: $\sum_{i \in \mathcal{D}_t \cap \mathcal{F}_a} |u_i^*| = V_{at}$. By summing over $t \in \mathcal{T}$, we obtain: $\sum_{i \in \mathcal{F}_a} |u_i^*| = \sum_{t \in \mathcal{T}} \sum_{i=y_t+1}^{y_{t+1}} \mathbb{1}(\Psi_i = a)$, that is $\sum_{i \in \mathcal{F}_a} |u_i^*| = \sum_{i=y_1+1}^{y_{T+1}} \mathbb{1}(\Psi_i = a) = \sum_{i=1}^{\Delta^*} \mathbb{1}(\Psi_i = a)$ (as $y_1 = 0$ and $y_T = \Delta^*$). Therefore, the solution (w^*, u^*) solves Problem (EQ). \square

Intuitively, under the conditions presented above, the efficiency-maximizing problem (EFF) admits a relatively large number of solutions. So one can choose, among this solution set, a solution that satisfies a secondary objective, such as maximizing inter-airline equity. In other words, the central decision-maker enjoys significant flexibility regarding *which* flights to reschedule, and in the case where airline schedules of flights exhibit some regularity, this selection can be done in a way that ensures inter-airline equity.

6.3.2 Cases of Efficiency/Equity Trade-off

An important question consists then of examining the conditions under which efficiency and equity cannot be jointly maximized. Based on the discussion above, this occurs through (i) inter-airline variations in flight schedules, (ii) inter-airline variations in network connectivities, and (iii) inter-airline variations in flight valuations.

We first provide an example that shows that weighted efficiency and equity may not be jointly maximized in the presence of differentiated airline schedules. Figure 6-3 shows a hypothetical original schedule in a 7-period case with 2 airlines and 26 flights per airline,

and a simple capacity constraint that ensures that no more than 10 flights may be scheduled per period. Figure 6-3a (resp. Figure 6-3b) shows which flights are rescheduled to later or earlier times for an efficiency-maximizing solution (resp. an equity-maximizing solution). We assume that airline 1’s flights (shown in red) are concentrated at earlier periods, and airline 2’s flights (shown in green) are concentrated at later periods. Note that the conditions of either Proposition 1 or 2 are not satisfied here. The capacity constraint is only violated during period 5, when all flights scheduled are airline 2’s flights. In this case, every efficiency-maximizing solution displaces 4 flights from airline 2 to later times, by 1 period each (shown as “+1”s in Figure 6-3a). The resulting total displacement is equal to 4 periods, and the airline disutilities are equal to 0 for airline 1 and to $4/26$ for airline 2. In contrast, every equity-maximizing solution displaces 3 flights from airline 1 and 3 flights from airline 2 to earlier times, by 1 period each (shown as “-1”s in Figure 6-3b). The resulting total displacement is equal to 6 periods, and each airline’s disutility is equal to $3/26$. In turn, the set of efficiency-maximizing solutions and the set of equity-maximizing solutions have no overlap.

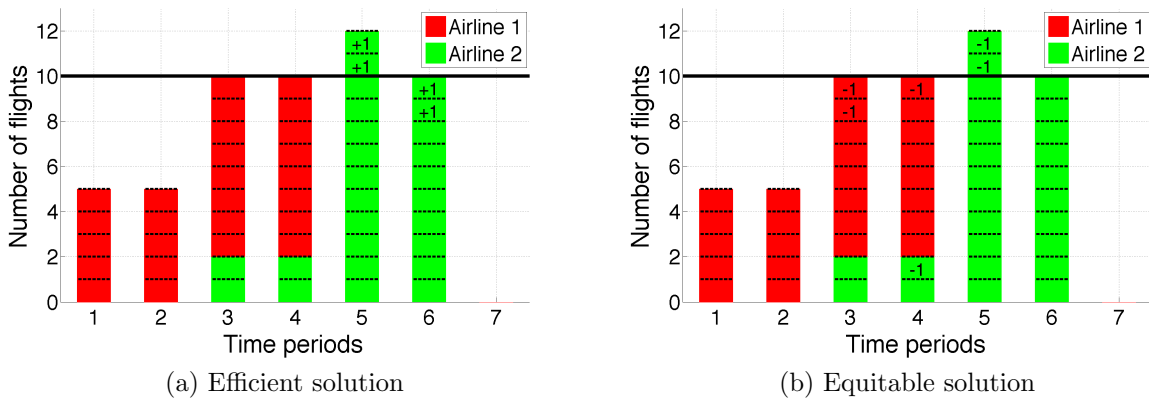


Figure 6-3: Trade-off between efficiency and equity due to variations in airline schedules

We now provide an example that shows that weighted efficiency and equity may not be jointly optimized in the presence of differentiated airline network connectivities. Intuitively, if one airline’s network is significantly more connected than another airline’s, then the flights from the former airline are likely to be more “costly” to reschedule. In turn, maximizing

efficiency may involve assigning more displacement to the latter airline's flights rather than the former's, at some equity loss. Figure 6-4 shows such an example with 5 periods, 2 airlines with 13 flights each, and a capacity of 6 flights per period. Note that the conditions of both Propositions 1 and 2 could be verified in the absence of network connections. But airline 1's network involves a number of connections, whereas airline 2's network is not connected. We represent connections by dashed, gray "links" between flight pairs, and we assume that each connection requires a 2-period interval between the flights in the connection at a minimum. In this case, every efficiency-maximizing solution displaces 4 flights from airline 2 (the airline with no connections) by 1 period each. The resulting total displacement is equal to 4 periods, and the airline disutilities are equal to 0 for airline 1 and to $4/13$ for airline 2. In contrast, every efficiency-maximizing solution displaces 3 flights from airline 1 and 3 flights from airline 2, by 1 period each. The resulting total displacement is equal to 6 periods, and each airline's disutility is equal to $3/26$. Again, the set of efficiency-maximizing solutions and the set of equity-maximizing solutions have no overlap.

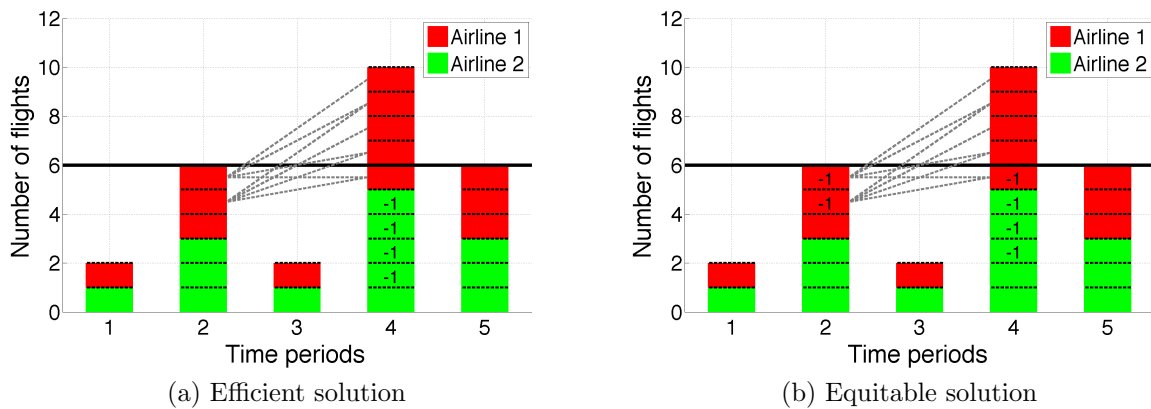


Figure 6-4: Trade-off between efficiency and equity due to variations in airline connectivities

Finally, we provide an example that shows that weighted efficiency and equity may not be jointly optimized in the presence of differentiated flight valuations. Figure 6-5 shows an example with 5 periods, 2 airlines with 10 flights each, and a capacity of 6 flights per period. Note that the conditions of both Propositions 1 and 2 could be verified under uniform flight valuations. But we assume that every flight, except the 6 flights scheduled by airline 2 in

period 3, has a value equal to $v_i = 1$. Specifically, three of them have a value $v_i = 0.1$ each, and three others have a value $v_i = 1.9$ (the average value of airline 2's flights is equal to 1). Every efficiency-maximizing solution displaces the three flights of value $v_i = 0.1$ and three flights of value $v_i = 1$. The optimal value of the weighted displacement is equal to 3.3 and the airline disutilities are equal to 3/10 for airline 1 and to 0.3/10 for airline 2. In contrast, every efficiency-maximizing solution displaces two flights of airline 1 and four flights of airline 2. The weighted displacement is equal to 4.2 and the airline disutilities are equal to 2/10 for airline 1 and to 2.2/10 for airline 2. Again, the set of efficiency-maximizing solutions and the set of equity-maximizing solutions have no overlap.

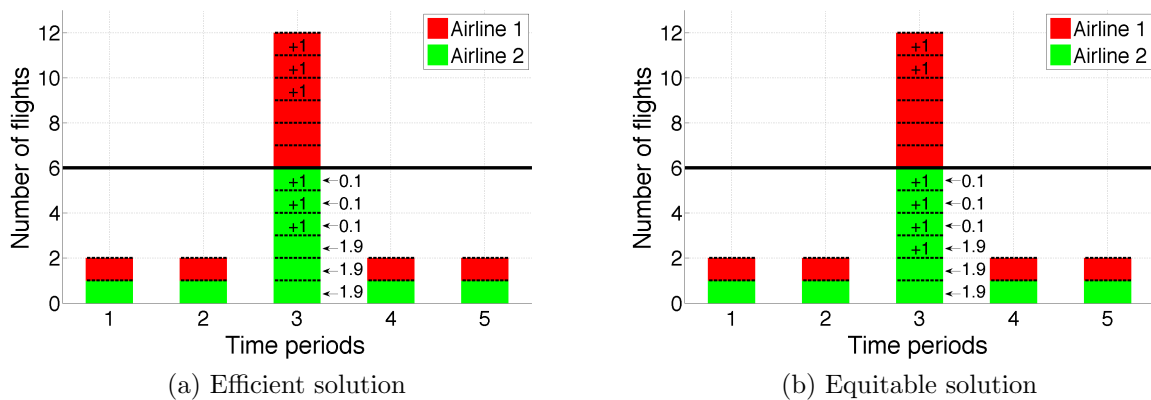


Figure 6-5: Trade-off between efficiency and equity due to variations in flight valuations

6.4 Mechanisms for Airport Scheduling Interventions

The models for scheduling interventions developed in this chapter rely on scheduling, connections, and flight valuations data. However, in the current scheduling environment at US airports, only the scheduling data (i.e., \mathcal{F} , \mathcal{F}^{arr} , \mathcal{F}^{dep} , \mathcal{F}_a , \mathcal{S}^{arr} , and \mathcal{S}^{dep}) are provided by the airlines. Connections data (i.e., \mathcal{C} , \mathcal{C}_a , t^{min} , and t^{max}) and flights' timetabling flexibility (i.e., flight valuations data v) are unknown to the central decision-maker at the time of scheduling interventions. Aircraft and crew connections are private information of the airlines (aircraft connections are only revealed on the day of operations), and so are flight valuations.

Passenger connections depend on air travelers' booking decisions and are thus subject to uncertainty before these decisions are made. In this section, we develop several mechanisms for scheduling interventions that will define the scheduling inputs provided by the airlines and the subsequent scheduling interventions optimized by the central decision-maker. In order to be implementable, we ensure that the considered mechanisms are non-monetary, transparent, and computationally tractable.

First, we define below a *Baseline Scheduling Mechanism with Implied Network considerations* (BSM-IN). The BSM-IN is the mechanism that is assumed by the ICUSM (Chapter 5). Under this mechanism, the airlines provide only their schedules of flights. The central decision-maker then implies the network connections to maintain (e.g., using historical data on aircraft routings and passenger itineraries), and assumes uniform flight valuations (i.e., $v_i = 1, \forall i \in \mathcal{F}$). Scheduling interventions are optimized based on efficiency and on-time performance objectives only (i.e., we solve Problems **P1** and **P2**). Though simple, this mechanism is consistent with current practice at the busiest US airports.

Mechanism 1 Baseline Scheduling Mechanism with Implied Network considerations (BSM-IN)

The airlines submit their preferred schedules of flights $\rightarrow \mathcal{F}, \mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}, \mathcal{F}_a, S^{\text{arr}}, S^{\text{dep}}$

The central decision-maker implies the network connections to maintain $\rightarrow \mathcal{C}, \mathcal{C}_a, t^{\text{min}}, t^{\text{max}}$

The central decision-maker assumes uniform flight valuations $\rightarrow v_i = 1, \forall i \in \mathcal{F}$

The central decision-maker solves Problems **P1** and **P2**

Our second mechanism is similar to the BSM-IN, but considers inter-airline equity objectives in the optimization of scheduling interventions. As in the BSM-IN, the airlines submit only their preferred schedule of flights, and the central decision-maker implies the network connections to maintain and assumes uniform flight valuations. However, scheduling interventions are optimized with respect to efficiency, equity, and on-time performance objectives (i.e., we solve Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$). We name this mechanism the Equitable Scheduling Mechanism with Implied Network considerations (ESM-IN).

We introduce next an Equitable Scheduling Mechanism with Specified Network considerations (ESM-SN) that enables the airlines to provide the network connections to maintain

Mechanism 2 Equitable Scheduling Mechanism with Implied Network considerations (ESM-IN)

The airlines submit their preferred schedules of flights $\rightarrow \mathcal{F}, \mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}, \mathcal{F}_a, S^{\text{arr}}, S^{\text{dep}}$

The central decision-maker implies the network connections to maintain $\rightarrow \mathcal{C}, \mathcal{C}_a, t^{\text{min}}, t^{\text{max}}$

The central decision-maker assumes uniform flight valuations $\rightarrow v_i = 1, \forall i \in \mathcal{F}$

The central decision-maker solves Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$

(instead of these connections being implied by the central decision-maker). The ESM-SN allows each airline a to reveal their pairs of connecting flights (i.e., \mathcal{C}_a), with corresponding bounds on the connecting times (i.e., t^{min} and t^{max}). Note that these connections can be requested to enable an aircraft to fly both flight legs, a passenger to connect between the inbound flight and the outbound flight, or a crew to serve both flight legs. Any connection requested adds one constraint to the model of scheduling interventions and may thus result in lower efficiency and/or lower equity, for any given on-time performance targets. For this reason, the proposed mechanism may add a “cost” for each connection requested, reflected through lower flight valuations. In order to reflect this cost without involving any monetary transfer, the ESM-SN is based on the allocation of (non-monetary) “credits” to the different airlines. Specifically, each airline receives a given number of “credits” per flight, denoted by \bar{v} , and loses β “credits” per requested connection. In turn, each airline receives a total of $\bar{v}|\mathcal{F}_a| - \beta|\mathcal{C}_a|$ credits. As in the previous mechanisms, the central decision-maker assumes that, for each airline, all flights are equally valued, i.e., all flights of airline a are given a valuation $v_i = \bar{v} - \beta \frac{|\mathcal{C}_a|}{|\mathcal{F}_a|}$. Scheduling interventions are then optimized with respect to efficiency, equity, and on-time performance objectives by solving Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$). We discuss the calibration of the parameter β in Section 6.5.2.

Finally, the last mechanism builds upon the ESM-SN, but enables the airlines to provide the relative valuation of their flights through credit allocation. In other words, each airline can allocate its $\bar{v}|\mathcal{F}_a| - \beta|\mathcal{C}_a|$ credits to its flights to signal their relative scheduling flexibility. The more inconvenient, or costly, the rescheduling of any flight i , the more credits are allocated to flight i , subject to the credit-availability constraint $\sum_{i \in \mathcal{F}_a} v_i \leq \bar{v}|\mathcal{F}_a| - \beta|\mathcal{C}_a|, \forall a \in$

Mechanism 3 Equitable Scheduling Mechanism with Specified Network considerations (ESM-SN)

The airlines submit their preferred schedules of flights $\rightarrow \mathcal{F}, \mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}, \mathcal{F}_a, S^{\text{arr}}, S^{\text{dep}}$

The airlines submit the network connections to maintain $\rightarrow \mathcal{C}, \mathcal{C}_a, t^{\text{min}}, t^{\text{max}}$

Each airline $a \in \mathcal{A}$ receives a total of $\bar{v} |\mathcal{F}_a| - \beta |\mathcal{C}_a|$ credits

The central decision-maker assumes uniform flight valuations, $\rightarrow v_i = \bar{v} - \beta \frac{|\mathcal{C}_a|}{|\mathcal{F}_a|}, \forall i \in \mathcal{F}_a, \forall a \in \mathcal{A}$

The central decision-maker solves Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$

\mathcal{A} . These credits are then used by the central decision-maker to optimize scheduling interventions, i.e., to solve Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$). We refer to this mechanism as Equitable Credit-based Scheduling Mechanism with Specified Network considerations (ECSM-SN).

Mechanism 4 Equitable Credit-based Scheduling Mechanism with Specified Network considerations (ECSM-SN)

The airlines submit their preferred schedules of flights $\rightarrow \mathcal{F}, \mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}, \mathcal{F}_a, S^{\text{arr}}, S^{\text{dep}}$

The airlines submit the network connections to maintain $\rightarrow \mathcal{C}, \mathcal{C}_a, t^{\text{min}}, t^{\text{max}}$

Each airline $a \in \mathcal{A}$ receives a total of $\bar{v} |\mathcal{F}_a| - \beta |\mathcal{C}_a|$ credits

The airlines allocate their credits to their individual flights $\rightarrow v_i, \forall i \in \mathcal{F}$, subject to the constraint: $\sum_{i \in \mathcal{F}_a} v_i \leq \bar{v} |\mathcal{F}_a| - \beta |\mathcal{C}_a|, \forall a \in \mathcal{A}$

The central decision-maker solves Problems **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$

Note that the proposed mechanisms introduce an increasing extent of collaboration. While the BSM-IN relies on flight scheduling data exclusively and optimizes scheduling interventions from a centralized, efficiency-based perspective, equity considerations and airline scheduling preferences are incorporated gradually to the mechanisms. Specifically, the ESM-IN introduces equity considerations, the ESM-SN enables the airlines to additionally provide the network connections that need to be maintained, and the ECSM-SN enables the airlines to additionally provide the degree of relative scheduling flexibility of their flights.

6.5 Computational Results

We now implement the mechanisms proposed in Section 6.4 at JFK. We consider data from September 18, 2007, whose number of flights corresponds to the median of the number of

daily flights at JFK in 2007. We group partner airlines together, as major airlines typically coordinate planning and scheduling decisions with their subsidiaries, and passengers can connect between flights operated by partner airlines. Under these groupings, all passenger connections are intra-airline connections, which is consistent with our modeling framework. Specifically, we consider four groups of airlines at JFK: (i) Delta Airlines (DAL) and its regional partners (which operated a total of 320 flights on 09/18/2007), (ii) American Airlines (AAL) and its regional partners (260 flights), (iii) JetBlue Airways (JBU) (174 flights), and (iv) all other airlines, each of which represents a smaller share of traffic at JFK (408 flights combined). These scheduling data were used to construct \mathcal{F} , \mathcal{F}^{arr} , \mathcal{F}^{dep} , \mathcal{F}_a , \mathcal{S}^{arr} , and \mathcal{S}^{dep} .

Regarding network connections, even though they are not made available at the time of the scheduling interventions, we reconstructed aircraft and passenger connections using *ex post* operational data, as described in Chapter 5. Because of data unavailability, we do not reconstruct crew connections here. In the remainder of this chapter, the network connections data (i.e., the sets and parameters \mathcal{C} , \mathcal{C}_a , t^{min} , and t^{max}) will be based either on these *ex post* aircraft and passenger connections (which we will refer to as “baseline connections data”) or on alternative sets of network connections that we shall construct.

We consider the same set of on-time performance targets A_{MAX} and D_{MAX} as in Chapter 5. With the schedule of 09/18/2007, all the on-time performance targets considered are met with a maximal flight displacement $\delta^* = 1$ period, i.e., without displacing any flight by more than 15 minutes. The results reported in this section will thus focus on weighted efficiency Δ (which, for clarity, we refer to simply by “efficiency” in the remainder of this section) and on inter-airline equity Φ (obtained from the vectors of airline disutilities σ), for any set of on-time performance targets. Unless otherwise specified, we use values of $A_{\text{MAX}} = 11$ aircraft and $D_{\text{MAX}} = 15$ aircraft.

6.5.1 Inter-Airline Equity: The ESM-IN

We compare here the results of the Baseline Scheduling Mechanism with Implied Network considerations (BSM-IN) and the Equitable Scheduling Mechanism with Implied Network

considerations (ESM-IN). They are both based on airlines’ preferred schedules of flights, assuming perfect information regarding network connections (using *ex post* “baseline” connections data), and uniform flight valuations ($v_i = 1, \forall i$). They differ by the objective of the scheduling interventions: the BSM-IN maximizes efficiency, under on-time performance targets (Problems **P1** and **P2**), while the ESM-IN incorporates inter-airline equity objectives into decision-making (Problems $\widehat{\mathbf{P3}}(\rho)$). Comparing the outcomes of these two mechanisms thus informs on the extent to which inter-airline equity can be achieved in scheduling interventions in the baseline setting.

Note that the solution of Problem **P2** is arbitrarily “chosen” by the solver among the set of optimal solutions. In order to characterize the equity range among efficiency-maximizing solutions, we determine the solution which *minimizes* inter-airline equity, i.e., which lexicographically *maximizes* airline disutilities, for the optimal value of efficiency. In other words, we characterize the efficiency-maximizing solution that performs the worst in terms of inter-airline equity. We denote this problem by $\overline{\mathbf{P2}}$.

Table 6.1 shows the schedule displacement faced by each airline (i.e., the number of its flights displaced by 15 minutes each, as $\delta^* = 1$), and each airline’s disutility (i.e., its average per-flight displacement) for Problems **P2**, $\overline{\mathbf{P2}}$ and $\widehat{\mathbf{P3}}(\rho^*)$. It also reports the ratio of the largest to smallest airline disutility. Note, first, that Problem $\overline{\mathbf{P2}}$ results in max-min ratios $\frac{\max_a \sigma_a}{\min_a \sigma_a}$ ranging between 10 and 50. For the cases considered, AAL and JBU tend to be much more significantly penalized than DAL, which is reflected through more of their flights being rescheduled and through higher disutility values. The set of efficiency-maximizing solutions thus contains very inequitable outcomes. Problem **P2** does not result in the most inequitable outcome in that set, but provides solutions that still impact some airlines (here, AAL, JBU and the “other” airlines) more negatively than others (here, DAL). Inter-airline equity is only achieved by solving Problem $\widehat{\mathbf{P3}}(\rho^*)$. In that case, airline disutilities are much closer to each other than by solving Problems $\overline{\mathbf{P2}}$ and **P2**. Note that the differences in airlines’ schedules of flights and network connectivities result in all airlines *not* having the exact same disutility, but differences are very small (i.e., the max-min ratio $\frac{\max_a \sigma_a}{\min_a \sigma_a}$ is very close to 1) under

the equitable solution. Most important, the equity-maximizing solution (Problem $\widehat{\mathbf{P3}}(\rho^*)$) results in the same total displacement as the efficiency-maximizing solution (Problem $\mathbf{P2}$) in the setting considered. Only the distribution of schedule displacement across the airlines is modified. In other words, efficiency and equity can be jointly maximized and the price of equity is zero (i.e., $\rho^* = 0$).

Table 6.1: Results of the BSM-IN and ESM-IN

On-time targets		Model	Number of flights displaced					Disutility: $\sigma_a = \frac{1}{ \mathcal{F}_a } \sum_{i \in \mathcal{F}_a} u_i $				
A_{MAX}	D_{MAX}		DAL	AAL	JBU	Others	All	DAL	AAL	JBU	Others	$\frac{\max_a \sigma_a}{\min_a \sigma_a}$
14	23	$\overline{\mathbf{P2}}$	1	13	1	5	20	0.3%	5.0%	0.6%	1.2%	16.00
		$\mathbf{P2}$	1	9	2	8	20	0.3%	3.5%	1.1%	2.0%	11.08
		$\widehat{\mathbf{P3}}(\rho^*)$	4	5	3	8	20	1.3%	1.9%	1.7%	2.0%	1.57
13	20	$\overline{\mathbf{P2}}$	1	29	9	7	46	0.3%	11.2%	5.2%	1.7%	35.69
		$\mathbf{P2}$	7	18	8	13	46	2.2%	6.9%	4.6%	3.2%	3.16
		$\widehat{\mathbf{P3}}(\rho^*)$	13	10	7	16	46	4.1%	3.8%	4.0%	3.9%	1.06
12	18	$\overline{\mathbf{P2}}$	1	28	27	9	65	0.3%	10.8%	15.5%	2.2%	49.66
		$\mathbf{P2}$	10	27	10	18	65	3.1%	10.4%	5.7%	4.4%	3.32
		$\widehat{\mathbf{P3}}(\rho^*)$	18	14	10	23	65	5.6%	5.4%	5.7%	5.6%	1.07
11	15	$\overline{\mathbf{P2}}$	37	113	39	17	206	11.6%	43.5%	22.4%	4.2%	10.43
		$\mathbf{P2}$	50	57	32	67	206	15.6%	21.9%	18.4%	16.4%	1.40
		$\widehat{\mathbf{P3}}(\rho^*)$	57	46	31	72	206	17.8%	17.7%	17.8%	17.6%	1.01

Therefore, joint optimization of efficiency and equity seems to be achievable under current schedules of flights, uniform flight valuations and the “baseline” network connections. In light of the results from Section 6.3, this suggests that inter-airline variations in flight schedules and network connectivities are relatively weak and do not create, by themselves, a trade-off between efficiency and equity. This is due to the fact that peak-hour schedules typically include flights from several airlines and the schedules of all airlines exhibit network connections (so the situations depicted in Figures 6-3 and 6-4 are not typical of actual scheduling patterns at busy airports). Under these conditions, incorporating inter-airline equity objectives in scheduling interventions can thus yield significant benefits by balancing scheduling adjustments more fairly among the airlines at no efficiency losses.

6.5.2 Network Connectivities: The ESM-SN

Under the Equitable Scheduling Mechanism with Specified Network considerations (ESM-SN), airlines provide their preferred schedule of flights and their network connections. If $\beta > 0$, requesting connections comes at the cost of some credit loss. From a central decision-maker’s standpoint, the two primary objectives of the mechanism are (i) preserving the connectivity of airlines’ networks of flights, and (ii) ensuring that airlines do not have incentives to request more connections than needed to reduce the displacement of their own schedule (thus minimizing the potential for airline strategic behaviors in the submission of their network connections).

We simulate the sets of connections requested by the airlines by determining, for each airline a , the set of connections that may become infeasible due to a minimum connecting time that may be violated, which we denote by $\widehat{\mathcal{C}}_a$. It is equal to the set of connections that would not satisfy the minimum connecting time constraint if the earlier flight was rescheduled to a *later* time by δ^* periods and if the later flight was rescheduled to an *earlier* time by δ^* periods, i.e.:

$$\widehat{\mathcal{C}}_a = \left\{ (i, j) \in \mathcal{C}_a \mid \sum_{t \in \mathcal{T}} (w_{jt}^{\text{dep}} - w_{it}^{\text{arr}}) - 2\delta^* < t_{ij}^{\text{min}} \right\}$$

We assume that all connections requested by airline a are in the set $\widehat{\mathcal{C}}_a$. If $\beta > 0$, each airline a can request up to $\max\left(\frac{\bar{v}|\mathcal{F}_a|}{\beta}, |\widehat{\mathcal{C}}_a|\right)$ connections, and if $\beta = 0$, each airline a can request up to $|\widehat{\mathcal{C}}_a|$ connections. We vary the number of connections requested by each airline one at a time, by increments of 100 connections, under the assumption that all the other airlines request their “baseline” connections. Note that the objective is *not* to predict accurately which connections the airlines would prioritize and how many connections they would request under the proposed mechanism, for different values of β , but instead to construct various sets of connections and to characterize their impacts on the outcome of scheduling interventions.

Table 6.2 shows the ESM-SN results (with $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$) for different numbers of connections requested by DAL and AAL (similar results are obtained for the

other airlines). It reports the measure of inequity Φ (Equation (6.4)) for Problems $\widehat{\mathbf{P3}}(\mathbf{0})$ and $\widehat{\mathbf{P3}}(\rho^*)$, as well as the price of equity (i.e., ρ^*) and the price of efficiency (i.e., the difference between $\Phi^*(0)$ and $\Phi^*(\rho^*)$). Note, first, that the prices of equity and efficiency are small in the case of $\beta = 0$, but larger with $\beta > 0$. If $\beta = 0$, flight valuations are uniform, so a trade-off between efficiency and equity may arise solely from inter-airline variations in flight schedules and network connectivities. However, it is much weaker than under differentiated flight valuations when $\beta > 0$. This is consistent with the theoretical results of Section 6.3. Moreover, for the different values of β and $|\mathcal{C}_a|$ considered, the price of equity, i.e., the efficiency loss required to maximize inter-airline equity, was found to be within the range of 0% to 10%. In all our experiments summarized in Table 6.2, on-time performance targets are achieved by displacing roughly 200 flights by 15 minutes each, so a 0% to 10% efficiency loss corresponds to displacing 0 to 20 additional flights. In contrast, the price of efficiency can be significant, i.e., *not* displacing these additional flights can result in large equity losses (up to slightly over 100% in some cases). These numerical results show that a substantial improvement in equity can be achieved at small efficiency losses. This is consistent with the theoretical bounds obtained with general utility functions by Bertsimas et al. [16, 17].

Figure 6-6 shows the total displacement for each individual airline under various numbers of connections submitted by DAL and AAL with $\beta = 0$. The three bars in each group represent the outcomes of Problems $\mathbf{P2}$, $\widehat{\mathbf{P3}}(\mathbf{0})$ and $\widehat{\mathbf{P3}}(\rho^*)$, respectively. As expected from Table 6.2, improving inter-airline equity does not induce significant increases in the total displacement, but can change significantly the *distribution* of the displacement among the airlines. Moreover, variations in DAL's and AAL's number of requested connections induce only marginal increases in the total displacement, and do not impact significantly the distribution of the displacement among the airlines. This is an important result in view of the two objectives of the mechanism outline above, i.e., preserving network connectivities and minimizing the potential for strategic behavior. If an airline faced a significantly larger displacement by requesting more connections, then the mechanism would create disincentives for the airlines to request connections and could thus impede network connectivity. If, on

Table 6.2: ESM-SN results: Inequity Φ , price of equity and efficiency

β	$ \mathcal{C}_a $	DAL				AAL			
		$\Phi^*(0)$	$\Phi^*(\rho^*)$	P_{eq}	P_{eff}	$\Phi^*(0)$	$\Phi^*(\rho^*)$	P_{eq}	P_{eff}
0	0	0.18	0.18	0%	0%	0.18	0.18	3.3%	2.1%
	100	0.18	0.18	0%	0%	0.18	0.18	1.9%	3.2%
	200	0.18	0.18	0%	0%	0.19	0.18	4.3%	4.3%
	300	0.18	0.18	0%	0%	0.18	0.18	2.4%	1.7%
	400	0.18	0.18	0.5%	0.3%	0.19	0.18	2.8%	4.0%
	500	0.18	0.18	2.4%	2.7%	0.19	0.18	4.3%	4.1%
	600	0.19	0.18	3.8%	3.0%	0.19	0.19	4.1%	2.6%
	700	0.19	0.18	6.7%	4.0%	0.19	0.19	3.2%	3.2%
	800	0.19	0.19	6.2%	3.9%	0.20	0.19	3.6%	4.6%
	900	0.19	0.19	7.1%	2.7%	0.19	0.19	2.7%	2.6%
	1000	0.20	0.19	6.2%	3.5%	0.20	0.19	1.8%	3.0%
	1100	0.20	0.19	6.1%	3.9%	–	–	–	–
0.1	0	0.18	0.18	1.4%	0.3%	0.30	0.17	5.5%	82.2%
	100	0.34	0.17	2.3%	102.1%	0.30	0.17	6.9%	82.2%
	200	0.22	0.16	1.0%	34.0%	0.30	0.16	4.9%	87.6%
	300	0.20	0.16	1.1%	24.6%	0.27	0.16	7.0%	66.5%
	400	0.27	0.16	1.4%	75.3%	0.28	0.17	7.9%	68.8%
	500	0.29	0.16	6.8%	76.2%	0.29	0.16	6.9%	77.0%
	600	0.28	0.16	9.4%	74.0%	0.27	0.16	8.9%	64.7%
	700	0.28	0.16	6.8%	76.1%	0.25	0.15	9.5%	60.0%
	800	0.29	0.17	5.1%	71.5%	0.24	0.15	8.4%	53.5%
	900	0.29	0.17	6.8%	75.0%	0.23	0.15	8.9%	51.2%
	1000	0.30	0.17	7.5%	74.0%	0.25	0.15	10.0%	65.1%
	1100	0.29	0.17	5.1%	69.5%	–	–	–	–
0.2	0	0.29	0.16	5.2%	84.3%	0.27	0.15	8.5%	73.8%
	100	0.30	0.15	6.1%	93.8%	0.28	0.15	8.5%	87.9%
	200	0.23	0.15	6.1%	51.7%	0.28	0.14	8.6%	90.3%
	300	0.29	0.15	6.1%	100.7%	0.23	0.14	11.7%	61.4%
	400	0.24	0.14	6.1%	67.4%	0.23	0.14	12.0%	66.1%
	500	0.25	0.15	6.1%	74.5%	0.22	0.14	11.6%	56.6%
	600	0.24	0.14	6.1%	71.7%	0.20	0.14	8.4%	40.0%
	700	0.24	0.15	6.1%	68.7%	0.20	0.13	15.4%	58.8%
	800	0.25	0.15	6.1%	65.5%	0.19	0.12	12.6%	61.8%
	900	0.26	0.15	6.1%	67.5%	0.21	0.11	7.6%	91.3%
	1000	0.26	0.15	6.1%	72.1%	0.19	0.10	12.5%	94.5%
	1100	0.25	0.15	6.1%	70.8%	–	–	–	–

the other hand, increasing an airline's number of requested connections led to significant reductions in the displacement faced by this airline, then the mechanism would create incen-

tives for the airlines to request *more* connections than needed, and would thus increase the potential for strategic behavior. In this case, an increase in β would be warranted to correct this effect. The fact that variations in airlines’ requested connections affect only marginally the displacement faced by each airline suggests that the mechanism does achieve the objectives of preserving network connectivities and minimizing the gaming potential with a value of $\beta = 0$. This can be explained by the fact that, since only a fraction of the flights are rescheduled, the algorithm can find *which* flights to reschedule in order to satisfy various network constraints, as requested by the airlines. The resulting mechanism (with $\beta = 0$) offers the advantage of being easy to calibrate and implement.

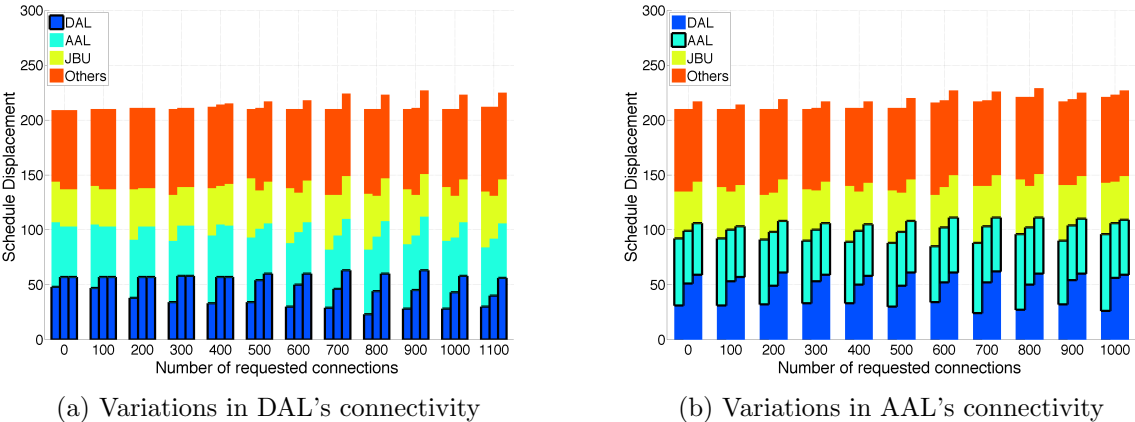


Figure 6-6: ESM-SN Results: Displacement per airline under **P1**, **P2**, and $\widehat{\mathbf{P3}}(\rho)$

Overall, these results suggest that the ESM-SN with $\beta = 0$ (i) maintains the connectivity of airlines’ networks, (ii) does not create significant incentives for untruthful revelations of network connections by the airlines, (iii) achieves inter-airline equity at minimal efficiency losses, and (iv) is easily implementable. In the next section, we use this value of $\beta = 0$ in the ECSM-SN, where we allow the airlines to provide the relative scheduling flexibility of their flights through credit allocation.

6.5.3 Credit Allocation: The ECSM-SN

Under the Equitable Credit-based Scheduling Mechanism with Specified Network considerations (ECSM-SN), the airlines provide their preferred schedule of flights, their network connections, and the relative scheduling flexibility of their flights through credit allocation. We assume that $\beta = 0$ and, without loss of generality, that $\bar{v} = 1$. In turn, each airline a receives $|\mathcal{F}_a|$ credits.

We consider the schedule of flights at JFK on 09/18/2007 and the “baseline” connections. We vary flight valuations for one airline at a time, i.e., we assume that all airlines allocate their credits uniformly across all their flights (i.e., each flight receives exactly $\bar{v} = 1$ credit), except one airline a . We partition its set of flights \mathcal{F}_a in two subsets $\mathcal{F}_a^{(1)}$ and $\mathcal{F}_a^{(2)}$ such that $\mathcal{F}_a^{(1)} \cap \mathcal{F}_a^{(2)} = \emptyset$ and $\mathcal{F}_a^{(1)} \cup \mathcal{F}_a^{(2)} = \mathcal{F}_a$. We can think of $\mathcal{F}_a^{(1)}$ (resp. $\mathcal{F}_a^{(2)}$) as the set of the most flexible flights (resp. the least flexible flights) of airline a . We choose to represent the valuations of the flights in $\mathcal{F}_a^{(1)}$ (resp. $\mathcal{F}_a^{(2)}$) by a Gamma distribution $\Gamma_1(\mu_1, k)$ (resp. $\Gamma_2(\mu_2, k)$) of mean μ_1 (resp. μ_2) and shape parameter k , with $\mu_1 < \mu_2$. We adjust the shape parameter of these distributions such that the 95th percentile of the former distribution coincides with the 5th percentile of the latter. These choices provide a transparent bimodal characterization of positive flight valuations such that the valuations of flights in $\mathcal{F}_a^{(1)}$ are, in most cases, lower than the valuations of flights in $\mathcal{F}_a^{(2)}$. Finally, we set the values of flights in $\mathcal{F}_a^{(1)}$ (resp. $\mathcal{F}_a^{(2)}$) equal to $\Theta_1^{-1} \left(\frac{1}{(|\mathcal{F}_a^{(1)}|+1)} \right), \Theta_1^{-1} \left(\frac{2}{(|\mathcal{F}_a^{(1)}|+1)} \right), \dots, \Theta_1^{-1} \left(\frac{|\mathcal{F}_a^{(1)}|}{(|\mathcal{F}_a^{(1)}|+1)} \right)$ (resp. $\Theta_2^{-1} \left(\frac{1}{(|\mathcal{F}_a^{(2)}|+1)} \right), \Theta_2^{-1} \left(\frac{2}{(|\mathcal{F}_a^{(2)}|+1)} \right), \dots, \Theta_2^{-1} \left(\frac{|\mathcal{F}_a^{(2)}|}{(|\mathcal{F}_a^{(2)}|+1)} \right)$), where Θ_1 (resp. Θ_2) denotes the cumulative distribution function of $\Gamma_1(\mu_1, k)$ (resp. $\Gamma_2(\mu_2, k)$). This ensures that the resulting set of flight valuations is distributed “smoothly” across the distributions considered without sampling these values multiple times. For each airline, we vary two parameters: (i) the fraction of flights in $\mathcal{F}_a^{(1)}$, denoted by $\eta = \frac{|\mathcal{F}_a^{(1)}|}{|\mathcal{F}_a|}$ (so that $1 - \eta = \frac{|\mathcal{F}_a^{(2)}|}{|\mathcal{F}_a|}$), and (ii) the mean valuations of flights in $\mathcal{F}_a^{(1)}$, i.e., μ_1 (so that $\eta\mu_1 + (1 - \eta)\mu_2 = \bar{v}$). We sort flights from the least valuable to the most valuable using 10 random permutations. In other words, the 10 tests have the same sets of flight valuations, but differ in terms of which flights are more and less flexible.

Table 6.3 shows ECSM-SN results (with $A_{\text{MAX}} = 11$ and $D_{\text{MAX}} = 15$) under different sets of flight valuations provided by DAL (left) and AAL (right). In the top half, we assume that $\mathcal{F}_a^{(1)}$ and $\mathcal{F}_a^{(2)}$ both comprise 50% of the flights from DAL or AAL, and we progressively increase the valuation differential $\mu_2 - \mu_1$. In the bottom half, we fix $\mu_1 = 0.75$ and we progressively decrease the proportion of flights in $\mathcal{F}_a^{(1)}$ (and we thus increase μ_2 to ensure that $\eta\mu_1 + (1 - \eta)\mu_2 = 1$). The table reports, in each scenario, the average schedule displacement $\sum_{i \in \mathcal{F}_a} |u_i|$ of *each* airline a obtained in the equity-maximizing scenario (i.e., Problem $\widehat{\mathbf{P3}}(\rho^*)$), as well as the average prices of equity and efficiency, taken across all 10 samples. The observations from variations in $\mu_2 - \mu_1$ (top of the table) and in η (bottom of the table) are threefold. First, as an airline’s flight valuations become more differentiated, the displacement of this airline’s schedule increases. In turn, each airline faces a trade-off between selecting *which* flights get rescheduled, on the one hand, and minimizing its overall displacement, on the other. Second, as the variance in any airline’s flight valuations increases, other airlines’ displacements do not change significantly (if anything, they seem to decrease). In other words, the model can account for any airline’s scheduling preferences without impacting the other airlines too much. Third, the price of equity seems much smaller than the price of efficiency. This is consistent with the results shown in Section 6.5.2, and suggests that, even in the case of differentiated flight valuations, inter-airline equity can be achieved at comparatively small efficiency losses. This motivates attempts to achieve equity improvements in scheduling interventions.

Finally, Figure 6-7 shows the disutilities of each airline (i.e., $\sigma_a = \frac{1}{|\mathcal{F}_a|} \sum_{i \in \mathcal{F}_a} v_i |u_i|$) in the scenarios shown in Table 6.3. The full bars correspond to the results of Problem $\widehat{\mathbf{P3}}(\rho^*)$ when flight valuations data v are considered, and the empty bars correspond to the results of Problem $\widehat{\mathbf{P3}}(\rho^*)$ when flight valuations data v are ignored (i.e., $v_i = 1, \forall i$). As expected, accounting for the flight valuations provided by any airline reduces its disutility. In other words, prioritizing its most flexible flights in the scheduling interventions offsets the potential increase in its overall displacement shown in Table 6.3. As importantly, accounting for flight valuations provided by any airline reduces the disutilities of the other airlines as

Table 6.3: ECSM-SN results: Displacement per airline, price of equity and efficiency

Scenario			Variations in DAL's flight valuations						Variations in AAL's flight valuations					
μ_1	μ_2	η	DAL	AAL	JBU	Oth.	P_{eq}	P_{eff}	DAL	AAL	JBU	Oth.	P_{eq}	P_{eff}
0.9	1.1	50%	57.7	44.0	36.2	70.0	4.4%	15%	55.5	46.9	36.0	69.5	2.8%	25%
0.8	1.2	50%	57.0	44.7	36.4	70.5	4.5%	15%	53.4	50.0	34.6	67.9	2.4%	27%
0.7	1.3	50%	58.1	44.1	35.9	70.2	5.6%	15%	52.2	50.5	33.1	66.6	2.0%	23%
0.6	1.4	50%	59.8	43.5	36.2	69.1	6.8%	14%	51.5	59.1	32.7	65.1	2.1%	16%
0.5	1.5	50%	63.5	43.4	35.3	68.8	7.2%	14%	49.9	64.0	32.1	63.4	2.3%	12%
0.4	1.6	50%	67.3	43.5	34.9	68.9	8.1%	13%	48.8	70.7	31.4	61.4	2.0%	8%
0.3	1.7	50%	73.2	42.3	34.0	67.1	9.3%	13%	46.5	78.4	31.7	59.0	2.7%	10%
0.2	1.8	50%	77.4	42.4	34.8	67.2	11.2%	13%	44.4	87.5	29.9	56.5	2.8%	11%
0.1	1.9	50%	84.1	42.9	34.0	67.8	13.5%	12%	42.0	95.8	29.6	53.4	3.8%	16%
0.75	1.03	90%	57.6	44.4	36.6	70.4	4.1%	17%	49.3	42.3	32.2	62.4	2.4%	20%
0.75	1.06	80%	57.4	44.7	36.6	70.7	4.4%	17%	54.9	48.7	35.2	69.0	2.2%	20%
0.75	1.11	70%	57.3	44.4	36.3	70.4	4.3%	17%	53.5	49.6	35.1	68.4	2.2%	18%
0.75	1.17	60%	57.2	44.7	36.2	70.7	4.5%	16%	54.3	50.8	35.1	68.2	2.6%	18%
0.75	1.25	50%	57.5	44.1	36.4	69.7	5.1%	15%	53.0	52.0	33.8	67.6	1.9%	21%
0.75	1.37	40%	57.6	44.3	36.0	70.5	5.5%	14%	52.7	54.2	34.4	66.9	2.7%	24%
0.75	1.58	30%	58.6	44.1	36.5	70.1	6.2%	12%	51.9	55.8	33.4	65.9	2.8%	33%
0.75	2.00	20%	59.7	44.0	35.6	69.6	6.6%	11%	51.0	58.2	33.0	65.0	2.8%	36%
0.75	3.25	10%	66.6	43.1	35.4	67.9	6.9%	10%	50.0	62.6	31.8	63.3	2.7%	34%

well, consistently with results shown in Table 6.3. In sum, integrating flight valuations in the optimization of scheduling interventions can better satisfy *all* airlines' scheduling preferences, thus improving the outcome of the scheduling interventions.

6.5.4 Summary

Computational results presented in this section have shown that:

- i. Inter-airline equity can be achieved at no, or small, losses in efficiency.
- ii. Network connections can be maintained through the scheduling interventions.
- iii. Accounting for flight valuations can improve the outcome of scheduling interventions.

Based on these results, we propose an airport scheduling interventions mechanism (Mechanism 5) that is based on the Equitable Credit-based Scheduling Mechanism with Specified Network considerations (ECSM-SN) introduced in Section 6.4, but does not penalize the sub-

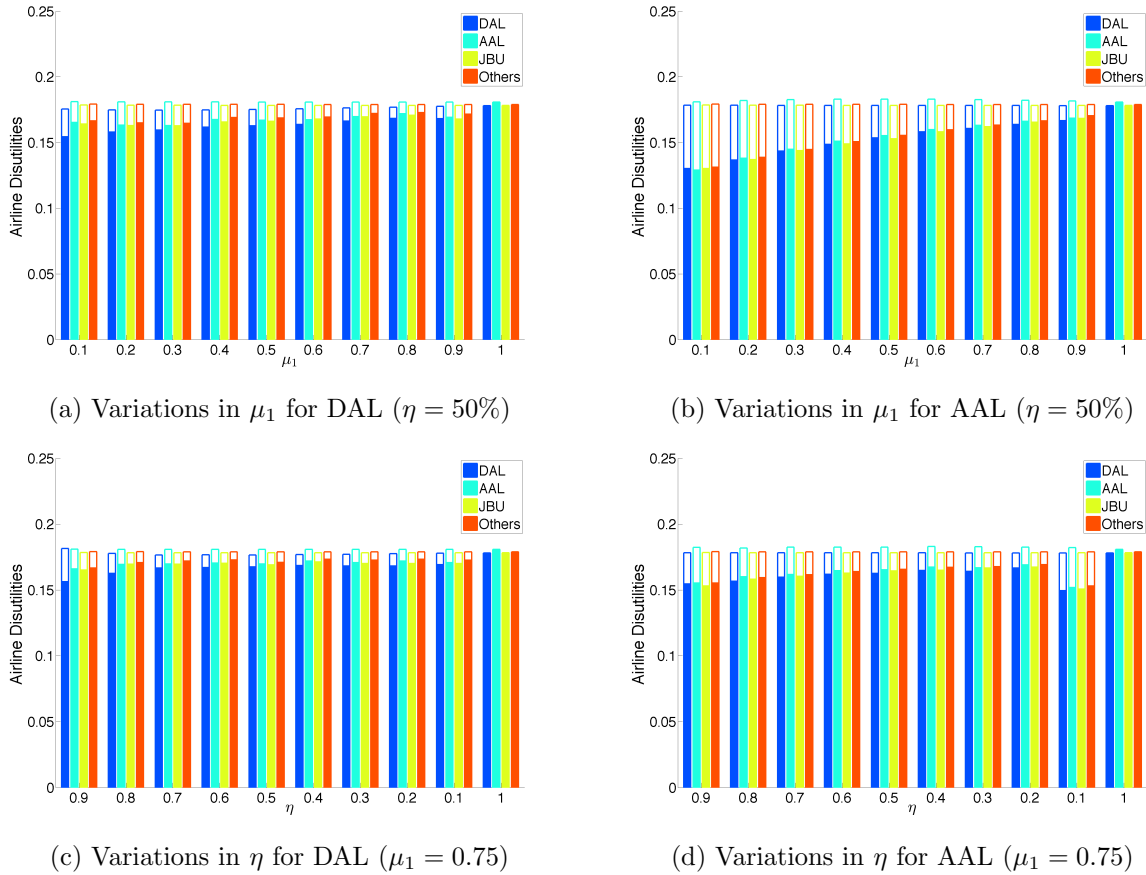


Figure 6-7: Airline utilities when flight valuations are considered (in full) vs. ignored

mission of network connections (i.e., $\beta = 0$). It optimizes scheduling interventions to attain optimal equity at minimal efficiency losses (i.e., it solves Problem $\widehat{\mathbf{P3}}(\rho^*)$). This mechanism is easily implementable as it is non-monetary, and considers only marginal scheduling adjustments that are consistent with the current practice at US airports. This mechanism is flexible as it does not rely on pre-determined schedule limits, but considers instead airline scheduling preferences as its starting point. This mechanism is equitable as it does not prioritize flights from any airline, and balances the scheduling adjustments fairly among the airlines. Finally, this mechanism is collaborative as it enables the airlines to provide their preferred schedules of flights, the network connections to maintain, and the scheduling flexibility of their flights in a simple and transparent way, and it accounts for these preferences in the modification of flight schedules. In turn, this mechanism optimizes scheduling interventions based on

system performance objectives (congestion mitigation) and on distributed airline objectives (satisfying airline scheduling preferences).

Mechanism 5 Proposed mechanism

The airlines submit their preferred schedules of flights $\rightarrow \mathcal{F}, \mathcal{F}^{\text{arr}}/\mathcal{F}^{\text{dep}}, \mathcal{F}_a, S^{\text{arr}}, S^{\text{dep}}$

The airlines submit the network connections to maintain $\rightarrow \mathcal{C}, \mathcal{C}_a, t^{\text{min}}, t^{\text{max}}$

Each airline $a \in \mathcal{A}$ receives a total of $\bar{v} |\mathcal{F}_a|$ credits

The airlines allocate their credits to their individual flights $\rightarrow v_i, \forall i \in \mathcal{F}$, subject to the constraint: $\sum_{i \in \mathcal{F}_a} v_i \leq \bar{v} |\mathcal{F}_a|, \forall a \in \mathcal{A}$

The central decision-maker solves Problem $\widehat{\mathbf{P3}}(\rho^*)$

6.6 Conclusion

We have developed, optimized and assessed mechanisms for airport scheduling interventions that build upon the methodology developed in Chapter 5. For the first time, these mechanisms incorporate inter-airline equity considerations in scheduling interventions and enable airline collaboration through the sharing of network connections and scheduling preferences. They rely on an original lexicographic modeling architecture that integrates on-time performance objectives into the decision-making framework for scheduling interventions, and optimizes scheduling interventions based on efficiency and inter-airline equity objectives.

The implementation of these scheduling interventions mechanisms has yielded three main results. First, under the standard paradigm that “a flight is a flight”, i.e., that all flights are equally inconvenient to reschedule, inter-airline equity can be achieved at no (or minimal) efficiency losses. Second, the proposed mechanism can maintain airlines’ network connections to enable aircraft, passengers, or crews to connect between flights. Third, the proposed mechanism can account for the relative scheduling flexibility of flights in determining which flights to reschedule, and which flights’ schedules to leave unchanged. In summary, the outcome of airport scheduling interventions could be improved by (i) ensuring inter-airline equity, and (ii) enabling the airline community to provide their scheduling preferences.

From a practical standpoint, the results from this chapter suggest that significant benefits

could be achieved through the creation of a collaborative environment between the airline community and the central decision-makers in charge of designing scheduling interventions (e.g., the Federal Aviation Administration (FAA) in the US or, outside the US, the airport schedule facilitators or coordinators). Under the proposed environment, the airlines would provide their scheduling preferences, the network connections to maintain, and the relative scheduling flexibility of their flights. The central decision-maker would then optimize scheduling interventions, based on efficiency, equity, and on-time performance objectives. Such a collaborative environment would produce delay-mitigating scheduling adjustments that are more consistent with airlines' scheduling preferences than achieved otherwise.

Chapter 7

Conclusion

In this thesis, we developed operating procedures and scheduling mechanisms to mitigate airport congestion. At airports where demand exceeds capacity and where capacity expansion is not feasible in the short- or medium term, congestion mitigation requires (i) improvements in the utilization of airport capacity, and/or (ii) scheduling interventions to allocate airport capacity to the airlines in ways that limit over-capacity scheduling. The former involves selecting runway configurations and balancing arrival and departure service rates to minimize congestion costs, for any given schedule of flights. It is a tactical decision, implemented over each day of operations. The latter involves producing schedules of flights that control the number of flights scheduled at peak hours, while minimizing interference with airline competitive scheduling. It is a strategic intervention, implemented months in advance of the day of operations. As detailed in Chapter 1, these two problems are strongly interdependent, but they had been treated separately in previous research. We developed the first integrated approach to airport congestion mitigation that jointly optimizes airport capacity utilization procedures and the design of airport capacity allocation mechanisms. In turn, it provides decision-making support for improving operating procedures and demand management policies at the busiest airports worldwide. In this last chapter, we summarize the results and contributions of the thesis, we discuss their implications for policy and practice, and we present several avenues for future research.

7.1 Summary of Research

Our approach, described in Chapter 2, started with a descriptive model of airport operations, namely a stochastic and dynamic $M(t)/E_k(t)/1$ queuing model. This model quantifies flight delays as a function of demand rates (determined by flight schedules) and service rates (constrained by airport capacity). In this thesis, we integrated this *descriptive* model of airport congestion into *decision-making* models of airport capacity utilization (that impact arrival and departure *service* rates) and of airport capacity allocation (that impact arrival and departure *demand* rates) for congestion mitigation. First, we addressed the problem of capacity utilization under operating stochasticity. Chapter 3 developed a decision-making model to optimize capacity utilization procedures, and Chapter 4 integrated it to a macroscopic model of airport congestion. Second, we addressed the problem of capacity allocation. Chapter 5 developed our modeling architecture that integrates the queuing model of airport capacity utilization and our model of airport congestion into a model of scheduling interventions. Chapter 6 extended this approach to incorporate inter-airline equity considerations and airline scheduling preferences into the design of scheduling mechanisms.

In Chapter 3, we developed a decision-making model that dynamically optimizes capacity utilization procedures under operating stochasticity. It bridges the gap between the literature on descriptive airport congestion modeling, which has extensively employed *stochastic* queuing models to describe airport operations, and the literature on the optimization of airport capacity utilization, which has mostly relied on *deterministic* queuing models of airport congestion. Given a schedule of flights and estimates of airport capacity, our model controls the selection of the runway configuration and the balance of arrival and departure service rates at the beginning of each 15-minute time period to minimize congestion costs. The control is exercised as a function of observed arrival and departure queue lengths, of the runway configuration in use, and of operating conditions. We implemented exact and approximate Dynamic Programming algorithms that, in combination, enable the real-time implementation of the model. Results showed that (i) optimal capacity utilization policies are path-dependent, i.e., depend on prior decisions and the prior stochastic evolution of the

system; (ii) our decision-making model can result in significant congestion costs savings, compared to advanced heuristics aimed to replicate decisions made in practice, (iii) our decision-making model provides significant improvements in capacity utilization policies, compared to an alternative approach based on deterministic queuing dynamics. Therefore, the integration of the stochasticity of airport operations into the decision-making framework can improve airport capacity utilization policies.

In Chapter 4, we approximated our Dynamic Programming model of airport capacity utilization that we integrated into a macroscopic queuing model of airport congestion. The resulting model quantifies in a computationally efficient way the relationship between flight schedules, airport capacity and flight delays at the strategic level, while accounting for the way airport capacity utilization procedures may vary at the tactical level to minimize congestion costs. It bridges the gap in the literature between models of airport capacity utilization, which adjusts arrival and departure service rates to best operate scheduled flights, and models of airport congestion, which typically consider single-value estimates of airport capacity that fail to capture the endogeneity of airport operating procedures with respect to flight schedules and airport capacity. The application of the model showed that it estimates quite accurately the average departure queue lengths, the variability of departure queue lengths and the average arrival and departure delays at JFK, EWR and LGA. Moreover, it suggested that the large delay reductions observed at these airports between 2007 and 2011 were, to a large extent, explained by the comparatively small changes in flight schedules over the period. In summary, we found that the model captures the magnitude of delays and their non-linear dynamics, as observed at three airports with very different capacity and scheduling patterns.

Chapters 3 and 4, combined, show that decisions of air traffic managers that affect arrival and departure service rates over each day of operations can have a significant impact on the dynamics and magnitude of arrival and departure queues at the airport. The optimization of such decisions is therefore of crucial importance to enhance the efficiency of airport operations. On the other hand, even the “optimal” capacity utilization may not be

sufficient to keep delays within reasonable bounds if the number of flights scheduled exceeds airport capacity by any significant margin. This motivated the consideration of scheduling interventions for congestion mitigation.

In Chapter 5, we developed the *Integrated Capacity Utilization and Scheduling Model (ICUSM)* that optimizes airport scheduling interventions, while accounting for the endogeneity of airport operations, i.e., the impact of scheduling changes on airport capacity utilization procedures and on airport congestion. This chapter bridges the gap between models of airport capacity utilization that optimize operating procedures for a given schedule of flights, and models of airport demand management that typically consider simplified operational settings. The model starts with the airlines' preferred schedule of flights, and modifies it in a way that (i) reduces the imbalances between airport demand and capacity to meet on-time performance objectives, (ii) eliminates no flight planned by the airlines, (iii) preserves the connectivity of aircraft and passenger itineraries, and (iv) minimizes the changes to flight schedules. It integrates an Integer Programming model of airport scheduling interventions, the Dynamic Programming model of airport capacity utilization developed in Chapter 3, and the Stochastic Queuing Model of airport congestion considered in this thesis. We developed a solution algorithm that iterates between these three models, until convergence. Computational results at JFK yielded two important results. First, substantial delay reductions can be achieved through limited interference with airlines' schedules of flights. We found that peak expected arrival and departure delays at JFK could be reduced by over 30% and 50%, respectively, by rescheduling approximately 20% of the flights at JFK by 15 or 30 minutes each. Second, our integrated approach to congestion mitigation was shown to provide significant benefits, compared to a typical sequential approach where flight schedules and airport operations are treated consecutively.

Motivated by these results, we extended the ICUSM in Chapter 6 to design the first scheduling interventions mechanisms that incorporate inter-airline equity objectives and airline scheduling preferences. First, we developed a multi-criteria modeling architecture that characterizes the trade-space between efficiency (defined as the mechanism's ability to meet

airline scheduling preferences), inter-airline equity (defined as the mechanism’s ability to balance scheduling adjustments fairly among the airlines), and on-time performance (defined as the mechanism’s ability to mitigate airport congestion). We formulated a lexicographic architecture that first fixes on-time performance targets, then maximizes efficiency under on-time performance constraints, and finally maximizes equity under on-time performance and efficiency constraints. We showed that under certain conditions, efficiency and equity can be jointly maximized, for given on-time performance targets, and discussed the conditions under which a trade-off between these two objectives may arise. We then designed non-monetary mechanisms that enable the airlines to submit their preferred schedule of flights, their network connections, and the relative scheduling flexibility of their flights through credit allocation. Computational results at JFK suggested that (i) inter-airline equity can be achieved at no, or small, losses in efficiency, (ii) network connections can be maintained through scheduling interventions, and (iii) accounting for the relative preferences of the airlines regarding the rescheduling of different flights can significantly improve the outcome of scheduling interventions.

In summary, Chapters 5 and 6 showed that scheduling interventions can mitigate airport congestion significantly through adjustments in airlines’ schedules of flights which are minimal, which consider the interdependencies between flight schedules and airport operations, which treat the airlines equitably, and which can account for the relative scheduling flexibility of different flights.

7.2 Practical Implications

The mitigation of airport congestion can be achieved through (i) capacity planning, i.e., the expansion of available airport capacity, (ii) capacity allocation, i.e., scheduling interventions to manage airport demand and limit over-capacity scheduling, and (iii) capacity utilization, i.e., improvements in airport operating efficiency. This thesis showed that these three interventions are strongly interdependent. Airport daily operations depend on available capacity and the schedule of flights to operate and, conversely, strategic decisions related to airport

capacity planning or demand management are motivated by the airport’s operating performance. The assessment of any congestion mitigation measure thus requires the comprehensive evaluation of its impacts on airport capacity, flight schedules, air traffic operations, and airport on-time performance.

There is a wide consensus in the US airline industry that capacity planning should be considered as the primary mechanism to accommodate demand growth and mitigate congestion at airports where infrastructure expansion is feasible [34, 22, 101]. As any infrastructure investment, airport capacity expansion plans should be assessed and weighed against alternatives based on factors such as capital costs, economic impacts and environmental impacts [78]. At many of the world’s busiest airports, however, capacity expansion is *not* feasible in the short- or medium-term because of the density of the surrounding urban areas and airspaces. As this thesis showed, significant delay reductions can be achieved even in the absence of infrastructure expansion through improvements in operating and scheduling practices *given airport capacity constraints*.

First, improvements in airport capacity utilization can mitigate congestion through better use of infrastructural resources to operate scheduled flights. In this thesis, we focused on the utilization of runway capacity, i.e., the dynamic selection of runway configuration and the balancing of the arrival and departure loads operated at busy airports. In combination with other Air Traffic Control and Air Traffic Flow Management initiatives, these interventions can reduce the magnitude of air traffic delays and thus improve the environmental and economic performance of air transportation systems. In contrast with capacity expansion, these improvements involve small costs of implementation. If designed effectively, operational enhancements require minimal infrastructural investments, maintain safety levels, and reduce air traffic controller workload [108]. They should therefore be considered systematically to “push the envelope” of airport operations.

In instances where operating enhancements are insufficient to prevent delays from reaching undesirable levels, scheduling interventions may be required to allocate airport capacity to the airlines in a way that reduces over-capacity scheduling. In contrast with capacity plan-

ning and utilization, these interventions do not affect airport supply, but airport demand, and thus interfere with airlines' scheduling decisions. For this reason, they should be used only when airport demand exceed capacity by a significant margin, and when there is little prospect for quantum capacity increases or improvements in operating efficiency. In these instances, this thesis showed that significant improvements in airport on-time performance can be achieved through minimal scheduling interventions that involve only temporal shifts in demand (i.e., the rescheduling of flights to distribute demand more evenly over the day of operations), and no reduction in the overall demand (i.e., no reduction in the total number of flights scheduled in a day). This result has important implications for the design of demand management policies at busy airports worldwide.

In the United States, the consideration of scheduling interventions at congested airports is recommended. We have been collaborating with the Port Authority of New York and New Jersey (PANYNJ) to develop for LaGuardia Airport a scheduling procedure based on the methodology described in this thesis. At the majority of airports, where no schedule limit is in place, any decision regarding the implementation of scheduling interventions should be based on the delay reductions that could be achieved through limited interference with airline competitive scheduling. At the airports where some scheduling interventions are in place, the formalization and optimization of the scheduling process is recommended. A formal process for scheduling interventions would involve the airlines providing their scheduling preferences to a central decision-maker (e.g., the FAA, in collaboration with airport operators), who would then suggest some minor scheduling adjustments to reduce demand-capacity imbalances. Such adjustments would consist of rescheduling 5% to 20% of the flights at the airport to later or earlier times by no more than 30 minutes each (in the great majority of cases, by 15 minutes). They would eliminate no flight, maintain all aircraft and passenger connections, and treat the airlines equitably. This approach would leave the determination of the total number of flights scheduled at the airport up to the airlines, and be limited to the exact timing of the flights. The large delay reductions that could be achieved would lower airline operating costs, improve passenger welfare, and enhance airport attractiveness. The

success of this approach would require its implementation in a collaborative and transparent manner and the engagement of the FAA, the airline community and airport operators. The success of the Collaborative Decision Making (CDM) paradigm in Air Traffic Flow Management provides a good precedent for enhancing collaboration between stakeholders at a more strategic level through modest scheduling interventions at selected airports.

At schedule-coordinated airports located outside the United States, there may exist opportunities for improving the efficiency of the capacity allocation processes. The reliance on flat, or nearly-flat, “declared capacities” set around airports’ IMC capacities may result in access restrictions and in the under-utilization of airport infrastructure at peak-hours (especially under good weather conditions). Instead, a more flexible approach to capacity allocation is recommended. The approach would start with some scheduling inputs provided by the airlines, and then determine a schedule of flights based on these scheduling preferences as well as capacity patterns at the airport. It would integrate on-time performance objectives in the determination of the number of flights scheduled per time period, instead of relying on an administratively determined value of declared capacity. It would also permit variations in the number of flights scheduled from one time period of the day to another, consistently with intra-day variations in airline scheduling preferences and passenger demand. Similarly to the US case, data exchanges and communication between schedule coordinators and the airline community are recommended to improve the outcome of such schedule coordination.

7.3 Future Research

The potential of operating and scheduling interventions for air traffic congestion mitigation motivates further research in this area. This section identifies some of the most promising directions to extend this research and its findings.

Airport Capacity Utilization

The decision-making model of capacity utilization developed in Chapter 3 relied on a number of modeling assumptions. An important extension would involve testing the robustness

of our operating policies with respect to the underlying models of airport operations that have been considered. For instance, we could relax the characterization of the airport as an $M(t)/E_k(t)/1$ queuing system [84, 105], integrate more realistic models of airport weather conditions [74, 27], and consider alternative representations of the costs of runway configurations changes but a time period of idleness [120]. More broadly, the use of ground surveillance data could improve the representation of airport operations in our decision-making approach to congestion mitigation. Another extension would regard integrating, in addition to expected delay objectives, delay variability objectives into the design of control strategies in order to reduce delay variability and improve schedule reliability.

In addition, the scope of this work could be extended in two major ways. First, the integration of recent empirical studies analyzing the runway configuration selection process based on discrete choice models into the design of control strategies could compare systematically decisions recommended by our model to actual decisions made in practice, quantify decision-making models' benefits in a more comprehensive way than done in Chapter 3, and improve the realism of our modeling approach. Second, our approach focused on the utilization of the capacity of an airport's runway system. In future research, this could be integrated with other air traffic management initiatives such as Ground Delay Programs and traffic metering to improve the design of dynamic operating procedures. An important challenge lies in the consideration of the stochasticity of air traffic operations in the optimization of such procedures. While computationally challenging, this research direction holds the potential to unlock further reductions in congestion costs, and thus further improvements in the operational, economic and environmental performance of air transportation systems.

Airport Capacity Allocation

In Chapters 5 and 6, we quantified the benefits of scheduling interventions by the reductions in flight delays at the airport considered, and their costs by the displacement from the airlines' preferred schedule of flights, respectively. In fact, the benefits and costs of such mechanisms include other dimensions that were left out of scope of this research. On the benefits side,

a natural extension would integrate descriptive models of network-wide congestion [95] and descriptive models of passenger delays [11] into our decision-making approach to airport scheduling interventions. On the costs side, future research could combine our model of scheduling interventions with models of airline schedule design (see, e.g., [75, 10]) and models of passenger welfare [69, 114]. These efforts would provide more comprehensive assessments of the effects of airport scheduling interventions on air transportation stakeholders.

A related avenue for future research would consist of integrating models of airline scheduling into our frameworks of scheduling interventions to better model airlines' scheduling inputs into the proposed mechanisms. Such models of airlines' profit-maximizing behaviors would also create the opportunity to analyze strategic interactions among the airlines and the potential for gaming under the proposed mechanisms. A first direction could investigate the potential for strategic behaviors in the submission of airline scheduling preferences. A second direction could build upon recent studies in the context of Air Traffic Flow Management that have proposed mechanisms to gather airlines' operating preferences and to integrate them into the broader decisions related to the duration, magnitude and scope of the ATFM initiatives [113, 6, 37]. Similar mechanisms could be applicable to the design of scheduling interventions at the strategic level (e.g., to coordinate decisions such as the number of flights to reschedule and the corresponding delay reductions to expect (see Figure 5-5)).

The integrated modeling architecture developed in this thesis could also be extended to systematically compare different airport capacity allocation mechanisms. First, additional practical considerations or policy requirements could be easily integrated into the proposed frameworks, e.g., slot grandfathering, pool of new entrants, secondary trading, etc. Second, market-based mechanisms based on congestion pricing and slot auctions could be modeled using a similar framework of flight scheduling and airport operations. The combination of the integrated framework of scheduling interventions and airport operations developed in this thesis and economic models of demand management would lead to better understanding of the effects of various capacity allocation mechanisms on airport operations, airline competition, and social welfare.

Toward an Integrated Approach to Infrastructure Management

The interdependencies between airport capacity, flight scheduling and airport operations motivate comprehensive approaches to the management of air transportation systems. Exciting research opportunities lie in the extension of the frameworks developed in this thesis to study several interventions that were been left out of its scope and modeled here as exogenous parameters. Two promising areas are infrastructure expansion and the system-wide deployment of new aircraft and air traffic control technologies. These two interventions could increase system capacity (e.g., by enabling more simultaneous operations, or by reducing the separation requirements between aircraft), or improve air traffic handling procedures (e.g., by reducing their stochasticity). They would, in turn, impact the need for airport scheduling interventions (i.e., capacity allocation) and Air Traffic Control and Air Traffic Flow Management decisions (i.e., capacity utilization). The combination of our operational and scheduling frameworks with approaches quantifying the impact of added infrastructure or new technologies on airport operations, based on system simulations [8] or empirical analyses [104], could assess their ultimate benefits on airport on-time performance, airline economics and passenger welfare. This could guide decisions regarding investments in airport infrastructure and the design of incentives for technology adoption, and compare the effectiveness of alternative interventions for congestion mitigation. In turn, this would provide a holistic approach to the design, management and operation of air transportation systems over the course of their lifecycles.

Finally, the problems of capacity planning, allocation, and utilization analyzed in the air transportation context are common to other complex, large-scale, and multi-stakeholder infrastructure systems such as ground transportation systems [56, 91] power systems [65, 92], telecommunications systems [71], etc. The performance of such systems also depends on infrastructure design, on the mechanisms employed to allocate scarce capacity among competing users and operators, and on the efficiency of day-to-day operations. Therefore, many research questions and methods are similar across these systems, but, at the same time, engineering and institutional particularities require the design of domain-specific approaches.

Looking ahead, the development of integrated frameworks and methodologies holds considerable promise to improve the planning, allocation, and utilization of infrastructure capacity across industrial domains and to ultimately promote more efficient, more reliable and more sustainable engineering systems in various infrastructure-intensive industries.

Appendix A

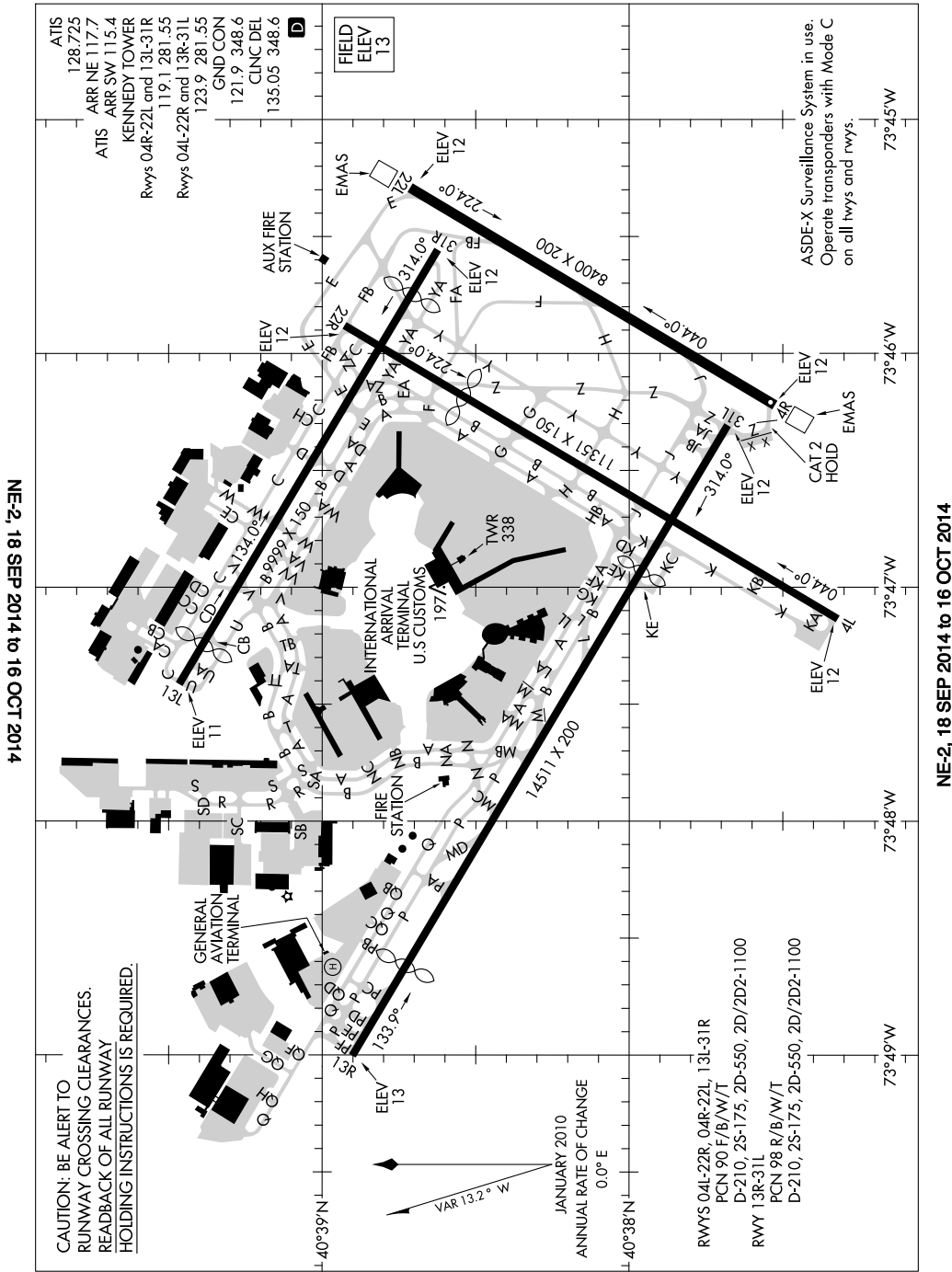
Airport Diagrams

14261

AIRPORT DIAGRAM

AL-610 (FAA)

JOHN F. KENNEDY INTL (JFK)
NEW YORK, NEW YORK



AIRPORT DIAGRAM

14261

NEW YORK, NEW YORK
JOHN F. KENNEDY INTL (JFK)

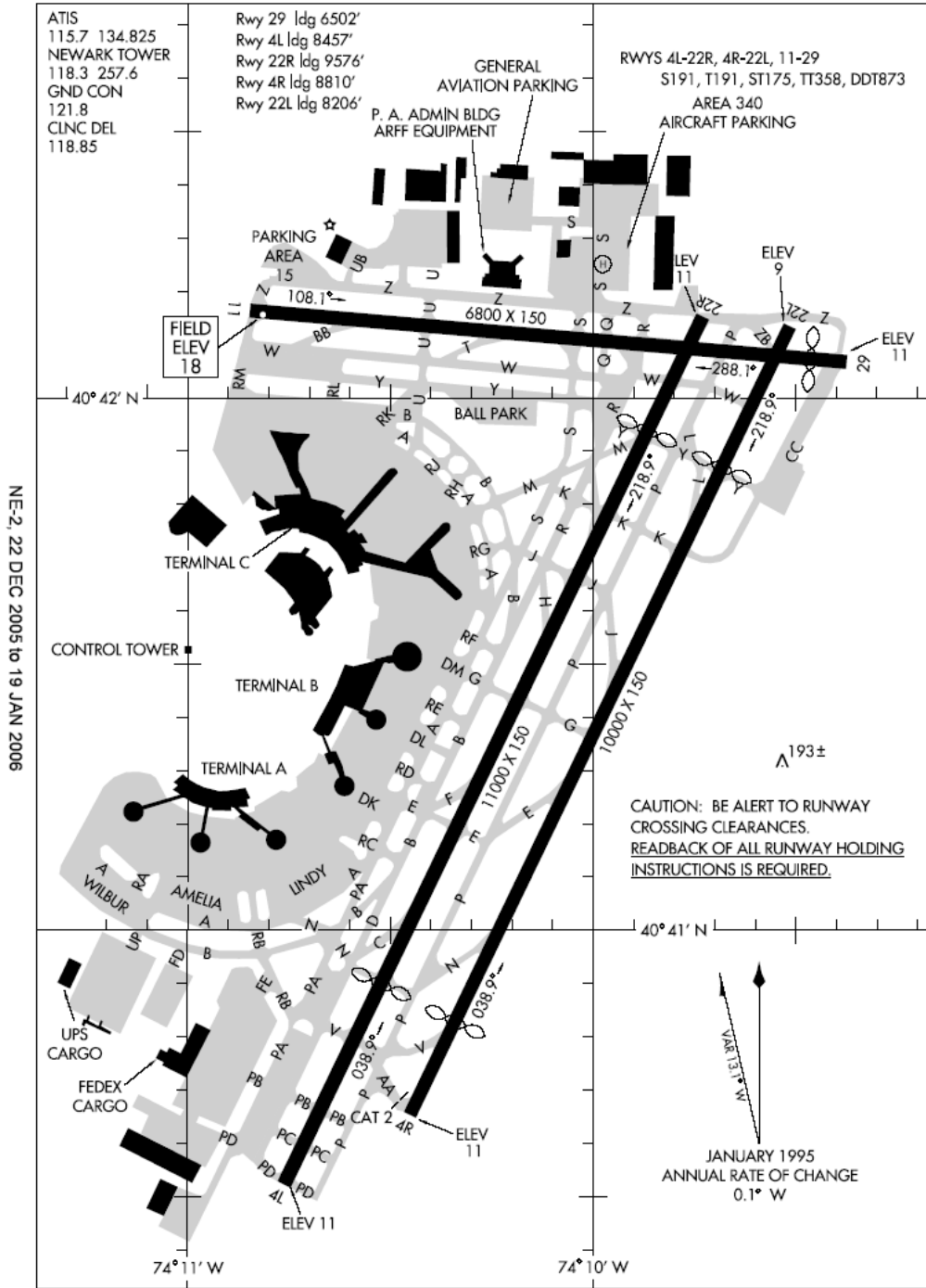
Figure A-1: Airport Diagram: JFK

05300

AIRPORT DIAGRAM

AL-285 (FAA)

NEWARK LIBERTY INTL (EWR)
NEWARK, NEW JERSEY



NE-2, 22 DEC 2005 to 19 JAN 2006

NE-2, 22 DEC 2005 to 19 JAN 2006

AIRPORT DIAGRAM

05300

NEWARK, NEW JERSEY
NEWARK LIBERTY INTL (EWR)

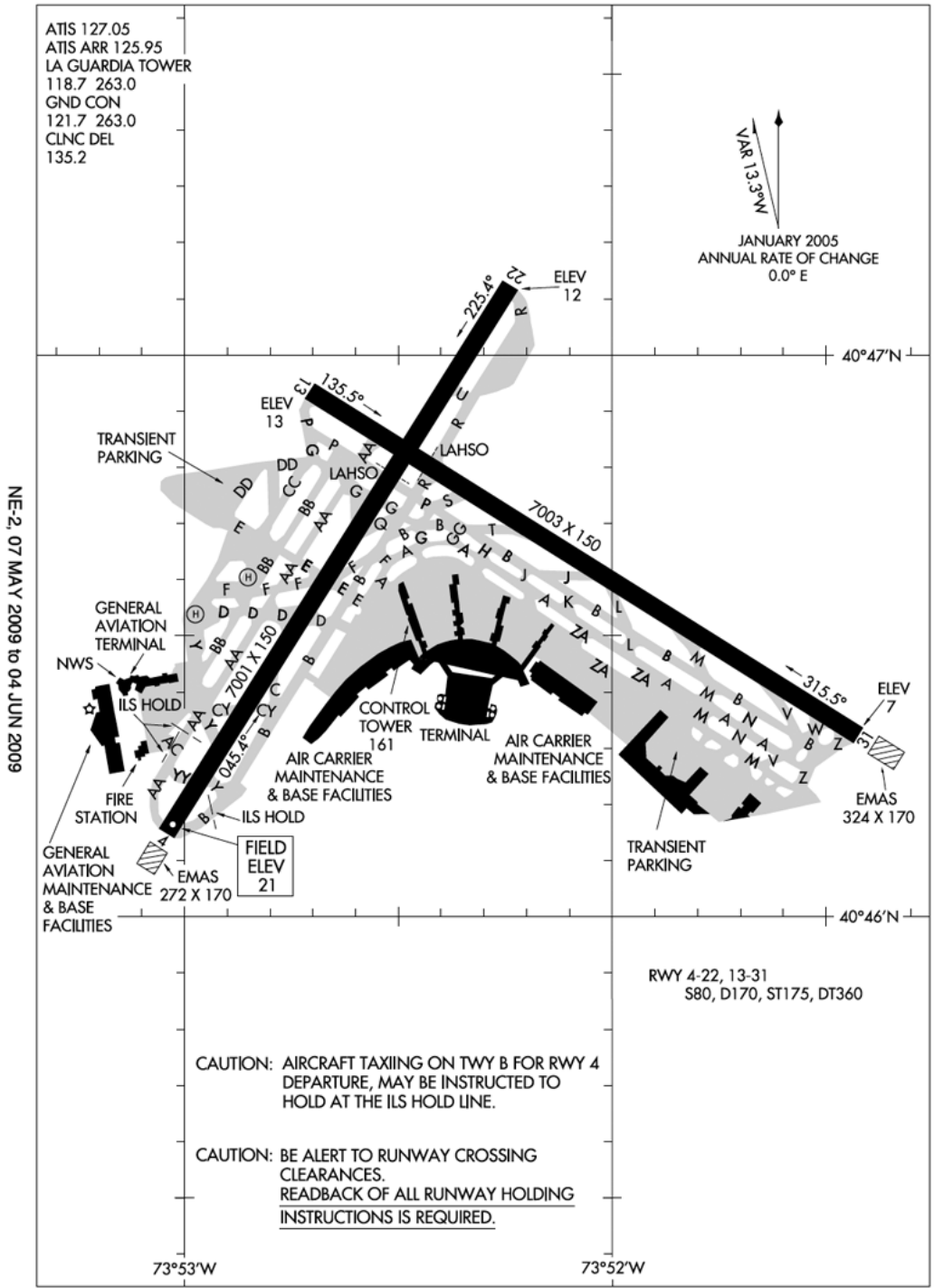
Figure A-2: Airport Diagram: EWR

09071

AIRPORT DIAGRAM

AL-289 (FAA)

NEW YORK/LA GUARDIA (LGA)
NEW YORK, NEW YORK



NE-2, 07 MAY 2009 to 04 JUN 2009

NE-2, 07 MAY 2009 to 04 JUN 2009

AIRPORT DIAGRAM

09071

NEW YORK, NEW YORK
NEW YORK/LA GUARDIA (LGA)

Figure A-3: Airport Diagram: LGA

Appendix B

Tail Number Reconstruction

The scheduling and operating data used in this thesis were obtained from the Aviation System Performance Metrics (ASPM) database. This database, maintained by the Federal Aviation Administration (FAA), provides extensive information on most commercial flights operated in the United States. Among other metrics, it reports, for each flight taking off or landing at the main airports in the United States, its origin and destination, its scheduled takeoff and landing times, the airline operating the flight, the type of aircraft used (*e.g.*, A320, B777, etc.), the flight number and the aircraft tail number.

However, the aircraft tail number is missing for approximately 40% of the flights reported in the database, including all international flights. This information is needed to identify aircraft connections and to define the related constraints in our models of airport scheduling interventions developed in Chapters 5 and 6. For this reason, we implemented a simple Tail Number Reconstruction Model that, given available information in the ASPM database, imputes missing tail number data by reconstructing aircraft itineraries. It is based on the models developed in [23]

Let us consider an airline, denoted by AL , and an aircraft type, denoted by AC . From the information available in the ASPM database, we identify the set of flights operated by airline AL and flown by aircraft type AC during a given week of operations. We denote this set by \mathcal{F} . The Tail Number Reconstruction Model determines aircraft itineraries that minimize

the number of aircraft used to operate the set of flights in \mathcal{F} . It is a simplified version of the extensively studied Fleet Assignment Model (see, *e.g.*, [52, 75]), applied to the set of flights operated by airline AL with an aircraft type AC . In other words, the assignment of aircraft types to the scheduled flights is assumed to have already been performed by the airline and we determine the routing of aircraft that minimizes the aircraft count.

We introduce the following two parameters. In the definitions below, a connection between flight i and flight j means that flights i and j are flown by the same aircraft and that flight j is the immediate successor of flight i .

$$A_{ij}^{\text{TNR}} = \begin{cases} 1 & \text{if a connection between flight } i \text{ and flight } j \text{ is feasible} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{ij}^{\text{TNR}} = \begin{cases} 1 & \text{if a connection between flight } i \text{ and flight } j \text{ was actually planned} \\ 0 & \text{otherwise} \end{cases}$$

A connection between flights i and j is assumed to be “feasible” if flight j departs from the airport where flight i arrived and if the difference between flight j ’s departure time and flight i ’s arrival time is larger than the minimum time required for the aircraft to turn around. Minimum turnaround times are obtained from [94]. A connection between flights i and j is identified as “actually planned” if the tail numbers of flights i and j are reported in the ASPM database, if they are identical to each other and if the difference between flight j ’s departure time and flight i ’s arrival time is small enough to ensure that the aircraft has not flown other flights between flight i and flight j .

We introduce the following binary variable:

$$z_{ij}^{\text{TNR}} = \begin{cases} 1 & \text{if a connection between flight } i \text{ and flight } j \text{ is assigned by the model} \\ 0 & \text{otherwise} \end{cases}$$

The number of aircraft used, denoted by N , is equal to the number of trips minus the

number of aircraft connections z , i.e.:

$$N + \sum_{i=1}^n \sum_{j=1}^n z_{ij} = |\mathcal{F}|$$

To see this, we partition the set of flights \mathcal{F} into subsets of flights flown by the same aircraft. For each aircraft, the number of flights it is assigned to is equal to the number of connections between these flights plus 1. For instance, if an aircraft is assigned to three flights in a given period of time, we count two connections between these three flights. By summing over all aircraft, we obtain the relationship above.

As a result, minimizing fleet size is equivalent to maximizing the number of connections between flights (since the total number of flights $|\mathcal{F}|$ is fixed). The resulting problem is formulated as follows:

$$\max \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{F}} z_{ij}^{\text{TNR}} \tag{B.1}$$

$$\text{st } z_{ij}^{\text{TNR}} \leq A_{ij}^{\text{TNR}} \quad \forall i, j \in \mathcal{F} \tag{B.2}$$

$$z_{ij}^{\text{TNR}} \geq B_{ij}^{\text{TNR}} \quad \forall i, j \in \mathcal{F} \tag{B.3}$$

$$\sum_{i \in \mathcal{F}} z_{ij}^{\text{TNR}} \leq 1 \quad \forall j \in \mathcal{F} \tag{B.4}$$

$$\sum_{j \in \mathcal{F}} z_{ij}^{\text{TNR}} \leq 1 \quad \forall i \in \mathcal{F} \tag{B.5}$$

$$z_{ij}^{\text{TNR}} \in \{0, 1\} \quad \forall i, j \in \mathcal{F} \tag{B.6}$$

Equation (B.1) maximizes the number of connections, thus minimizes fleet size. Constraint (B.2) states that there cannot be a connection between any pair of flights if the connection is not feasible. Constraint (B.3) maintains the connections that were actually planned. Constraint (B.4) (resp. Constraint (B.5)) states that any flight can be immediately preceded (resp. followed) by at most one flight.

For each day of operations considered (e.g., 05/25/2007), we applied the Tail Number Reconstruction model over the week spanning from three days before to three days after (e.g.,

between 05/22/2007 and 05/28/2007). We reconstructed the networks of flights operated by all airlines that operated at least 1% of all flights scheduled at the airport considered (e.g., JFK) on the day considered. For each of these airlines, we consider the aircraft types that operated at least 100 flights during the week considered. We applied the Tail Number Reconstruction Model for each of these airlines and for each of these aircraft types during the considered week. We reconstructed the remaining itineraries using a manual procedure similar to the one outlined by Pyrgiotis [94].

This procedure optimizes the routing of aircraft to minimize the number of aircraft used over a week of operations. It is thus expected to approximate the fleet assignment decisions made by the airlines. It is likely, however, that the resulting routing will not match exactly the one that was used by the airlines on the considered day. For instance, considerations such as aircraft maintenance routing and crew scheduling are not included in the model. Nonetheless, the procedure is quite conservative: since it maximizes the number of aircraft connections, it constrains the models of scheduling interventions as much as possible. Therefore, the solution to our models of scheduling interventions with any alternative routing of aircraft is expected to induce the same, or a smaller, displacement than the solution that we obtain with this optimized routing.

Appendix C

Scheduling Interventions at LGA

In Chapter 5, we developed the Integrated Capacity Utilization and Scheduling Model (ICUSM) and applied it to JFK Airport. The model finds a schedule of flights that meets on-time performance objectives, while minimizing the displacement from the airlines' preferred schedule. Results shown in Chapter 5 showed that significant delay reductions could be achieved at JFK through limited changes in flight schedules. In this appendix, we apply the same model at LaGuardia Airport (LGA). This application is motivated by two main factors. From a modeling standpoint, existing flight schedules are much evenly distributed over the day at LGA than at JFK. This motivates the extent to which delays could be reduced through smoothing the demand further. From a practical standpoint, LGA is undergoing a Redevelopment Program that includes a number of projects aimed to enhance the airport's ability to serve airlines and passengers. In this context, interactions with the Port Authority of New York and New Jersey (PANYNJ) have suggested that the potential for implementation of scheduling interventions in the short term is likely to be greater at LGA than at JFK.

Based on interactions with the PANYNJ and the major airlines operating at LGA, we have selected the day of July 17, 2014 as a representative busy Summer day for the analysis. Figure C-1 shows the schedule of flights on 07/17/2014 at LGA, aggregated per hour (Figure C-1a) and per 15-minute period (Figure C-1b). As expected, the number of flights

scheduled is relatively constant over the day at LGA. The hourly schedule falls within the range of capacity values at LGA, as a result of the imposition of the flight caps since the phase-out of the High Density Rule. However, the schedule exhibits significant variability on a 15-minute basis, with most of the flights being concentrated at XX:00 and XX:30. In turn, above-capacity scheduling occurs even under optimal conditions (e.g., VMC) when one looks at the 15-minute schedule. Similarly to JFK, these imbalances between scheduling and capacity are even greater when capacity is sub-optimal (e.g., in IMC).

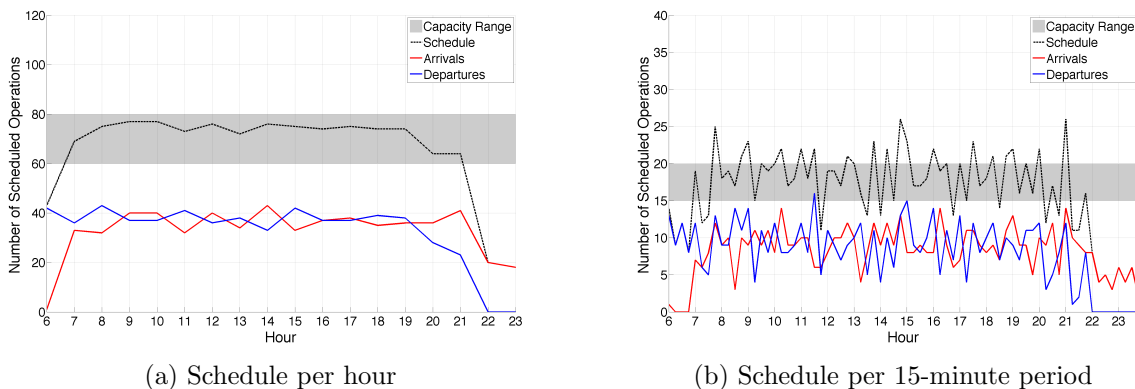


Figure C-1: Flight schedule at LGA on 07/17/2014

Figure C-2 shows the impact of scheduling interventions on the schedule of flights, obtained with expected queue length limits equal to $A_{\text{MAX}} = 9$ and $D_{\text{MAX}} = 18$. Similarly to the case of JFK, the modified schedule is smoother than the original schedule of flights produced by the airlines on 07/17/2014. Note that the arrival schedule is almost evenly distributed over the day. In contrast, the departure schedule still exhibits variations from one 15-minute period to another. This confirms the insights obtained in the case of JFK that the model can maintain peaks and valleys in the schedule of flights, consistently with airline scheduling preferences.

Table C.1 defines four sets of expected queue length limits A_{MAX} and D_{MAX} , and reports the associated displacement and expected delay reductions. Figure C-3 shows the corresponding expected queue length over the course of the day of operations. Note that significant delay reductions can be achieved through limited displacement in the airline schedule

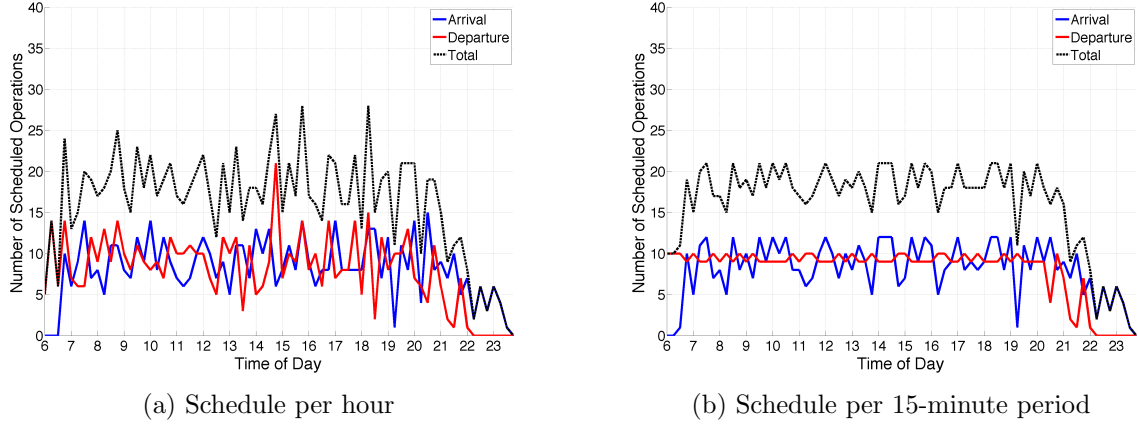


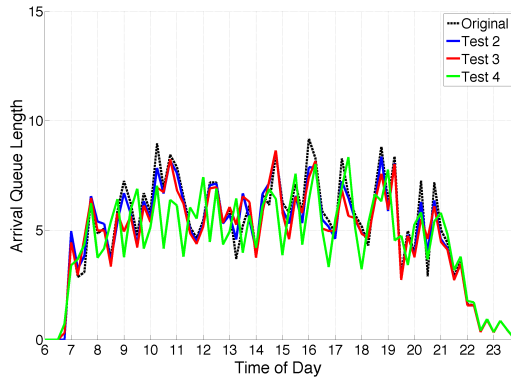
Figure C-2: Original and modified schedules at LGA

of flights. All the expected queue length limits considered can be reached without displacing any flight by more than 15 minutes, and by displacing 100-400 flights in total. On the delay side, this induces reductions in peak expected departure queue lengths by 20%-30% and reductions in average departure delays by 10%-25%. Similarly to the JFK results shown in Chapter 5, these delay reductions are achieved by smoothing the demand, which results in delays forming earlier in time, but increasing at a more controlled rate. Finally, the impact on arrival delays is positive, but smaller in magnitude than the impact on departure delays.

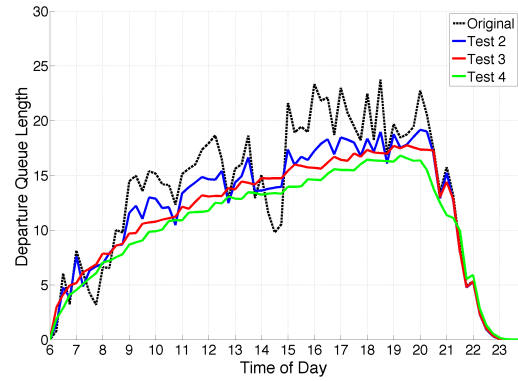
Table C.1: Displacement and delays at LGA for different expected queue length targets

Test	Targets		Optimal Disp.		Peak Queue Lengths		Average Delay	
	A_{MAX}	D_{MAX}	δ^* (periods)	Δ^* (periods)	Arrival (A/C)	Departure (A/C)	Arrival (mins.)	Departure (mins.)
1	∞	∞	0	0	9.2	23.8	9.8	26.1
2	9	20	1	95	8.6	19.2	9.4	22.9
3	9	18	1	198	8.6	17.7	9.2	21.5
4	8.5	17	1	421	8.3	16.8	8.9	19.8
					(-5.8%)	(-19.4%)	(-4.0%)	(-12.2%)
					(-5.8%)	(-25.3%)	(-6.7%)	(-17.4%)
					(-9.1%)	(-29.2%)	(-9.2%)	(-24.1%)

In summary, we find that the approach to airport scheduling interventions developed in this thesis can also provide significant benefits at LGA. This underscores that many airports,



(a) Arrival queue lengths



(b) Departure queue lengths

Figure C-3: Expected queue lengths at LGA before and after the scheduling interventions

even with different scheduling patterns, may benefit from the type of scheduling interventions considered in this thesis if scheduling exhibits significant variability on a 15-minute basis. The approach developed in Chapters 5 and 6 can be applied to different airports based on their scheduling and capacity patterns, and quantify the potential delay reductions as a function of the schedule displacement.

Bibliography

- [1] Airlines for America. Annual Results U.S. Airlines. Accessed February 12, 2015. Available at: <http://airlines.org/data/>, 2015.
- [2] I. Ater. Internalization of Congestion at US Hub Airports. *Journal of Urban Economics*, 72(2-3):196–209, 2012.
- [3] H. Balakrishnan and B. Chandran. Algorithms for Scheduling Runway Operations Under Constrained Position Shifting. *Operations Research*, 58(6):1650–1665, 2010.
- [4] M. Ball, L. Ausubel, F. Berardino, P. Cramton, G. Donohue, M. Hansen, and K. Hoffman. Market-based Alternatives for Managing Congestion at New York’s LaGuardia Airport. In *AirNeth Annual Conference*, 2007.
- [5] M. Ball, C. Barnhart, M. Dresner, M. Hansen, K. Neels, A. Odoni, E. Peterson, L. Sherry, A. Trani, and B. Zou. Total Delay Impact Study. Technical report, National Center of Excellence for Aviation Operations Research, 2010.
- [6] M. Ball, C. Barnhart, M. Hansen, L. Kang, Y. Liu, P. Swaroop, V. Vaze, and C. Yan. Distributed Mechanisms for Determining NAS-Wide Service Level Expectations: Final Report. Technical report, NEXTOR II, 2014.
- [7] M. Ball, C. Barnhart, G. Nemhauser, and A. Odoni. Air Transportation: Irregular Operations and Control. In C. Barnhart and G. Laporte, editors, *Handbook in Operations Research & Management Science*, volume 14, pages 1–67. Elsevier, 2007.
- [8] A. Barnett, M. Ball, G. Donohue, M. Hansen, A. Odoni, and A. Trani. Los Angeles International Airport North Airfield Safety Study. Technical report, North Airfield Safety Advisory Committee, 2010.
- [9] C. Barnhart, D. Bertsimas, C. Caramanis, and D. Fearing. Equitable and Efficient Coordination in Traffic Flow Management. *Transportation Science*, 46(2):262–280, 2012.
- [10] C. Barnhart, A. Farahat, and M. Lohatepanont. Airline Fleet Assignment with Enhanced Revenue Modeling. *Operations Research*, 57(1):231–244, 2009.

- [11] C. Barnhart, D. Fearing, and V. Vaze. Modeling Passenger Travel and Delays in the National Air Transportation System. *Operations Research*, 62(3):580–601, 2014.
- [12] L. Basso and A. Zhang. Pricing vs. Slot Policies when Airport Profits Matter. *Transportation Research Part B: Methodological*, 44:381–391, 2010.
- [13] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.
- [14] D. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, 3rd edition, 2005.
- [15] D. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.
- [16] D. Bertsimas, V. Farias, and N. Trichakis. The Price of Fairness. *Operations Research*, 59(1):17–31, 2011.
- [17] D. Bertsimas, V. Farias, and N. Trichakis. On the Efficiency-Fairness Trade-Off. *Management Science*, 58(12):2234–2250, 2012.
- [18] D. Bertsimas, M. Frankovich, and A. Odoni. Optimal Selection of Airport Runway Configurations. *Operations Research*, 59(6):1407–1419, 2011.
- [19] D. Bertsimas and S. Gupta. Fairness and Collaboration in Network Air Traffic Flow Management: An Optimization Approach. *Transportation Science*, Articles in advance, 2015.
- [20] D. Bertsimas and S. Stock Patterson. The Air Traffic Flow Management Problem with Enroute Capacities. *Operations Research*, 46(3):406–422, 1998.
- [21] K. Bilimoria, B. Sridhar, G. Chatterji, K. Sheth, and S. Grabbe. FACET: Future ATM Concepts Evaluation Tool. In *3rd USA/Europe Air Traffic Management R&D Seminar*, 2000.
- [22] P. Bonnefoy. *Scalability of the Air Transportation System and Development of Multi-Airport Systems: A Worldwide Perspective*. PhD thesis, Massachusetts Institute of Technology, 2008.
- [23] P. Bonnefoy, A. Jacquillat, and A. Fan. Investigation of Benefits and Implications from Fleet of Aircraft with Alternative Mission (Range) Specifications. In *8th Triennial Symposium on Transportation Analysis*, 2013.
- [24] J. Brueckner. Airport congestion when carriers have market power. *American Economic Review*, 92(5):1357–1375, 2002.
- [25] J. Brueckner. Price vs. quantity-based approaches to airport congestion management. *Journal of Public Economics*, 93:681–690, 2009.

- [26] Bureau of Transportation Statistics. TranStats. Accessed March 26, 2013. www.transtats.bts.gov, 2013.
- [27] G. Buxi and M. Hansen. Generating Probabilistic Capacity Profiles from Weather Forecast: A Design-of-experiment Approach. In *9th USA/Europe Air Traffic Management R&D Seminar*, 2011.
- [28] A. Carlin and P. Park. Marginal Cost Pricing of Airport Runway Capacity. *American Economic Review*, 60(3):310–319, 1970.
- [29] A. Churchill, D. Lovell, A. Mukherjee, and M. Ball. Determining the Number of Airport Arrival Slots. *Transportation Science*, 47(4):526–541, 2012.
- [30] A. Czerny. Airport Congestion Management under Uncertainty. *Transportation Research Part B: Methodological*, 44:371–380, 2010.
- [31] A. Czerny, P. Forsyth, D. Gillen, and H. Niemeier. *Airport Slots*. Ashgate Publishing Ltd, 2008.
- [32] J. Daniel. Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues. *Econometrica*, 62(2):327–370, 1995.
- [33] J. Daniel and Harback. (When) Do Hub Airlines Internalize their Self-imposed Congestion Delays? *Journal of Urban Economics*, 63(2):583–612, 2008.
- [34] R. de Neufville and A. Odoni. *Airport Systems: Planning, Design and Management*. McGraw-Hill, 2nd edition, 2013.
- [35] P. Dell’Olmo and G. Lulli. A Dynamic Programming Approach for the Airport Capacity Allocation Problem. *IMA Journal of Management Mathematics*, 14(3):249, 2003.
- [36] G. Douglas and J. Miller. Economic regulation of domestic air transport: theory and economic regulation of domestic air transport: Theory and policy. Technical report, Brookings Institution, 1974.
- [37] A. Evans, V. Vaze, and C. Barnhart. Airline-Driven Performance-Based Air Traffic Management: Game Theoretic Models and Multicriteria Evaluation. *Transportation Science*, 2015.
- [38] T. Fan. *Market-based Airport Demand Management - Theory, Model and Applications*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [39] T. Fan and A. Odoni. Airport Demand Management: A Practical Perspective. *Air Traffic Control Quarterly*, 10(3):285–306, 2002.
- [40] Federal Aviation Administration. Airport Design. Advisory Circular 150/5300-13A, 2012.

- [41] Federal Aviation Administration. Aviation System Performance Metrics (ASPM) database. Accessed April 4, 2013. Available at: <https://aspm.faa.gov/apm/sys/main.asp>, 2013.
- [42] Federal Aviation Administration. NextGen Implementation Plan. Technical report, 2014.
- [43] Federal Aviation Administration. Operational Network (OPSNET) database. Accessed February 12, 2015. Available at: <https://aspm.faa.gov/opsnet/sys/main.asp>, 2015.
- [44] E. Feron, J. Hansman, A. Odoni, R. Cots, B. Delcaire, X. Feng, W. Hall, H. Idris, A. Muharremoglu, and N. Pujet. The Departure Planner: A Conceptual Discussion. Technical report, International Center for Air Transportation, MIT, 1997.
- [45] S. George, G. Satapathy, V. Manikonda, F. Wieland, M. Refai, and R. Dupee. Build 8 of the Airspace Concept Evaluation System. In *AIAA Modeling and Simulation Technologies Conference*, 2011.
- [46] E. Gilbo. Airport Capacity: Representation, Estimation, Optimization. *IEEE Transactions on Control Systems Technology*, 1(3):144–154, 1993.
- [47] E. Gilbo. Optimizing Airport Capacity Utilization in Air Traffic Flow Management subject to Constraints at Arrival and Departure Fixes. *IEEE Transactions on Control Systems Technology*, 5(5):490–503, 1997.
- [48] E. Gilbo and K. Howard. Collaborative Optimization of Airport Arrival and Departure Traffic Flow Management Strategies for CDM. In *3rd USA/Europe Air Traffic Management R&D Seminar*, 2000.
- [49] Government Accountability Office. Slot-Controlled Airports: FAA’s Rules Could be Improved to Enhance Competition and Use of Available Capacity. Technical Report GAO-12-902, 2012.
- [50] S. Gupta. Transient Analysis of D(t)/M(t)/1 Queuing System with Applications to Computing Airport Delays. Master’s thesis, Massachusetts Institute of Technology, 2010.
- [51] W. Hall. *Efficient Capacity Allocation in a Collaborative Air Transportation System*. PhD thesis, Massachusetts Institute of Technology, 1999.
- [52] C. Hane, C. Barnhart, E. Johnson, R. Marsten, G. Nemhauser, and G. Sigismondi. The Fleet Assignment Problem: Solving a Large-scale Integer Program. *Mathematical Programming*, 70:211–232, 1995.
- [53] M. Hansen. Micro-level Analysis of Airport Delay Externalities using Deterministic Queuing Models: a Case Study. *Journal of Air Transport Management*, 8(2):73–87, 2002.

- [54] M. Hansen, A. Odoni, D. Lovell, T. Nikoleris, and K. Vlachou. Use of Queuing Models to Estimate Delays Savings from 4D Trajectory Precision. In *8th USA/Europe Air Traffic Management R&D Seminar*, 2009.
- [55] P. Harsha. *Mitigating Airport Congestion: Market Mechanisms and Airline Response Models*. PhD thesis, Massachusetts Institute of Technology, 2009.
- [56] C. Hendrickson and S. Matthews. *Civil Infrastructure Planning, Investment and Pricing*. Available at <http://cspbook.ce.cmu.edu>, 2011.
- [57] G. Hengsbach and A. Odoni. Time Dependent Estimates of Delays and Delay Costs at Major Airports. Technical Report 75-4, MIT Flight Transportation Laboratory, 1975.
- [58] International Air Transport Association. Worldwide Slot Guidelines. 5th Edition, 2013.
- [59] A. Jacquillat. A Queuing Model of Airport Congestion and Policy Implications at JFK and EWR. Master’s thesis, Massachusetts Institute of Technology, 2012.
- [60] A. Jacquillat and A. Odoni. Congestion Mitigation at JFK: The Potential of Schedule Coordination. *Transportation Research Record: Journal of the Transportation Research Board*, 2400(28-36), 2014.
- [61] A. Jacquillat and A. Odoni. An Integrated Scheduling and Operations Approach to Airport Congestion Mitigation. *Operations Research*, under minor revision, 2015.
- [62] A. Jacquillat and A. Odoni. Endogenous Control of Arrival and Departure Service Rates in Dynamic and Stochastic Queuing Models with Application at JFK and EWR. *Transportation Research Part E: Logistics and Transportation Review*, 73(1):133–151, 2015.
- [63] A. Jacquillat, A. Odoni, and M. Webster. Dynamic Control of Runway Configurations and of Arrival and Departure Service Rates at JFK Airport under Stochastic Queue Conditions. *Transportation Science*, under minor revision, 2015.
- [64] A. Jacquillat and V. Vaze. Inter-airline Equity and Airline Collaboration in Airport Scheduling Interventions. *in preparation*, 2015.
- [65] P. Joskow and R. Schmalensee. *Markets for Power: An Analysis of Electric Utility Deregulation*. MIT Press, 1983.
- [66] H. Khadilkar and H. Balakrishnan. Network Congestion Control of Airport Surface Operations. *AIAA Journal of Guidance, Control and Dynamics*, 37(3):933–940, 2014.
- [67] P. Kivestu. Alternative Methods of Investigating the Time Dependent M/G/k Queue. Master’s thesis, Massachusetts Institute of Technology, 1974.
- [68] R. Klein, H. Luss, and D. Smoth. A Lexicographic Minimax Algorithm for Multiperiod Resource Allocation. *Mathematical Programming*, 55(2):213–234, 1992.

- [69] F. Koppelman, G. Coldren, and R. Parker. Schedule Delay Impacts on Air-travel Itinerary Demand. *Transportation Research Part B: Methodological*, 42:263–273, 2008.
- [70] I. Kwan and M. Hansen. US Flight Delay in the 2000’s: an Econometric Analysis. In *90th Transportation Research Board Annual Meeting*, 2010.
- [71] J. Laffont and J. Tirole. *Competition in Telecommunications*. MIT Press, 2001.
- [72] L. Le, G. Donohue, K. Hoffman, and C. Chen. Optimum Airport Capacity Utilization under Congestion Management: The Case of New York LaGuardia Airport. *Transportation Planning and Technology*, 31(1):93–112, 2008.
- [73] L. Li and J. P. Clarke. A Stochastic Model of Runway Configuration Planning. In *AIAA Guidance, Navigation, and Control Conference*, 2010.
- [74] P. B. Liu, M. Hansen, and A. Mukherjee. Scenario-based Air Traffic Flow Management: From Theory to Practice. *Transportation Research Part B: Methodological*, 42:685–702, 2008.
- [75] M. Lohatepanont and C. Barnhart. Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment. *Transportation Science*, 38(1):19–32, 2004.
- [76] D. Lovell, A. Churchill, A. Odoni, A. Mukherjee, and M. Ball. Calibrating Aggregate Models of Flight Delays and Cancellation Probabilities at Individual Airports. In *7th USA/Europe Air Traffic Management R&D Seminar*, 2007.
- [77] C. Mayer and T. Sinai. Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil. *American Economic Review*, 93(4):1194–1215, September 2003.
- [78] R. Miller and D. Lessard. *The Strategic Management of Large Engineering Projects: Shaping Institutions, Risks, and Governance*. MIT Press, 2001.
- [79] A. Molnar. Congesting the Commons: A Test for Strategic Congestion Externalities in the Airline Industry. Working Paper, 2013.
- [80] T. Morisset and A. Odoni. Capacity Delays and Schedule Reliability at Major Airports in Europe and the United States. *Transportation Research Record*, 2214:85–93, 2011.
- [81] S. Morrison and C. Winston. Another Look at Airport Congestion Pricing. *American Economic Review*, 97(5):1970–1977, 2007.
- [82] S. Morrison and C. Winston. The Effect of FAA Expenditures on Air Travel Delays. *Journal of Urban Economics*, 63(2):669–678, 2008.
- [83] NERA. Study to assess the effects of different slot allocation schemes. Technical report, National Economic Research Associates, 2004.

- [84] T. Nikoleris and M. Hansen. Queuing Models for Trajectory-Based Aircraft Operations. *Transportation Science*, 46(4):501–511, 2012.
- [85] A. Odoni. *The Global Airline Industry*, chapter The International Institutional and Regulatory Environment, pages 19–46. John Wiley & Sons Ltd., 2009.
- [86] A. Odoni, T. Morisset, W. Drotleff, and A. Zock. Benchmarking Airport Airside Performance: FRA vs. EWR. In *9th USA/Europe Air Traffic Management R&D Seminar*, 2011.
- [87] Office of Inspector General. New York Flight Delays Have Three Main Causes, but More Work Is Needed To Understand Their Nationwide Effect. Technical Report AV-2011-007, U.S. Department of Transportation, 2010.
- [88] W. Ogryczak, M. Piore, and A. Tomaszewski. Telecommunications network design and max-min optimization problem. *Journal of Telecommunications and Information Technology*, 3:43–56, 2005.
- [89] P. Pellegrini, L. Castelli, and R. Pesenti. Secondary Trading of Airport Slots as a Combinatorial Exchange. *Transportation Research Part E: Logistics and Transportation Review*, 48(5):1009–1022, 2012.
- [90] E. Pels and E. Verhoef. The economics of airport congestion pricing. *Journal of Urban Economics*, 55:257–277, 2004.
- [91] M. Pena-Alcaraz. *Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems*. PhD thesis, Massachusetts Institute of Technology, 2015.
- [92] I. Pérez-Arriaga. *Regulation of the Power Sector*. Springer-Verlag, 2013.
- [93] N. Pujet, B. Delcaire, and E. Feron. Input-Output Modeling and Control of the Departure Process of Congested Airports. In *AIAA Guidance, Navigation, and Control Conference*, 1999.
- [94] N. Pyrgiotis. *A Stochastic and Dynamic Model of Delay Propagation Within an Airport Network For Policy Analysis*. PhD thesis, Massachusetts Institute of Technology, 2011.
- [95] N. Pyrgiotis, K. Malone, and A. Odoni. Modelling Delay Propagation within an Airport Network. *Transportation Research Part C: Emerging Technologies*, 27:60–75, 2013.
- [96] N. Pyrgiotis and A. Odoni. On the Impact of Scheduling Limits: A Case Study at Newark International Airport. *Transportation Science*, Articles in advance, 2015.
- [97] N. Pyrgiotis and I. Simaiakis. An Analytical Queuing Model of Airport Departure Processes for Taxi Out Time Prediction. In *AIAA Aviation Technology, Integration, and Operations*, 2010.

- [98] V. Ramanujam and H. Balakrishnan. Data-driven Modeling of the Airport Configuration Selection Process. *IEEE Transactions on Human-Machine Systems*, forthcoming, 2015.
- [99] S. Rassenti, V. Smith, and R. Bulfin. A Combinatorial Auction Mechanism for Airport Time Slot Allocation. *Bell Journal of Economics*, 13(2):402–417, 1982.
- [100] N Rupp. Do Carriers Internalize Congestion Costs? Empirical Evidence on the Internalization Question. *Journal of Urban Economics*, 65(1):24–37, 2009.
- [101] M. Ryerson and A. Woodburn. Build Airport Capacity or Manage Flight Demand? How Regional Planners Can Lead American Aviation Into a New Frontier of Demand Management. *Journal of the American Planning Association*, 80(2):138–152, 2014.
- [102] SESAR. European ATM Master Plan: The Roadmap for Sustainable Air Traffic Management. Technical report, 2012.
- [103] R. Shumsky. *Dynamic Statistical Models for the Prediction of Aircraft Take-off Times*. PhD thesis, Massachusetts Institute of Technology, 1995.
- [104] I. Simaiakis. *Analysis, Modeling and Control of the Airport Departure Process*. PhD thesis, Massachusetts Institute of Technology, 2012.
- [105] I. Simaiakis and H. Balakrishnan. Probabilistic Modeling of Runway Inter-departure Times. *AIAA Journal of Guidance, Control and Dynamics*, 37(6), 2014.
- [106] I. Simaiakis and H. Balakrishnan. A Queuing Model of the Airport Departure Process. *Transportation Science*, Forthcoming, 2015.
- [107] I. Simaiakis, M. Sandberg, and H. Balakrishnan. Dynamic Control of Airport Departures: Algorithm Development and Field Evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):285–295, 2014.
- [108] I. Simaiakis, M. Sandberg, H. Balakrishnan, T. Reynolds, and J. Hansman. Demonstration of Reduced Airport Congestion through Pushback Rate Control. *Transportation Research Part A: Policy and Practice*, 66:251–267, 2014.
- [109] G. Skaltsas. Analysis of Airline Schedule Padding on U.S. Domestic Routes. Master’s thesis, Massachusetts Institute of Technology, 2011.
- [110] G. Solveling, S. Solak, J. P. Clarke, and E. Johnson. Runway Operations Optimization in the Presence of Uncertainties. *AIAA Journal of Guidance, Control and Dynamics*, 34(5):1373–1382, 2011.
- [111] N. Sood and F. Wieland. Total Airport and Airspace Model (TAAM) Parallelization Combining Sequential and Parallel Algorithms for Performance Enhancement. In *2003 Winter Simulation Conference*, 2003.

- [112] A. Sun. *Advances in Electric Power Systems: Robustness, Adaptability, and Fairness*. PhD thesis, Massachusetts Institute of Technology, 2011.
- [113] P. Swaroop. *Problems and Models in Strategic Air Traffic Flow Management*. PhD thesis, University of Maryland, 2013.
- [114] P. Swaroop, B. Zou, M. Ball, and M. Hansen. Do More US Airports Need Slot Controls? A Welfare Based Approach to Determine Slot Levels. *Transportation Research Part B: Methodological*, 46(9):1239–1259, 2012.
- [115] V. Vaze and C. Barnhart. Modeling Airline Frequency Competition for Airport Congestion Mitigation. *Transportation Science*, 46(4):512–535, 2012.
- [116] V. Vaze and C. Barnhart. The Role of Airline Frequency Competition in Airport Congestion Pricing. *Transportation Research Record: Journal of the Transportation Research Board*, 2266(69-77), 2012.
- [117] V. Vaze and C. Barnhart. An Assessment of the Impact of Demand Management Strategies for Efficient Allocation of Airport Capacity. *2012*, 6(1/2):5–27, International Journal of Revenue Management.
- [118] T. Vossen and M. Ball. Slot Trading Opportunities in Collaborative Ground Delay Programs. *Transportation Science*, 40(1):29–43, 2006.
- [119] T. Vossen, R. Hoffman, and A. Mukherjee. Air Traffic Flow Management. In *Quantitative Problem Solving Methods in the Airline Industry*, volume 169 of *International Series in Operations Research & Management Science*, pages 385–453. Springer US, 2012.
- [120] C. Weld, M. Duarte, and R. Kincaid. A Runway Configuration Management Model with Marginally Decreasing Transition Capacities. *Advances in Operations Research*, 2010.
- [121] M. Wittman and W. Swelbar. Capacity Discipline and the Consolidation of Airport Connectivity in the United States. *Transportation Research Record: Journal of the Transportation Research Board*, 2449:72–78, 2014.
- [122] N. Xu. *Method for Deriving Multi-factor Models for Predicting Airport Delays*. PhD thesis, George Mason University, 2007.
- [123] K. Zografos, Y. Salouras, and M. Madas. Dealing with the Efficient Allocation of Scarce Resources at Congested Airports. *Transportation Research Part C: Emerging Technologies*, 21:244–256, 2012.