A Stochastic and Dynamic Model of Delay Propagation Within an Airport Network For Policy Analysis

by
Nikolaos Pyrgiotis

M.Eng., Imperial College London (2005)
S.M., Massachusetts Institute of Technology (2008)

Submitted to the Department of Aeronautics and Astronautics in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

© Massachusetts Institute of Technology 2012. All rights reserved.

Author

Department of Aeronautics and Astronautics

September 15, 2011

Certified by
Amedeo R. Odoni
Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by
Cynthia Barnhart
Professor of Civil and Environmental Engineering

Doctoral Committee Member

Certified by
Hamsa Balakrishnan
Professor of Aeronautics and Astronautics

Doctoral Committee Member

Accepted by
Eytan Modiano
Professor of Aeronautics and Astronautics
Chairman, Department Committee on Graduate Theses
A Stochastic and Dynamic Model of Delay Propagation
Within an Airport Network For Policy Analysis

by

Nikolaos Pyrgiotis

Submitted to the Department of Aeronautics and Astronautics
on September 15, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

As demand for air travel increases over the years and many busy airports operate close
to their capacity limits, congestion at some airports on any given day can quickly
spread throughout the National Aviation System (NAS). It is therefore increasingly
important to study the operation of large networks of airports as a group and to
understand better the interactions among them under a wide range of conditions.
This thesis develops a fundamental tool for this purpose, enhances it with several
capabilities designed to address issues of particular interest, and presents some early
insights and observations on the system-wide impacts of various scenarios of network-
wide scope.

We first describe an analytical queuing and network decomposition model for the
study of delays and delay propagation in a large network of airports. The Airport
Network Delays (AND) model aims to bridge the gap in the existing modeling tools
between micro-simulations that track aircraft itineraries, but require extensive re-
sources and computational effort, and macroscopic models that are simple to use, but
typically lack aircraft itinerary tracking capabilities and credible queuing models of
airport congestion. AND operates by iterating between its two main components: a
queuing engine (QE), which is a stochastic and dynamic queuing model that treats
each airport in the network as a $M(t)/E_k(t)/1$ queuing system and is used to compute
delays at individual airports and a delay propagation algorithm (DPA) that updates
flight schedules and demand rates at all the airports in the model in response to the
local delays computed by the QE.

We apply AND to two networks, one consisting of the 34 busiest airports in the
United States and the other of the 19 busiest in Europe. As part of the development
of AND, we perform a statistical analysis of the minimum ground turn-around times
of aircraft, one of the fundamental variables that determine delay propagation. In
addition, we show that the QE, with proper calibration, can model very accurately
the airport departure process, predicting delays at two major US airports within
10% of observed values. We also validate the AND model on a network-wide scale
against field data reported by the FAA. Finally, we present insights into the complex
interactions through which delays propagate through a network of airports and the often-counterintuitive consequences.

In the third part of the thesis, we present two important extensions of the AND model designed to expand its usability and applicability. First, in order to provide a more accurate representation of NAS operations, we develop an algorithm that replicates quite accurately the execution of Ground Delay Programs (GDPs). The algorithm operates consistently with the rules of the Collaborative Decision-Making (CDM) process under which GDPs are currently conducted in the United States. The second extension is the implementation in AND of a deterministic queuing engine (\(D(t)/D(t)/1\)) which can be used as an alternative to the original stochastic QE. This deterministic model can be used to study delay-related performance in a future system that operates at a higher level of predictability than the current one, as the one envisioned by FAA in the Next Generation Air Transportation System.

In the final part of the thesis we describe a Mixed Integer optimization model for studying the impact of introducing slot controls at busy airports. The model generates new flight schedules at airports by reducing the number of available slots, while respecting all existing aircraft itineraries and preserving all passenger connections. We test the model at Newark Airport (EWR) and conclude that, with a small schedule displacement (less than 30 minutes for any flight during the day), it is possible to obtain a feasible schedule that obeys slot limits that are as low as the IFR capacity of the airport. We test the new schedule in AND and find that the local delay savings that would result from “slot-controlling” EWR in this way are of the order of 10% for arrivals and of 50% for departures, while we may also expect a reduction of 23% in propagated delays to the rest of the US network of airports.

Thesis Supervisor: Amedeo R. Odoni
Title: Professor of Aeronautics and Astronautics
Acknowledgments

My first and foremost thank you goes to my advisor Amedeo Odoni. Over the last four years his academic (and non-academic) guidance and advice, and constant encouragement and support have been amongst the most important reasons for the successful completion of this thesis. I want to express my sincere gratitude for his countless remarks, comments, input and editing during the write-up of this dissertation. His deep knowledge and expertise in the area of air transportation and operations research have kept me motivated at all times and helped me define research ideas. I’m also grateful for Amedeo’s continuous interest and all the help he provided towards my future career plans.

I would like to thank my other two committee members, Cynthia Barnhart and Hamsa Balakrishnan for their guidance, feedback and continuous interest on my research work. I feel immense gratitude towards the MIT-Portugal program for funding my research and providing me with the opportunity to collaborate with professors and students from Portugal. In particular, I want to express my sincere gratitude to Antonio Antunes for our great collaboration in the AirNets project, for his hospitality in all of our visits to Portugal and his sense of humor.

I’m thankful to all the faculty, students and staff in Aero-Astro and the ICAT lab, and everyone related to air transportation research that I have had the chance to interact with. In particular, I want to thank Joao Pita, Miguel Santos and Vikrant Vaze for our collaboration through the AirNets project and for the good times we had in Portugal. I would also like to thank Claire Cizaire, Diana Michalek and everyone else in the ICAT lab for all the talks (for research or otherwise) and the fun we had, creating a lively and friendly environment in our office.

All of my friends at MIT have made the past five years in Cambridge an unforgettable experience. I want to thank in particular my beloved friends and roommates Kostas Bimpikis and Ruben Lobel for all the fun we had in the “church” and my friend and lab-mate Yannis Simaiakis for the infinite discussions in (and out) of the lab and for our research collaboration. In addition, I want to thank my friends Gerry
Tsoukalas, Hoda Eudgahi, Gerasimos Skaltsas and Ioannis Bertatsos that filled my life in grad school with so much fun and my oldest friend Spyros Kallipolitis for the great weekends in New York and for his cheerfulness and support. Moreover, I want to express my appreciation to everyone in our basketball team for all the excitements we went through during the last four IM seasons.

Finally, I would like to express my deepest appreciation to my parents, Fotini and Yannis, who have always been standing next to me. This thesis is dedicated to them.
3 Calibration, Validation and Testing of the AND model

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Development of AND</td>
<td>72</td>
</tr>
<tr>
<td>3.1.1 Separating arrival and departure queues</td>
<td>75</td>
</tr>
<tr>
<td>3.1.2 Estimation of Aircraft Turnaround Times</td>
<td>77</td>
</tr>
<tr>
<td>3.1.3 Padding of Scheduled Block Times</td>
<td>87</td>
</tr>
<tr>
<td>3.2 Validation of the $M(t)/E_k(t)/1$ queuing system at the airport departure process</td>
<td>90</td>
</tr>
<tr>
<td>3.2.1 Model structure</td>
<td>90</td>
</tr>
<tr>
<td>3.2.2 Unimpeded taxi-out times and Ramp and taxiway interactions</td>
<td>92</td>
</tr>
<tr>
<td>3.2.3 The runway service process</td>
<td>93</td>
</tr>
<tr>
<td>3.2.4 Predictions at EWR and BOS</td>
<td>98</td>
</tr>
<tr>
<td>3.3 Validation of AND</td>
<td>104</td>
</tr>
<tr>
<td>3.4 A Detailed Example and Insights from the AND Model</td>
<td>111</td>
</tr>
<tr>
<td>3.4.1 Overall Network Behavior</td>
<td>112</td>
</tr>
<tr>
<td>3.4.2 Local versus upstream delays</td>
<td>114</td>
</tr>
<tr>
<td>3.4.3 The effect of delay propagation</td>
<td>115</td>
</tr>
<tr>
<td>3.5 AND European Air Transportation Network</td>
<td>121</td>
</tr>
<tr>
<td>3.5.1 A Comparison Between the US and EU Airport Hubs</td>
<td>123</td>
</tr>
</tbody>
</table>

4 Extensions to the Airport Network Delays Model

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Modeling Ground Delay Programs</td>
<td>127</td>
</tr>
<tr>
<td>4.1.1 Ground Delay Program Algorithm</td>
<td>131</td>
</tr>
<tr>
<td>4.1.2 GDP initiation</td>
<td>132</td>
</tr>
<tr>
<td>4.1.3 Slot assignment</td>
<td>133</td>
</tr>
<tr>
<td>4.1.4 Continuation and Termination of the GDP</td>
<td>136</td>
</tr>
<tr>
<td>4.1.5 Link with AND</td>
<td>137</td>
</tr>
<tr>
<td>4.1.6 Validation and Results</td>
<td>138</td>
</tr>
<tr>
<td>4.2 Alternative Queuing Engine:</td>
<td></td>
</tr>
<tr>
<td>Deterministic Queues</td>
<td>143</td>
</tr>
</tbody>
</table>
4.2.1 Deterministic Queuing Engine ........................................ 144
4.2.2 Comparison of the Stochastic and the Deterministic Airport
   Network Model ............................................................ 147
4.2.3 Local Comparisons ..................................................... 147
4.2.4 Network-wide comparison ............................................. 152
4.3 Summary ................................................................. 154

5 On the Effect of Slot Control at Busy US Airports ..................... 159
5.1 The Slot Control System ................................................ 160
5.2 Analysis of demand and capacity at Newark Liberty Int’l Airport ... 161
   5.2.1 Newark Liberty Int’l Airport .................................... 161
   5.2.2 Capacity and Schedule at EWR ............................. 164
5.3 The Demand Smoothing Model ........................................... 172
   5.3.1 Input ................................................................. 173
   5.3.2 Parameters .......................................................... 174
   5.3.3 Variables ............................................................ 176
   5.3.4 Objective ............................................................ 176
   5.3.5 Constraints .......................................................... 177
   5.3.6 Model Development and Computation of Model Solution .... 180
   5.3.7 Experimental Setup .............................................. 183
5.4 Optimization results ..................................................... 186
5.5 Testing in AND ............................................................ 189
   5.5.1 Local Delays at EWR ........................................... 191
   5.5.2 Network delays .................................................... 197
   5.5.3 Modified runway utilization ..................................... 198
5.6 Summary ................................................................. 202

6 Conclusions ........................................................................ 205
6.1 Summary of Thesis ....................................................... 205
6.2 Future Research ........................................................... 210
A  Estimated Minimum Turnaround Times  215

B  Estimated Nominal Block Time  219
List of Figures

1-1 Evolution of arrival delays in the years 2000-2010 [Source: ASPM [19]]. 21
1-2 Break-down of delays by cause in 2007. 21
1-3 On-time performance of flights classified according to the number of flights on aircraft itineraries.[50] 23

2-1 A schematic of AND iteration. 44
2-2 Conceptual representation of a single airport as a queuing system. 46
2-3 A state transition diagram for an $M(t)/E_3(t)/1$ with arrival rate $\lambda(t)$ and service rate $\mu(t)$, with capacity of 4 customers in the system. 47
2-4 Flight breakdown. 53
2-5 Maximum number of flights from a single origin per hour of day at Chicago O’Hare International Airport. 64
2-6 The family of Erlang distributions. 65

3-1 Airport map of the US AND model. 73
3-2 Use of the virtual airport in AND. 73
3-3 Box-plots of the turn time of the B737 family of aircraft (B737-300, B737-500 and B737-700) for CO and WN at their hubs and at all other airports. 81
3-4 Box-plots of the turn time of the A320 aircraft for UA and B6 at their hubs and at all other airports. 82
3-5 Histograms of the actual and the scheduled turn times of the A320 aircraft for UA and B6 outside their hubs. 83
3-6 Box-plots of the actual turn times of 6 aircraft types in DL’s fleet. 83
3-7 Sensitivity analysis results. ........................................ 87
3-8 Integrated model of the departure process. [47, 49] ............ 91
3-9 Takeoff rate as function of $N(t)$ [47] ............................ 95
3-10 Takeoff rate as function of $N(t)$ .................................. 96
3-11 Histogram of the inter-departure times in saturation and Gamma and
Erlang distribution fits .................................................. 97
3-12 Actual and modeled takeoff rate as a function of $N(t)$ ........ 100
3-13 Actual and modeled $N(t)$ histogram ............................ 101
3-14 Prediction of departure throughput, average taxi-out times and depart-
ture queue lengths in each 15-min interval over a 10-hour period on
July 22, 2007. ............................................................... 103
3-15 Average arrival delay per hour at ORD on 04/01/2007 .......... 107
3-16 Average arrival delay by airport on 04/01/2007 ................. 108
3-17 Average arrival delay of the final arrival of an aircraft at ORD as a
function of the number of total visits to ORD on 04/01/2007. ... 108
3-18 Average arrival delay per hour at the three New York airports on
05/10/2007. ............................................................... 109
3-19 Average arrival delay by airport on 05/10/2007. ............... 110
3-20 Percent of average flight delay caused by upstream delay at each of the
34 airports. ............................................................... 116
3-21 Demand and delay profiles at ORD in Scenario 2. .......... 117
3-22 Demand and delay profiles at LGA in Scenario 3. ............ 119
3-23 Flight arrival delays at ORD in Scenario 2 ....................... 121
3-24 European AND airport map. ..................................... 122
3-25 Flight share by airline at FRA and ORD. ....................... 123
3-26 Analysis at FRA and ORD. ..................................... 124

4-1 Average ground hold time as a function of flight time at ORD. ... 130
4-2 Heap representation of the arrival queue before and after the slot assignment in the period 15:00-15:15. (a) shows 10 non-exempt flights scheduled to arrive between 15:00 and 15:20; (b) shows the 5 exempt flights and their respective departure times in parenthesis.

4-3 Scheduled arrivals and Arrival Acceptance Rate per 15 minutes at Chicago O’Hare on 04/01/2007.

4-4 Average arrival delay per hour at ORD.

4-5 Average arrival delay by airport.

4-6 Average arrival delay at the final arrival of an aircraft at ORD as a function of the number of previous visits at ORD.

4-7 Demand and Capacity per period at a random airport.

4-8 Cumulative demand and throughput and the corresponding queue length and delay for the example shown in Figure 4-7.

4-9 Average local arrival delay by airport for each of the four scenarios, including linear regression lines.

4-10 Average local arrival delay by 15-minute period for LGA under each of the three scenarios.

4-11 Ratio of the deterministic over the stochastic AND local arrival delays at LGA.

4-12 Comparison of the deterministic to the stochastic AND local arrival delay estimates at LGA with a single queue shared by arrivals and departures.

4-13 Average total (local plus upstream) arrival delay by airport for each of the four scenarios, including a linear regression line. (Note the different scales at each graph.)

4-14 Total upstream delay by airport for each of the four scenarios, including a linear regression line.

5-1 Runway layout at EWR.

5-2 Utilization of the main runway configurations by hour of day at EWR.
5-3 Hourly scheduled operations at EWR in 2007; the bold line shows the hourly slot limit introduced in 2008.[note:TR, S and SN stand for Thursday, Saturday and Sunday respectively] ................................. 163
5-4 Scatter plots of observed arrival and departure counts per 15 minutes at EWR in 2007; red line indicates the expected capacity [48]. ............. 167
5-5 Scatter plots of scheduled arrival and departure counts per 15 minutes at EWR in 2007; lines indicate the expected capacity [48]. ............. 169
5-6 Cumulative frequency of scheduled operations counts per hour at EWR in 2007; lines indicate the expected capacity and the current slot limit. 169
5-7 Average departure and arrival schedule daily profile in 2007 at EWR.[note: TR stands for Thursday] ................................. 171
5-8 Distribution of total daily operations at EWR in 2007. ................. 185
5-9 Scheduled operations at EWR on 07/13/2007. ......................... 185
5-10 Original and modified schedule at EWR for Test 3, 4 and 5. ........ 190
5-11 Delay profiles at EWR in VFR conditions under the original schedule and the new schedule from Tests 3, 4 and 5. ......................... 193
5-12 Delay profiles at EWR in IFR conditions under the original schedule and the new schedules from Tests 3, 4 and 5. ......................... 195
5-13 Scheduled and adjusted demand profiles at EWR in IFR conditions under the new schedule from Test 5. ................................. 196
5-14 Flight arrival delay profile at EWR in VFR and IFR conditions for the four different schedules. ................................. 197
5-15 Delay profiles at EWR in VFR conditions under the original schedule and the new schedule from Test 5 and with a modified runway utilization during the day. ................................. 201
List of Tables

1.1 Models of air transportation systems. ........................................ 27

3.1 Turn time statistics. ................................................................. 85

3.2 Sensitivity analysis results. ...................................................... 88

3.3 Runway service time characteristics for two frequently-used runway configurations at BOS and at EWR .......................................................... 97

3.4 Time-dependent Erlang distribution for the runway service time for two frequently-used runway configurations at BOS and at EWR ........... 99

3.5 Aggregate taxi time predictions for BOS segment (VMC; 4L, 4R | 4L, 4R, 9) .................................................................................. 102

3.6 Aggregate taxi time predictions for EWR segment (VMC; 22L | 22R) 102

3.7 Individual taxi time predictions for BOS segment (VMC; 4L, 4R | 4L, 4R, 9) and EWR segment (VMC; 22L | 22R) .............................. 103

3.8 Evaluation of model predictions for July 22, 2007 at BOS. ........ 104

3.9 Average arrival delay (in minutes) at EWR, JFK and LGA on 05/10/2007. 111

3.10 The three scenarios tested in AND. ............................................ 112

3.11 Statistics for the 24-hour period for each of the airports in AND under each of the three scenarios. .................................................. 113

4.1 Average ground hold with respect to total GDP length, of flights with more than 4.5 hours flight time destined to ORD. .......................... 131
4.2 Schedule and delays of American Airlines’ N289AA aircraft flying through ORD (All times in GMT). “ZZZ” signifies an out-of-the-network airport. .......................... 142

4.3 Four scenarios for the comparison of the Stochastic and the Deterministic AND. .................................................. 147

4.4 Gradient of the regression lines from Figure 4-9. ......................... 148

4.5 Gradient of the regression lines from Figures 4-13 and 4-14. ........ 154

5.1 Runway configuration frequency of utilization at EWR under VFR and IFR conditions. .................................................. 162

5.2 Test scenarios for demand smoothing at EWR. .............................. 186

5.3 Optimization results for all flights in $F$ in EWR tests (1 period = 15 minutes). ........................................... 187

5.4 Optimization results for flights in $F$ that operate directly in/out of EWR (1 period = 15 minutes). ....................... 187

5.5 Percentage of flights operated by each airline at EWR tests. ....... 188

5.6 Optimization results: Number of flights displaced by airline in EWR tests. ........................................... 189

5.7 Optimization Results: Average displacement by airline in EWR tests. 189

5.8 Local delays at EWR in VFR (in parenthesis is shown the percentage change from the original schedule) .................. 192

5.9 Local delays at EWR in IFR ................................................. 194

5.10 Average flight arrival delay at EWR (shown in parenthesis is the percentage reduction from the original schedule) .......... 197

5.11 Total daily upstream delays (in minutes) at selected airports of the AND network for Scenarios 1 and 2 (shown in parenthesis is the percentage reduction from the original schedule) ..................... 199

5.12 Local delays at EWR in VFR (shown in parenthesis is the percent change from the original schedule) ......................... 200
5.13 Average flight arrival delay at EWR (in minutes), (shown in parenthesis is the percent reduction from the original schedule)........ 200

A.1 Average minimum turn times by airline, by aircraft type and by airport type .......................................................... 216

A.2 Average minimum turn times by airline, by aircraft type and by airport type .......................................................... 217

B.1 Nominal block time components by route (in minutes), as estimated by Skaltsas [50] ....................................................... 220

B.2 Nominal block time components by route (in minutes), as estimated by Skaltsas [50] ....................................................... 221

B.3 Nominal block time components by route (in minutes), as estimated by Skaltsas [50] ....................................................... 222
Chapter 1

Introduction

As demand for air travel has increased in the United States over the years, so have air traffic congestion and delays. Particularly noteworthy has been the growing phenomenon of delay propagation within the National Airspace System (NAS), with local congestion at some locations spreading quickly to generate delays in large sections of the NAS. The current capacity of the NAS is often being utilized to its limits and new infrastructure development and demand management schemes will soon be necessary, should demand resume an upward trend.

Tools that are able to assess the system-wide impacts of proposed alternatives for increasing capacity and of other measures that aim at relieving airport congestion are badly needed. In this thesis we develop an analytical, dynamic and stochastic model of large networks of airports that aims to facilitate the study of system-wide queuing phenomena, as they develop and propagate. As will be seen, this model makes it possible to perform macroscopic-level analyses at the network level quickly and to explore efficiently a broad range of alternatives.

In this Chapter, we first summarize some statistics about delays and delay propagation within the NAS and then introduce the most important available delay mitigation mechanisms. In Section 1.2, we present a literature review, focusing on existing modeling tools of the NAS as a whole, as well as on research concerned with the application of demand management policies at airports. Section 1.3 contains a brief review of the contributions of this thesis, in terms of (i) the development of a powerful
new tool, (ii) the enhancement of this tool with a number of important capabilities and (iii) the insights obtained through the exploration of a number of different scenarios using the new model. The final section of the chapter contains an outline of the structure and content of Chapters 2 through 6.

1.1 The Air Transportation System

1.1.1 Air traffic delays and delay propagation

Until the early 1990s, major air traffic delays were largely confined to a relatively small number of airports. However, the strong growth in the number of airport operations that took place during that decade and, again, between 2003 and 2007, has led to system-wide congestion problems. According to the Bureau of Transportation Statistics (BTS) [53] every year since 2004 at least 18% of all commercial flights in the United States arrived delayed by more than 15 minutes, with the figure peaking in 2007 at 25%. Figure 1-1 shows the evolution of arrival delays from 2000 to 2010 [19]. The average delay at the 35 busiest US airports increased from 10 minutes in 2002 to 16 minutes in 2007 (a 60% increase). To put this into perspective, an increase of 6 minutes of the system-wide average delay translates to roughly 1.2 million aircraft-hours of delays in a year. Assuming an aircraft utilization of 3500 hours per year, this increase is equivalent to 350 aircraft-years, or the entire fleet of United Airlines being grounded for a year. Four airports, including the three major airports of the New York Metroplex, have averaged more than 20 minutes of arrival delay throughout the three years from 2006 to 2008.

By 2007, the worst year for delays in aviation history, demand levels at many airports in the United States and, to a lesser extent, in Europe, were close to or exceeded the capacities of these airports for several hours each day, especially during the peak summer season. Observing Figure 1-2 we see that 12% of airport delays have been attributed directly to the imbalance between demand and capacity at airports. However, a large part of the 77% of weather-related delays can also be indirectly
Figure 1-1: Evolution of arrival delays in the years 2000-2010 [Source: ASPM [19]].

Figure 1-2: Break-down of delays by cause in 2007.
attributed to insufficient capacity at airports when weather deteriorates. As a result, not only did delays at individual airports reach record levels, particularly on days when less-than-ideal weather conditions prevailed, but congestion also spread readily on such days, propagating throughout airline fleets and affecting large parts of the airport system.

The net cost of congestion in this tightly inter-connected and over-scheduled network of airports is enormous. Estimates for the United States for 2007 range from $14 billion (Air Transport Association, 2009 [1]) to $41 billion (U.S. Congress Joint Economic Committee, 2008 [52])—with the latter estimate purporting to include both the direct costs of the delays to the airlines and their passengers and the indirect and induced costs that these delays cause to the airline industry (e.g., by forcing the industry to increase the scheduled gate-to-gate time of flights) and to other sectors of the economy.

Airline schedules include some “slack”, both in the planned gate-to-gate time-lengths of flights and in the “turnaround times” on the ground between consecutive flights of any given aircraft. For instance, while an aircraft may be scheduled to be on the ground between flights at some particular airport for 45 minutes, it may actually require only 35 minutes to turn around, thus providing a slack of 10 minutes. But these slacks are generally insufficient to absorb the longer delays that typically occur on a daily basis, thus leading to the propagation of delay. For example, if an aircraft is delayed by one hour when departing from airport A, it will almost certainly be late when arriving at its next airport B; the late arrival at B may also result in a late subsequent departure of that aircraft from B—leading to the dreaded announcement of a delay “due to a late-arriving aircraft”.

The effect of “delay propagation” has been increasingly important as more airports in the United States become congested. In fact, at least one airline has computed empirical “multipliers” (Beatty, 1998 [5]) for delays incurred during different parts of the day. For instance, that particular study claims that one hour of delay suffered by some flights early in the day may result in seven hours of delay for the airline’s entire fleet, as that initial delay propagates to other aircraft and airports later in the
day through late flight arrivals and late connections. Recently, Skaltsas (2011) [50] showed that, as aircraft execute their daily schedules, the on-time performance of the flights in their schedules deteriorates. For example, observing Figure 1-3, the average on-time performance of an aircraft that executes 3 flights daily, drops from 80% at the end of the first flight to 75% at the end of the third flight of the itinerary. In a similar study for the European air transportation system, Jetzki (2010) [27] claims that 40% of all departure delays in Europe can be attributed to delays due to a late arriving aircraft. Moreover, Jetzki showed that propagated delays starting in the morning have a higher impact on system-wide delays than the ones starting in the afternoon, since they propagate on average over more flight legs.

1.1.2 Delay mitigation

Strategic delay mitigation measures are important for the sustainable evolution of the air transportation system. We may classify actions to reduce delays into two categories, demand management policies and infrastructure improvements. Demand management policies aim at regulating the demand for access at busy airports. The two main demand management policies are slot control and congestion pricing. A slot
is defined as a specific time when an operation (arrival or departure) is allowed to be scheduled at an airport. In a slot control regime, the responsible authority allows only a specific number of arrivals and departures to be scheduled in any given period of time (typically of 5, 10, 15 or 60 minutes). Congestion pricing involves setting arrival and departure fees at levels designed to reduce demand to numbers comparable to the airport’s capacity. Demand management through congestion pricing relies on airlines’ willingness to pay for access at each specific airport.

Infrastructure improvements generally deal with increasing the capacity of the system, not managing the demand. These improvements mainly include airport expansion projects and implementation of new technologies that allow more efficient operations both in the terminal area and in the entire airspace. For example, the Federal Aviation Administration (FAA) through the Next Generation Air Transportation System (NextGen) [38] envisions an increase of the overall capacity of the NAS through the introduction of a combination of new technologies that will assist trajectory-based operations. With four-dimensional (4D) flight trajectories, the exact position of each aircraft will be accurately predicted and tracked in time and space, hence leading to reduced and more precise separations between aircraft, as well as better traffic coordination. The application of NextGen technologies requires the extensive upgrading of many Air Traffic Control (ATC) systems, as well as of aircraft avionics.

Infrastructure improvements typically require a long time to implement and are often associated with enormous capital costs. Hence, in order for a new technology to be adopted or a new runway to be approved the benefits have to be sufficiently large to outweigh the costs. Similarly, the benefits of demand management policies have to be very clear because such policies may restrict airlines from flying their preferred schedules. It is then clear that models of the air transportation system on which any of the aforementioned strategies for delay mitigation can be tested are highly desirable. Models of this type should be able to support studies of the impact of hypothetical developments such as (in increasing degree of difficulty):

- a new runway that boosts capacity at a hub airport;
a future air traffic management system (e.g., NextGen in the United States, SESAR in Europe) that increases capacity, to varying extents, at all airports;

- the imposition of “slot limits” at certain key airports, limiting the number of aircraft movements that can be scheduled per hour;

- the initiation of a congestion pricing scheme at a selected sub-set of airports.

1.1.3 Research goals

The primary objectives of this thesis are to:

1. Develop a model that includes representations of a large set of busy airports in the NAS and captures the interactions among them;

2. provide certain capabilities to the model that are motivated by important policy considerations; and

3. study network phenomena, related to how delays are created and how they propagate within an airport system.

We develop the Airport Network Delays model (AND), a stochastic and dynamic queuing model, designed to compute approximately delays at each of the individual airports in a network and, more important, how these delays propagate from one airport to another over the course of a day or other time period of interest. AND treats the airports in the network as a set of interconnected queuing systems. Delays at any specific airport may impact delays at other airports in the network, as aircraft execute their daily flight schedules (or “itineraries”) by flying from one airport to another.

AND employs a combination of a numerical queuing model (its “queuing engine”, QE) and a delay-propagation algorithm (DPA). The QE computes delays at individual airports and the DPA tracks the propagation of these delays and their impact on subsequent airline operations at all the other airports in the network. Because it is fast and simple to use, AND may be used to explore at a “macroscopic”, approximate
level a large number of scenarios, alternatives and policies. For example, we have been able to develop an approach for investigating the effect of introducing slot controls at some busy US airports on delays in the entire air transportation system.

1.2 Literature Review

1.2.1 Models of the Air Transportation System

From the technical point of view, problems involving network-wide airport congestion are difficult to analyze. Steady-state queuing models are inapplicable for all but the crudest approximations because airport demand typically varies strongly with time-of-day and the dynamic characteristics of the airport queues thus become dominant. Similarly, airport capacity often varies significantly over the course of the day in response to weather conditions. Deterministic flow models with bottlenecks fail to capture another essential aspect, namely the fundamentally stochastic nature of airport operations. Finally, continuous ("fluid") models with stochastic elements (e.g., heavy-traffic, diffusion approximations of queues at individual airports) are also of limited usefulness at the network level, especially if one is interested in tracking the delays experienced by individual aircraft executing their daily itineraries.

Generally, models of the Air Transportation System may be classified according to four fundamental criteria:

1. Simulations versus analytical models.

2. The level of detail in the representation of the system (microscopic, mesoscopic and macroscopic models).

3. The ability to capture randomness (deterministic versus stochastic models).

4. The ability to track the itineraries and the delays suffered by individual aircraft.

Simulations (deterministic or stochastic) require multiple runs in order to obtain statistically significant results that describe the operation of systems that are subject
to uncertainty or stochasticity ("randomness"). Given the same input variables, deterministic simulations provide identical results at every run of the model. To obtain meaningful results reflecting variability, the user must change some of the input parameters between runs of the simulation, in order to capture any randomness in the system. On the other hand, stochastic simulation models produce different results at every run of the simulation, even with the same input variables. Randomness is captured in the form of probability distribution functions (pdf) that describe each of the system's variables (such as airport capacities or flight times). During each simulation run all variables are assigned a value that is drawn from their corresponding pdf.

Models can also be distinguished by whether or not they track aircraft as they fly through the airport network. By tracking aircraft itineraries, such models may estimate the effects of delay propagation. Models that do not track aircraft itineraries—by not linking flights operated by the same aircraft—overlook a potentially enormous share of delays—i.e., the delays which can be attributed to earlier flights by the same aircraft. They therefore lack the ability to model interactions between airports in a network.

Existing models of the Air Transportation System cover most of the possible combinations of the aforementioned classes. In Table 1.1 some well-known existing models are classified according to their attributes.

Table 1.1: Models of air transportation systems.

<table>
<thead>
<tr>
<th></th>
<th>Simulations</th>
<th>Analytical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Micro</td>
<td>Macro</td>
</tr>
<tr>
<td></td>
<td>A/c tracking</td>
<td>A/c tracking</td>
</tr>
<tr>
<td>Stochastic</td>
<td>TAAM</td>
<td>ATM-NEEMO</td>
</tr>
<tr>
<td>Deterministic</td>
<td>ACES, FACET</td>
<td>NASPAC, DPAT</td>
</tr>
</tbody>
</table>

27
Simulation models

It is not surprising that most of the (few) available models of queuing in airport networks are simulations. The National Airspace System Performance Analysis Capability (NASPAC) of MITRE CAASD (Frolow and Sinnott, 1989 [20]) was one of the first NAS simulation models to be developed. This model, with many subsequent enhancements, is still being used by the FAA today. NASPAC is a macroscopic, deterministic simulation model that requires extensive effort on the input side to obtain meaningful statistical results. Wieland (1997) [56] describes an alternative model, the Detailed Policy Assessment Tool (DPAT), a low-level-of-detail simulation of the national airport network also developed at MITRE CAASD. Given a set of daily capacity and demand profiles at each airport in a network, DPAT estimates delays for any flight between any two airports. However, DPAT lacks input information on aircraft itineraries, i.e., the sequence of airports that each aircraft will visit on a given day, and thus does not fully capture the impact of delays at any one airport on delays in the rest of the system. In addition, the queuing model, which is used to model airport congestion, is very simple and approximate and will only produce delays when the demand exceeds the capacity at an airport.

Far more fine-grained simulation models of air traffic operations at the national level have also been developed. The state-of-the-art "microscopic" models in this category are ACES, the Airspace Concept Evaluation System (Raytheon, 2003 [43]) and FACET, the Future ATM Concepts Evaluation Tool (Bilimoria et al., 2000 [7]) both used currently by NASA and the FAA. These are agent-based simulation tools that model with high detail the entire NAS, but require a great amount of computation time, as well as extensive input preparation. For example, a run of ACES with a full-scale representation of the NAS typically requires several hours and it takes multiple runs to obtain statistically reliable estimates of the parameters of interest. Furthermore, as ACES is a deterministic simulation model, extensive input preparation would be required before each simulation run to capture any level of uncertainty. Finally, the Total Airspace and Airport Modeler (TAAM) [51] network model is a less
detailed network model than FACET and ACES but, still, considered a microscopic
model of the NAS. In fact, TAAM requires at least 30 minutes for the simulation of
only part of the NAS (approximately 6000 flights). These models are therefore not
well-suited to the types of policy-oriented, macroscopic issues that AND is designed
to address.

Analytical models

Turning to analytical models, Peterson, Bertsimas and Odoni (1995) [40, 41] have
used a combination of (i) a deterministic fluid-flow approach and (ii) a semi-Markov
model of airport capacities to model non-stationary queuing networks configured in a
hub-and-spoke configuration with a single hub airport at its center. Like AND, that
model accounts for changes in demand rates caused by earlier delays, but is limited
by the fact that it considers only a single hub.

Long et al. (1999) [32] developed a national-scale airport network model, LMINET,
which has some features similar to those of AND, including the modeling of individ-
ual airports as $M(t)/E_k(t)/1$ queuing systems. However, the approximation approach
they used to compute numerically the queuing statistics is completely different from
the one used in AND and does not obtain estimates of the state probabilities of the
queue at each airport.

To obtain the service characteristics at each airport, rather than estimating the
service rates at individual airports through historical data, Long et al. developed sepa-
rate airport capacity models to estimate service rates. These models generate arrival
and departure capacities as functions of surface meteorological conditions (ceiling,
visibility, wind speed and direction, and temperature) and the arrival and depar-
ture demand by introducing aircraft separations that meet all applicable rules—e.g.,
miles-in-trail requirements. Although, this approach is correct for the estimation of
the theoretical capacity of an airport, it is not accurate for the estimation of the ser-
vice rate of an $M(t)/E_k(t)/1$ queuing system. In addition, they use separate methods
to estimate the service rate and the Erlang order of the $M(t)/E_k(t)/1$ system.

Furthermore, LMINET does not use information on aircraft itineraries and there-
fore does not capture the propagation of delays through the network by tracking individual aircraft as they fly their daily routes. Recently, LMINET2 (Long et al., 2009 [33])—an upgraded version of the original LMINET—has addressed this particular issue, but the estimation of the queue statistics remains completely different from AND’s.

Other models

Sengupta et al. (2009) [45] recently described a family of queuing network models of the NAS designed to facilitate the study of the effectiveness of 4-dimensional trajectory-based operations in reducing traffic delays. Their models do not consider the propagation of delays due to the disruption of the scheduled itineraries of individual aircraft. Their paper also includes a good review of other simulation and queuing models of airport and airspace networks developed for various applications different from those of the AND model. Finally, adopting an entirely different methodological approach, Xu et al. (2008) [58] have used regression models and historical data to derive estimates of the amount of locally generated delay, locally “absorbed delay”, and propagated delay associated with individual major airports in the United States.

1.2.2 Modeling the effect of infrastructure improvements in the NAS

The main purpose of macroscopic models of the Air Transportation System is the evaluation of hypothetical developments such as proposed infrastructure improvements at a local, regional or network-wide scale and policies aimed at mitigating congestion. For instance, an extensive literature exists that attempts to assess the benefits of various NextGen and SESAR programs. We briefly describe here a small but representative sample of that literature in order to illustrate the type of research taking place in this area.

Long et al. (2009) [33] evaluated the delay savings resulting from proposed infrastructure improvements and NextGen technologies at major US airports, as envisioned
by the FAA, for a forecast flight schedule in 2025 using LMINET2. More recently, Post et al. (2011) [42] developed a policy-oriented framework for assessing the circumstances under which government financial incentives should be offered to airlines in order to accelerate equipping their aircraft with NextGen avionics. In order to perform a cost-benefit analysis they used the NASPAC simulation model to estimate the potential benefits to the system, in terms of reduced delays.

Much of the existing literature in this area, focuses on the impact of long term traffic growth on network delays. For this purpose, starting from traffic growth forecasts from national or multi-national Civil Aviation Authorities, such as the Federal Aviation Administration and the Eurocontrol, these studies develop ad-hoc networks of flights which they then test on one of the existing NAS models. One such study titled "Network Congestion 2030" sponsored by the Eurocontrol [16], shows that without any improvements to the EU system, the average departure delay will double by 2030 and the ability to recover from disrupted states will deteriorate significantly.

On the same topic, Kotegawa et al. (2011) [30] propose a series of algorithms which aim to forecast how the US airline network will evolve until 2020. The different network structures they develop, were tested with NASPAC to evaluate the impact on network-wide delays.

1.2.3 Modeling the effect of demand management policies

The topic of demand management has always been one of the main areas of research in the air transportation field. Numerous papers have addressed issues related to slot control, slot allocation and congestion pricing, from many different points of view. A number of books provide a reasonably comprehensive review of this body of work. For example, Czerny et al. (2007) [10] contains a compilation of papers on a broad range of research topics: a description of the current slot allocation system; determining the proper number of slots at an airport; the behavior of airlines in slot constrained environments; application of congestion pricing theory to airports; and using auctions for slot allocation.

Most of the existing literature, however, examines methods for applying demand
management policies or estimating the benefits of demand management at individual airports. Few researchers have looked at the impact of applying slot controls at an airport on system-wide delays and delay propagation. For example, Vaze (2011) [54] attempts to estimate the competitive responses of airlines to the application of slot limits at a busy US airport—in this case New York’s LaGuardia Airport. He adopts a game theoretic approach and estimates that the cost savings from the delay reductions resulting from low slot limits may be substantial for all airlines, while the revised airline schedules would still satisfy most of the existing demand.

Adopting a more general approach to the topic, Vaze and Barnhart (2010) [55] evaluated the extent to which delays can be reduced using demand management policies in the United States. For this purpose, they use integer linear programming to develop an airline network for only single airline that would serve all current passenger demand in the US. A bound on the minimum possible level of system-wide delays was obtained by testing the schedule of this single airline in the AND model (presented in this thesis), hence providing an estimate “of the inefficiencies in the usage of airport infrastructure in the domestic US due to competitive scheduling decisions by the airlines.”

Following a completely different approach, Morisset and Odoni (2011) [37] performed an extensive statistical comparison of delays and airport throughput performance in 2007 at a major European hub (Frankfurt) that operates under slot control and a major US hub (Newark) where no strict demand management policy was being enforced at the time.

1.3 Contributions of the Thesis

This thesis concentrates on modeling delays within an airport network and on assessing the system-wide effects of demand management policies and infrastructure improvements. The main contributions of our research are:

1. A stochastic and dynamic queuing network model (the AND model) for estimating delays and their propagation in large-scale airport networks such as the
ones in US and Europe.

2. An extensive calibration of the main components of AND and the model’s validation against real data.

3. New insights into the complex interactions that take place within networks of congested airports.

4. Extensions that expand in important ways the capabilities of the AND model to study features of the existing or future air traffic management systems.

5. A mathematical (mixed integer) optimization model for revising airline schedules in response to the application of slot controls at busy airports.

6. An example of a detailed assessment of the benefits of slot control using the AND model and the results of the optimization model above.

These contributions are summarized in the following sections.

1.3.1 Airport Network Delays model

The AND model developed in this research is aimed to bridge the gap between microsimulations that track aircraft itineraries but require extensive effort both in terms of input preparation and computation, and macroscopic models (simulations and analytical) that are simple to use, but typically lack aircraft itinerary tracking capabilities and credible queuing models of airport congestion.

The advantages and contributions of the AND model are several. First, because of its analytical queuing engine, AND does not require multiple runs to generate estimates of its various performance metrics. Along with the simplicity of its required inputs, this gives it a great computational advantage over simulations. This advantage, moreover, increases with network size. Because it is fast and has simple input requirements, the model can be used to explore at a “macroscopic”, approximate level a large number of scenarios, alternatives and policies.
In contrast to Long et al. (see Section 1.2.1), AND’s Queuing Engine not only approximates extremely well an exact \( M(t)/E_k(t)/1 \) queuing system, but has also been shown to accurately model congestion at individual airports. As will be shown in Chapter 3, AND’s Queuing Engine was used to model the departure process at two different busy US airports and provided very accurate estimates of delay. In this connection, we describe a new methodology for estimating the service characteristics of an airport that is modeled as an \( M(t)/E_k(t)/1 \) queuing system.

The Queuing Engine of AND also provides estimates of the state probabilities at each airport, so that one can estimate such statistics as the percent of flights that are more than 15 minutes late—a measure by which the on-time performance of airlines is often judged.

The Delay Propagation Algorithm is another contribution of the model. Its modular design and heap data structure has made it possible to develop several important extensions to the model without reducing the model’s computational efficiency. In fact, in Chapter 4 we implement an algorithm that emulates, at a high level of detail, the execution of Ground Delay Programs in today’s ATM system, without affecting AND’s computational efficiency.

Furthermore, the structure of AND is such that it allows the use of other analytical queuing models as its queuing engine, thus providing the opportunity to study the impacts of NextGen improvements that reduce stochasticity. In this respect, an entirely deterministic queuing engine has been implemented within AND as an alternative option to the stochastic queuing engine. Although there have been many studies in which each of these two types of queuing models have been used to obtain estimates of flight delays at individual airports (see Simaiakis and Pyrgiotis (2010) [49], Long et al. (1999) [32] and Malone (1995) [34] for stochastic models and Hansen (2002) [24] for deterministic ones), and a few studies where results from these two types of models have been compared (Hansen et al., 2009 [25]), there has never been a comparison within the context of a network. This is accomplished in this work in which we study a small set of different scenarios and compare the performance of the US airport network under the same conditions (flight schedules and airport
capacities) with the deterministic and with the stochastic versions of AND.

Another contribution of this research is the calibration of the AND model. We perform an extensive statistical analysis of the minimum ground turn-around times of aircraft, one of the fundamental variables that determine delay propagation. The analysis provides estimates of the minimum turn times of an aircraft, and of the slack embedded in airline schedules, as a function of the airline operating the aircraft, the type of airport where the connection takes place and the aircraft type. The data used in the analysis have been filtered carefully in order to focus on connections that have potentially tried to utilize the slack in their scheduled turn time. In contrast, other models of the NAS adopt rather simplistic approaches estimating minimum turn times. For example, LMINET2 estimates minimum ground turnaround times only as a function of a classification of aircraft with respect to their number of seats and sets the value of the minimum turnaround times equal to the lowest 5th percentile in published airline schedules. A very similar approach is adopted by NASPAC.

1.3.2 Insights and Applications

The AND model has generated results that provide insights into the complex interactions that take place when a significant subset of the system of airports operates under the strain of widespread congestion. These interactions may have quite surprising effects and may benefit certain types of flights or airlines, while penalizing others. For example, delay propagation tends to "smoothen" daily airport demand profiles and push more demands into late evening hours. Such phenomena are especially evident at hub airports, where certain flights may benefit considerably (by experiencing reduced delays) from the changes that occur in the scheduled demand profile as a result of delay propagation. Several of the observations detailed in an example presented in Chapter 3 of this thesis are made for the first time, to our knowledge.

In addition the application of AND to both the US and European air transportation systems, as presented in Section 3.5, offers a basis for the comparison of some fundamental operational differences between the two systems.

In order to provide a more accurate representation of the National Airspace Sys-
tem, we have also extended AND to capture a fundamental operational tool of the Air Traffic Management system. Ground Delay Programs (GDP) constitute a central coordination mechanism built on the idea that it is safer and less expensive to sustain delays on the ground, before takeoff, than in the air. When a GDP is initiated, flights are assigned a “ground-holding time” according to their projected time of arrival at the destination airport. The algorithm developed for AND adheres to the rules of the GDP process that is currently in place in the United States and replicates the steps of: initiating a GDP; assigning landing and takeoff slots to GDP-affected flights; and deciding on whether to terminate or continue the GDP. The slot assignment step includes the assignment of Expected Arrival Control Times and Expected Departure Control Times to all affected flights, slot swapping between flights of the same airline and schedule compression. The algorithm has been validated against real data and further improves the delay estimation capabilities of AND.

In another application the deterministic version of AND, described in the previous section has been used to provide upper bound estimates of potential delay savings that may result from one of the features associated with the Next Generation Air Transportation System (NextGen), the widespread implementation of 4D aircraft trajectories. A direct benefit of 4D trajectories is that the increased precision of flight paths and improved predictability of the instants when demands will occur at airports will lead to more efficient operations during periods of congestion. One way to study the effects of this increase in efficiency is by reducing the stochasticity of the queuing models of airports. While the current terminal airspace operations are reasonably accurately modeled by the stochastic queuing engine of AND, high precision 4D-trajectory operations, as envisioned in NextGen, may be better modeled through a deterministic queuing engine. In other words, by comparing the difference in the delay estimates generated by the stochastic and deterministic models, we can obtain an upper bound for the delay savings that may arise just from the reduction in stochasticity, without taking into account any alterations to airport capacities.
1.3.3 Slot control model

As noted earlier, existing research on the application of demand management policies at airports follows two principal directions: a) modeling the reaction of airlines, in terms of schedule changes at individual airports where a demand management policy is applied, and b) performing statistical analyses that compare otherwise similar airports, which are subject to different demand management policies. However, no previous work has examined the effect that demand management policies at a single or multiple airports may have on delays experienced throughout a national or international airport system.

In contrast to the rest of the world, most of the airports in the United States do not operate under a slot control system. Airlines are allowed to schedule an arrival or departure at any time they wish at all but a few airports. Our contribution to this topic begins with an extensive statistical analysis of the operations schedule at one of the busiest hubs in the United States, Newark Liberty International Airport, and the relationship between scheduled demand and airport capacity there. This leads to an argument suggesting the potential usefulness of slot-controls at that airport and to the conclusion that there should be separate slot limits for arrivals and for departures, rather than just limits on the total number of operations.

The main contribution of the thesis in this area is a Mixed Integer optimization model that generates a new flight schedule in response to reductions in the number of available slots at busy US airports. Given an initial schedule that is created through the airlines’ scheduling choices without any slot constraints (the current schedule), the model produces a feasible schedule that obeys the slot limits that are specified for an airport. At the same time, the model respects all aircraft itineraries, so that aircraft fly their original routes through the network. In addition, it preserves all passenger connections, in the sense that any passenger originally scheduled to connect at the slot-controlled airport will also be able to connect between exactly the same two flights under the new schedule.

We test the optimization model at Newark and conclude that, with a small sched-
ule displacement that never exceeds 30 minutes for any flight, it is possible to obtain a feasible schedule that obeys slot limits that are as low as the IFR capacity of the airport. Furthermore, we use the AND model to test the effect of several revised flight schedules, produced under various slot limits, on the local and network-wide delays under several different capacity conditions. It is shown that, when setting the slot limits near the IFR capacity of the airport, the delay savings, both local and system-wide, can be very large. This can be achieved without “sacrificing” any demand and by making only small changes to the original schedule at the busy airport.

1.4 Thesis outline

The main body of this thesis consists of five chapters: In Chapter 2 we provide a detailed description of the core of the AND model. Specifically, we describe the decomposition approach used in AND, that iterates between a queuing engine — that estimates delays through a queuing model — and a delay propagation algorithm. We also describe in detail the queuing engine and the structure of the delay propagation algorithm. Finally, we provide an extensive discussion of the main assumptions of the AND model.

In Chapter 3 we focus on the development of the various components of the AND model, on the validation of the model and on some of the insights obtained by testing several traffic and capacity scenarios. We begin by describing briefly the software development of AND and its application to the US airport network. We also revisit some of the assumptions discussed in Chapter 2 for a more detailed discussion. We then present a statistical analysis of the ground turnaround times of aircraft. We continue by modeling the departure process at two major airports as an $M(t)/E_k(t)/1$ queuing system, using the same approximation as the one used by AND's Queuing Engine. This serves as a validation of the Queuing Engine with real data. We then expand our validation efforts by testing AND for several days of operations at the 34 busiest US airports in 2007 and comparing the system-wide results from AND with actual observations of delays. We also discuss several important insights concerning
the propagation of delays in the network. Lastly, we describe the development of a European application of the AND model and perform a preliminary comparison of some performance characteristics of the US and EU networks.

Chapter 4 focuses on a couple of important extensions of the AND model. We describe, first, how we model Ground Delay Programs in AND, and demonstrate the performance of the GDP model through a set of tests. We also introduce an alternative queuing engine that estimates delays for AND under the assumption of deterministic demand and service rates. We compare the results of AND when run with this deterministic $(D(t)/D(t)/1)$ queuing engine against the results obtained with the original stochastic $(M(t)/E_k(t)/1)$ queuing engine. We also show how the two queuing engines could together provide approximate estimates of the benefits that might accrue from the use of 4D trajectories in the Next Generation Air Transportation System.

We begin Chapter 5 by summarizing current demand management practices at airports worldwide. We then perform an analysis of the relationship between the scheduled demand at Newark’s Liberty International Airport (EWR) and the airport’s capacity under different runway configurations and weather conditions. This analysis underlines the potential importance of reducing the number of slots available to airlines at EWR. We present next a Mixed Integer optimization model that generates a new flight schedule at an airport when the number of available slots is reduced, while respecting aircraft connectivity and passenger connectivity constraints. Finally, we use the AND model and the optimization model to test the effect of several flight schedules, produced under various slot limits, on local and network-wide delays for several different capacity conditions in the system.

Chapter 6 concludes this work by summarizing our findings and the main conclusions of our research. We also discuss possible future directions of this research.
Chapter 2

The Airport Network Delays Model

Air transportation systems rely increasingly on highly connected networks of airports where disruptions at one airport may affect other nodes of the network. The main goal of this research is to develop a macroscopic and fast, stochastic and dynamic model of a network of major airports, in order to obtain an assessment tool for network-wide impacts of policy changes and local or regional infrastructure improvements.

The Airport Network Delays (AND) model is based on the fundamental observation that airlines fly their aircraft on daily scheduled itineraries that require visits to a sequence of airports. Given the scheduled itineraries of all the commercial aircraft that fly within a regional or national system of airports, it should then be possible to trace the propagation of delays from airport to airport: if a particular aircraft is scheduled to fly from Airport A to Airport B and then to Airport C and departs from A with a long delay, part or all of that delay will be propagated downstream and result in a late arrival of the aircraft at B and subsequently at C. To operate properly, a model that captures this process should be able to (a) compute delays at each individual airport and (b) trace how delays at individual airports spread through the aircraft itineraries to the other airports in the system. Moreover (a) and (b) should be performed efficiently, if the model is to be of practical use. In this section we provide a detailed description of the core of the AND model, which is comprised of
a stochastic and dynamic delay estimator, derived from queuing theory and a delay propagation algorithm.

2.1 Operation of the Airport Network Model

The Air Traffic Management (ATM) system is highly complex with many components and participants affecting its performance. The AND model focuses on a subset of the ATM system that consists of the airports where most delays currently occur in the United States. AND models these airports in the ATM system as nodes in a queuing network, while the links of the network are modeled by the flight itineraries of aircraft. Arrivals at the queues consist of aircraft requesting permission to land or take off. The requested service is provided by each airport’s runway system.

Given the above queueing characterization of the ATM system, we provide a specific example of how a single aircraft travels through it. Consider the following partial itinerary. A Delta Airlines (DL) aircraft is scheduled to fly from New York’s LaGuardia Airport (LGA) to Atlanta’s Hartsfield Airport (ATL) via one of DL’s hubs, Detroit (DTW). Specifically, assume the aircraft is scheduled to depart LGA at 6:44 a.m. and arrive at DTW at 8:39 a.m. local time. This same aircraft is then scheduled to depart DTW at 9:35 a.m., arriving at ATL at 11:37 a.m. local time.

If this aircraft incurs little or no delay on takeoff or landing at any of the airports in this itinerary, this means there is little or no local congestion, nor any delay propagation from LGA to DTW or from DTW to ATL.

Consider now the case in which there is congestion at the hub airport in Detroit. This could be caused by the combination of a normal peak in demand in the morning, and a decreased service rate due to fog at the airport. Assume the flight takes off on time from LGA, but congestion at DTW causes the aircraft to land 20 minutes behind schedule at 8:59 a.m. Assuming that 10 minutes of slack are built into the fifty-six-minute scheduled time on the ground at DTW, the aircraft will join the queue of departing aircraft 10 minutes behind schedule at 9:45 a.m. Note that the aircraft does not physically need to be in the departure queue. It could be waiting at its gate,
or be shuttled off to a holding area. Assuming that the congestion has not yet cleared at DTW by 9:45am and the aircraft incurs another 30 minutes of delay waiting to depart, the aircraft departs 40 minutes later than its original scheduled departure time, at 10:25 a.m. Assuming 5 minutes can be made up in flight to ATL and there is no waiting for landing at ATL, the aircraft arrives there 35 minutes behind schedule, at 12:12 p.m.

An important point in this example is that although the aircraft arrives late at Atlanta, the aircraft’s lateness has nothing to do with the weather conditions or traffic congestion at ATL. The lateness is a result of the congestion at DTW, a hub airport. One hub’s congestion may potentially affect tens of other airports in North America. This type of interaction is a key characteristic of hub-and-spoke networks, and is modeled by AND.

2.1.1 Algorithmic Approach

Let $A$ be the set of airports in the network. We assume the demand rate at the network’s nodes is periodic, with period $T$. Typically, but not necessarily, $T$ is equal to a 24-hour period, beginning at a time when there is little air traffic activity throughout the network (such as 4 a.m. Eastern time in the United States). $T$ is subdivided into $m$ sub-periods, $h_1, h_2, \ldots, h_m$ of equal length. We have typically used $m = 96$, i.e., each sub-period is 15-minutes long, but other sub-period lengths may be more appropriate in different application contexts.

The AND model employs a combination of analytical and algorithmic approaches to estimate local delays at nodes in the network and to propagate delays system-wide. The model iterates between its analytical Queuing Engine (QE) and its algorithmic part, the Delay Propagation Algorithm (DPA), as illustrated in Figure 2-1. The analytical part, QE, employs a stochastic and dynamic queuing model, DELAYS, to compute delays at each airport, as if the airport operated in isolation. (Alternative QEs can be used, if desired, as described in Chapter 4.) The DPA “propagates” delays from each individual airport to the rest of the network by tracking how individual aircraft are affected by local delays and updating demand profiles at every airport.
affected by delays taking place at upstream airports.

In addition to the QE and DPA, the AND model includes a pre-processor which prepares all of the inputs for the QE and DPA, as shown in Figure 2-1. The aircraft itineraries contain information about how an aircraft travels through the network of airports over the period \( T \). Specifically, it contains scheduled arrival and departure times, the slack time at each airport, and the immediate predecessor and successor flights for each visit to an airport.

Starting at the beginning of the day, the QE is run for every airport in the network separately. Expected delays on takeoff and landing are calculated for every flight. The DPA determines whether delays incurred by any flight are significant enough to warrant propagation, i.e., to affect the original demand rates for any sub-period, \( h_j \), at any airport in the network. Let \( t^* \) be the earliest time at which this propagation condition is met. All airport demand profiles for all sub-periods \( h_j \) terminating before \( t^* \) are unaffected by delays occurring at or after \( t^* \). All flights scheduled to take place during all these unaffected periods \( h_j \) are then classified as “processed” by AND and will not be affected by any subsequent iterations of the algorithm. DPA then adjusts the arrival and departure times of all unprocessed flights affected by the propagation, as well as the demand rates at affected airports. The QE is then rerun using the updated demand rates and aircraft itineraries. This will, in turn, identify a new
earliest time, $t^{**}$, when the propagation condition is again met and a return to the DPA will be necessary. Note that $t^{**}$ must be greater than $t^{*}$, i.e., closer to the end of the time horizon, $T$. This alternating sequence of QE—DPA—QE—DPA— ... continues until all delays are eventually propagated through the system and the model reaches the end, $t = T$, of its time-horizon.

### 2.1.2 The Queuing Engine

The AND model treats individual airports as nodes of a queuing network. Arrivals to each queuing system are aircraft requesting service at the corresponding airport, i.e., permission to land or to take off. To avoid confusion, we refer to the rate of requests for landings and takeoffs as the demand rate.

The service at each queuing system is provided by each airport’s system of runways. Depending on the airport layout, each runway system, or runway configuration, is modeled either as a single server that serves both arrivals and departures (Figure 2-2b), or as two separate servers, one for arrivals and one for departures (Figure 2-2a), as it will be described in greater detail in Section 3.1.1. An aircraft requires two services at each airport it visits, one for landing and the other for takeoff. It is assumed that demands for access to the runway system are served according to a first-come-first-served (FCFS) discipline. The service rate at each airport/queuing system is equal to the expected number of landings and/or takeoffs that can take place there per unit of time under continuous demand conditions. This service rate is usually referred to as the capacity of the runway system. Infinite waiting line capacity for queuing aircraft is assumed to be available at each airport, so that no demands are ever denied access. Physically, aircraft waiting to take off are queued on the airport’s surface (taxiways and, possibly, apron stands), while aircraft waiting to land may be queued in or near the airport’s terminal airspace, as well as on the ground at other airports, when a Ground Delay Program for the airport of destination is in effect.

We now describe the QE in detail. In order to estimate delays at individual airports, we use a queuing model with a non-stationary Poisson arrival process, time-dependent $k^{th}$-order Erlang service-time distribution, a single-server, and in-
Figure 2-2: Conceptual representation of a single airport as a queuing system.
finite waiting room, denoted as a $M(t)/E_k(t)/1$ system in queuing theory. We write the equations describing the dynamic evolution of the queuing system and solve these equations numerically for any given profile of demand and service rates. We assume that both the arrival server and the departure server of each airport can be modeled as $M(t)/E_k(t)/1$, possibly with different service rates and Erlang orders. Hence, we only need to describe one queuing model.

To begin with, let us assume that the queuing system has a capacity of $N$ customers (i.e., landing and departing aircraft in our case), with one customer receiving service and up to $N-1$ waiting in queue. Such a system has $kN + 1$ states, with each state representing a “stage of work”. Figure 2-3 shows a state-transition diagram for a $M(t)/E_2(t)/1$ system (i.e., for the case $k = 2$) that can hold up to $N = 4$ customers. The system has 9 states: State “0” represents the empty system and State “2” the state in which there is only one customer in the system, who is occupying the server with no stage of work completed to this customer, i.e., with two stages of work still to be completed; State “5” indicates that (i) there are 3 customers in the system, one receiving service and two waiting and (ii) one of the two stages of work to the customer receiving service has been completed. Customers who find the queuing system in State “7” or in State “8” are rejected (and lost) because of lack of queuing space.

In general, given a demand rate $\lambda(t)$ and a service rate $\mu(t)$, the following set of first-order differential equations (often referred to as the “Chapman-Kolmogorov equations”) describe the evolution over time of the $M(t)/E_k(t)/1$ queuing system, when the system’s holding capacity is equal to $N$ customers (i.e., aircraft), including the one being served:
\[ P'_0(t) = -\lambda(t)P_0(t) + (k\mu(t))P_1(t) \]  
\[ P'_i(t) = -(\lambda(t) + \mu(t))P_i(t) + (k\mu(t))P_{i+1}(t), \quad 1 \leq i \leq k \]  
\[ P'_i(t) = \lambda(t)P_{i-k}(t) - (\lambda(t) + \mu(t))P_i(t) + (k\mu(t))P_{i+1}(t), \quad k < i \leq (N-1)k \]  
\[ P'_i(t) = \lambda(t)P_{i-k}(t) - (k\mu(t))(P_{i+1}(t) - P_i(t)), \quad (N-1)k \leq i \leq kN-1 \]  
\[ P'_{kN}(t) = \lambda(t)P_{k(N-1)}(t) - (k\mu(t))(P_{kN}(t)) \]

where \( P_i(t) \) is defined as the probability of the system being in stage \( i \) at time \( t \) and \( P'_i(t) \) denotes the derivative of \( P_i(t) \).

The above set of equations, 2.1, describes, again, a system that can hold up to \( N \) customers and thus has \( kN + 1 \) states, i.e., stages of work. In order to approximate an infinite-capacity system, one has to set the queue capacity, \( N \), of the system to a sufficiently large number so that the probability that the system is full at any time is very small. A practical problem exists, however, in the case of \( M(t)/E_k(t)/1 \) models: for large \( k \) (i.e., when the service times have a small coefficient of variation or, in practical terms, are nearly constant), the number of equations in 2.1 becomes very large, even for modest values of \( N \), and the numerical solution of the system of equations can be very time-consuming. (Note, as well, that the AND model may contain many such queuing systems, specifically as many as the airports that have been included in the network modeled.) For this reason, the Queuing Engine of AND uses an approximation scheme to the \( M(t)/E_k(t)/1 \) system, due to Kivestu [28], that solves a set of \( N + 1 \) difference equations (independent of \( k \)), instead of the system of \( kN + 1 \) Chapman-Kolmogorov equations. These resemble the classical difference equations that describe the evolution of \( M(t)/G(t)/1 \) systems, except that the “epochs”, as defined in the standard approach [31], are chosen differently.

Let \( p_j(t_i) \) denote the state probability that there are \( j \) aircraft in the queuing system at the instant when the \( l^{th} \) aircraft completes service. (Note that all \( j \) aircraft were waiting in queue, since a service has just been completed, and that the first one of these will occupy the single server next.) In addition, define \( \alpha_l(r) \) to be the
probability that \( r \) aircraft will arrive at the queuing system for the purpose of receiving service during the service time of the \( l^{th} \) aircraft. Then the difference equations we solve are:

\[
p_j(t_{i+1}) = p_0(t_i)\alpha_{l+1}(j) + \sum_{i=1}^{j+1} p_i(t_i) \ast \alpha_{l+1}(j - i + 1), \ j = 0, 1, \ldots, N
\]  

(2.2)

where

\[
t_{l+1} = t_l + \frac{k + 1}{k} \left( \frac{1}{\mu(t_i)} \right)
\]  

(2.3)

\[
\alpha_{l+1}(r) = \frac{(\lambda(t)/\mu(t))^r e^{-\lambda(t)/\mu(t)}}{r!}
\]  

(2.4)

The calculation of the state probabilities of the queuing system at any given airport \( a \in A \), leads to the estimation of the expected number of aircraft in the departure queue, \( L_{a,q}(t) \), and the expected waiting time in queue, \( W_{a,q}(t) \) from the following relationship:

\[
W_{a,q}(t) \approx \frac{L_{a,q}(t)}{\mu(t)} = \frac{\sum_{j=1}^{N} (j - 1)p_{a,j}(t)}{\mu(t)}
\]  

(2.5)

Under Kivestu’s approximation (and more specifically as implied by Equation 2.5) a flight arriving at the queue will be served according to the instantaneous service rate that exists at the airport at the moment it joins the queue. However, if the waiting time in the queue is long, the service rate of the airport might change during the time the flight is waiting in the queue. Hence, a fundamental assumption of the approximation is that the service rate at any given airport changes slowly relative to the length of time aircraft wait in queue. Typically, significant service rate changes at airports do indeed occur rather slowly over the course of a day. It may take more than two hours for capacity to go from a high value (VFR capacity) to a low one (low IFR capacity) or vice versa. By contrast, typical waiting times are less than 20 minutes and even under conditions of high congestion waiting times are of the order of thirty to ninety minutes. As well, it typically takes approximately 10-30 minutes for a change in runway configuration that will result in a large change in the service
rate. When such configuration changes take place for non-weather-related reasons (e.g., for noise-balancing purposes) the capacity change is typically small and the configuration change is planned for low demand periods.

The approximation will become inaccurate when long waiting times are combined with sharp changes in capacity. For example, when an airport is operating in bad weather for the whole morning and afternoon, creating very long delays, and the weather rapidly improves in the evening. In such cases an aircraft scheduled to arrive in the afternoon will be actually served at a different service rate than the one existing when it joined the queue. In order to correct for such sharp changes in capacity Gupta [23] has introduced the notion of effective service rate, $\mu_{eff}(t)$, in the estimation of delays in Kivestu’s approximation. Gupta deals with this issue by calculating the waiting time of aircraft $k$ in the queue according to an effective service rate, $\mu_{eff}^{k,j}(t)$, that takes into account the different service rates to be experienced by the $j$ aircraft already in the queue at the time of arrival of the $k^{th}$ aircraft.

Extensive computational experiments performed by Malone [34] and by Gupta [23] indicate that Kivestu’s approach approximates very accurately the exact solution of $M(t)/E_k(t)/1$ systems and is much faster than the exact (numerical) solution approach. The Kivestu approach has been implemented in the DELAYS software including Gupta’s effective service rate. It computes all the time-dependent delay statistics for a 24-hour period at a busy airport in only a few milliseconds on a typical modern laptop. In addition to approximating accurately the low moments and central moments of the queuing statistics, DELAYS also provides accurate estimates of the state probabilities $p_j(t)$ over time.

2.1.3 The Delay Propagation Algorithm

The Delay Propagation Algorithm (DPA) accounts for the interactions taking place in a network of airports by propagating any significant delays that are incurred at any particular airport to “downstream” flights and airports. Specifically, the DPA performs four functions:
1. Determines if significant delays occur (a delay is “significant” if it will propagate downstream).

2. Propagates delays between consecutive flights performed by the same aircraft.

3. Adjusts the arrival and departure times of any delayed flights.

4. Updates airport hourly demand rates.

In this section, the logic of these four functions is described in detail, after introducing some notation.

As noted already, we assume that the demand rates at the network’s nodes (the airports) are periodic, with period $T$ (usually 24 hours) and vary over the period $T$. The demand rates are based on the schedule of airport operations (see further below). For any given airport $a \in A$ in sub-period $h_j, j = 1, 2, \ldots, m$, the expected number of demands and the expected number of service completions when the server is continually busy are denoted by $\lambda_a(h_j)$ and $\mu_a(h_j)$, respectively. $\Delta h$ denotes the length of each sub-period.

Because the AND model performs calculations over a finer-grain time scale than $\Delta h$, we distinguish between the above demand and service rates for the sub-period $h_j$ and the instantaneous demand and service rates $\lambda(t)$ and $\mu(t)$, as used by the Queuing Engine. Define $index(t)$ as the function that maps the time of day, $t$, into the sub-period, $h_j$, to which it belongs. The relationship between $t$ and $h_j$, is given by:

$$ h_j = index(t) = \left\lfloor \frac{t}{\Delta h} \right\rfloor $$

(2.6)

An aircraft’s itinerary consists of a set of flights between pairs of consecutive airports. For example, the itinerary $\{A, B, C, D\}$ consists of a sequence of three flights, A-to-B, B-to-C, and C-to-D. (Note that, according to our definition and without loss of generality, each flight consists of a single “hop”.) $F$ is the set of all flights, $f$, that are scheduled to operate during period $T$ in the network. Figure 2-4 shows how each flight is accounted for in the context of AND. With reference to figure 2-4 we introduce the following nomenclature that will be used throughout this thesis:
\begin{align*}
o(f) & = \text{airport of origin of } f \\
d(f) & = \text{airport of destination of } f \\
SD(f) & = \text{scheduled departure time of } f \text{ from } o(f) \\
SA(f) & = \text{scheduled arrival time of } f \text{ at } d(f) \\
AD(f) & = \text{adjusted departure time of } f \text{ from } o(f) \\
AA(f) & = \text{adjusted arrival time of } f \text{ at } d(f)
\end{align*}

It should be noted that by “departure time” and “arrival time” we refer to the
time a flight will request permission to take off or land, respectively. This time does
not include any delay incurred in the process of landing or taking off. Thus, \(SD(f)\)
and \(SA(f)\) refer to the times when these requests would have been made according
to the original schedule of flight \(f\), while the adjusted times, \(AD(f)\) and \(AA(f)\), refer
to the times of the day when \(f\) will actually request to land or takeoff. While we
set initially \(AD(f) = SD(f)\) and \(AA(f) = SA(f)\) for all \(f \in F\), some flights may
have \(AD(f) \neq SD(f)\) and/or \(AA(f) \neq SA(f)\) at the end of an AND run, due to
delays suffered earlier in the day by the aircraft performing flight \(f\) within the network
of airports. Furthermore the departure and arrival times of every flight \(f \in F\) are
mapped into the respective sub-periods using Equation 2.6; e.g. the adjusted arrival
sub-period of \(f\) is given by \(h_{AA(f)} = index(AA(f)) = \left\lfloor \frac{AA(f)}{\Delta k} \right\rfloor\).

We assume that the immediate predecessor flight, \(f'\), of every flight \(f\) in every
aircraft’s itinerary is known, as are the parameters \(turn(f', f)\) and \(minturn(f', f)\).
The parameter \(turn(f', f) = SD(f) - SA(f')\) of the flight pair \((f', f)\) is the scheduled
“turnaround time” on the ground between the arrival of \(f'\) and the departure of \(f\).
Associated with each turnaround time there is also a “minimum turnaround time”,
\(minturn(f', f)\), which is the smallest amount of time necessary to “handle” (unload,
clean, refuel, load, etc.) the aircraft arriving as flight \(f'\) and get it ready to depart
as flight \(f\). The quantity \(slack(f', f) = turn(f', f) - minturn(f', f)\) indicates the
“slack” associated with the pair \((f', f)\). The flight \(f\) will still depart on time, as long

52
Figure 2.4: Flight breakdown.
as the arrival of $f'$ is delayed by less than $\text{slack}(f', f)$.

The "upstream" (departure or arrival) delay of $f$ refers to the portion of delay of $f$ that cannot be attributed to local congestion at the airport of arrival or departure but has been incurred at earlier destinations of the aircraft. Hence, we introduce variables $U_o(f)$ and $U_d(f)$ that correspond to the upstream delay of $f$ before departing from $o(f)$ and the upstream delay of $f$ before arriving at airport $d(f)$, respectively. Then, referring to Figure 2-4, the following set of equations describe the timeline of a flight in the context of AND:

\[
AD(f) = SD(f) + U_o(f) \tag{2.7}
\]

Final Departure Time($f$) = $AD(f) + W_{o(f), q}(AD(f))$ \tag{2.8}

\[
AA(f) = SA(f) + U_d(f) \tag{2.9}
\]

Final Arrival Time($f$) = $AA(f) + W_{d(f), q}(AA(f))$ \tag{2.10}

We turn next to Functions 1 and 2 in our list. We define sets $G$, $G_D$, and $G_A$, and $G'$, $G'_D$, and $G'_A$ such that $G \cup G' = G_D \cup G'_D = G_A \cup G'_A = F$. These sets indicate whether any particular flight $f$ has been "processed" by AND. Specifically, $G_D(f) = 1$ if the departure of $f$ has been "processed" by AND, and, in that case, $f \in G_D$; and $G_D(f) = 0$ if the departure of $f$ has not been processed and, in that case, $f \in G'_D$. Similarly $G_A(f) = 1$ if the arrival of $f$ has been processed (in which case $f \in G_A$) and 0 otherwise (in which case $f \in G'_A$). Finally, if both the departure and the arrival of $f$ have been processed, $G(f) = 1$ (and $f \in G$), otherwise $G(f) = 0$ and $f \in G'$. We call $G$ and $G'$, respectively, the sets of processed and unprocessed flights.

Whenever the DPA is called, e.g., at some time $t^*$, it must determine whether significant delays have occurred among the unprocessed departures and arrivals, $G'_D$ and $G'_A$. The delay is "significant" if there exists $f \in G'_D$ or $f \in G'_A$ such that $f'$'s departure or arrival time must be adjusted by an amount that will move the operation into a new sub-period, i.e. if $h_{(AA(f))} > h_{(SA(f))}$ or $h_{(AD(f))} > h_{(SD(f))}$. The necessary
adjustment to the departure time of \( f \) (that is to the time when take-off from an airport is requested) is determined by three quantities:

- \( AA(f') \), the adjusted arrival time of the flight \( f' \) immediately preceding \( f \) in the relevant aircraft’s itinerary

- \( W_{d(f'),q}(AA(f')) \), the expected delay that flight \( f' \) will incur at the arrival airport \( d(f') \)

- \( slack(f', f) \), the slack in the turnaround time between the arrival of \( f' \) and the departure of \( f \)

Then we have:

\[
AD(f) = \max [SD(f), SD(f) + (AA(f') - SA(f)) + W_{d(f'),q}(AA(f')) - slack(f', f)]
\]

Likewise, \( f \)'s adjusted arrival time depends on:

- \( AD(f) \), the adjusted departure time of \( f \)

- \( W_{o(f),q}(AD(f)) \), the delay that is incurred by \( f \) on takeoff

- The slack in the scheduled block time of \( f \), otherwise known as scheduled or (block time) “padding”.

Then we have:

\[
AA(f) = \max [SA(f), AD(f) + W_{o(f),q}(AD(f)) +
(scheduled\ block\ time\ of\ f) - (block\ time\ padding\ of\ f)]
\]

Given that at least one significant delay occurs, DPA determines when the earliest one of these delays takes place. All flights operating before the earliest significant delay will be unaffected by the delays to be propagated after this delay. Thus, all arrivals and departures occurring before this delay become processed arrivals and departures. The time and sub-period at which the earliest delay occurs is found as
follows:

\[ t^* = \min[\min_{f \in \mathcal{G}_{D}^P}(AD(f)), \min_{f \in \mathcal{G}_{A}^P}(AA(f))] \]  

\[ h^* = \text{index}(t^*) \]  

To carry out Function 3, DPA propagates delay by making one-step adjustments to the arrival and departure times of those flights that are immediate successors of the flights processed in the current iteration. That is, if flight \( f' \) is processed in the current iteration, and \( f \) is its immediate successor flight (and has not yet been processed), the departure time of \( f \) is adjusted using equation 2.11. Similarly, if only the departure of flight \( f' \) is processed in the current iteration, the arrival time of \( f' \) is adjusted using equation 2.12.

If the adjusted arrival time of \( f \) falls into a different sub-period \( h_j \) than the sub-period into which it was originally scheduled, then the demand rates at \( d(f) \) are adjusted, thus performing Function 4 of the DPA. Specifically, let \( s = \text{index}(SA(f)) \) and \( n = \text{index}(AA(f)) \), i.e., while the arrival of flight \( f \) was originally scheduled to take place in sub-period \( n \), this arrival is now expected to take place in sub-period \( s \), with \( n \geq s \). Then, if \( n > s \), we reduce \( \bar{\lambda}_{d(f)}(h_s) \) by one unit (one fewer expected arrival at the destination airport \( d(f) \) of flight \( f \) during sub-period \( s \)), and increase \( \bar{\lambda}_{d(f)}(h_n) \) by one unit. We perform an entirely analogous adjustment for departures, using \( o(f), SD(f) \) and \( AD(f) \). After the completion of the DPA process, the updated demand rates for every airport in AND are fed back to the QE which re-estimates delays for the remainder of \( T \), i.e. the remaining time period of interest (typically 24 hours).

\[ 2.1.4 \text{ The AND pseudo-code} \]

In order to integrate and summarize the functions performed within AND we present the following pseudo-code. It is divided into two parts the AND pre-processor and the main AND iteration, as shown earlier in figure 2-1.
begin AND Pre-processor:

Create aircraft itineraries:

- Connect flights operated by the same aircraft
- Initialize $SD(f)$, $SA(f)$, $slack(f', f)$, $padding(f)$

Set:

$$AD(f) = SD(f)$$

$$AA(f) = SA(f)$$

$$W_{o(f),q}(AD(f)) = W_{d(f),q}(AA(f)) = 0$$

$$GD(f) = GA(f) = G(f) = 0$$

Initialize airport demand rates $\bar{\lambda}_a(h)$, $h = 1, 2, \ldots, H$ and $\forall a \in A$

Initialize airport service rates $\mu_a(h)$, $h = 1, 2, \ldots, H$ and $\forall a \in A$

end AND Pre-processor
begin AND: Start at $t = t^* = h = h^* = 0$

while $t < T$

step 1: run QE $\forall a \in A$ and $t^* \leq t \leq T$ to obtain $W_{a,q}(t)$  \hspace{1cm} end step 1

step 2: run DPA

\begin{align*}
&\forall f \in G_D \text{ assign } W_{o(f),q}(AD(f)), \\
&\forall f \in G_A \text{ assign } W_{d(f),q}(AA(f)) \\
&\text{check if significant delays occur} \\
&\text{comment: is there a delay that cannot be absorbed} \\
&\text{by the ground slack or the "padding"} \\
&\text{if true then} \\
&\text{identify } t^* = \min[\min_{f \in G_D}(AD(f)), \min_{f \in G_A}(AA(f))] \\
&\text{process all operations (take-offs and landings) that occur before } t^*: \\
&\text{if } AD(f) \leq t^* \text{ \textbf{and} } AD(f) < t^* \textbf{ then} \\
&\quad G_D(f) = 1, G_A(f) = 0, G(f) = 0 \\
&\text{else if } AA(f) \leq t^* \\
&\quad G_D(f) = 1, G_A(f) = 1, G(f) = 1 \\
&\text{end if} \\
&\text{update:} \\
&\quad AD(f) \forall f \ni G_D(f) = 0 \\
&\quad AA(f) \forall f \ni G_A(f) = 0 \\
&\quad \text{according to:} \\
&\quad AD(f) = \max[SD(f), SD(f) + (AA(f') - SA(f)) + \\
&\quad \quad +W_{d(f'),q}(AA(f')) - \text{slack}(f', f)] \\
&\quad AA(f) = \max[SA(f), AD(f) + W_{o(f),q}(AD(f)) + \\
&\quad \quad +(\text{scheduled block time of } f) - (\text{block time padding of } f)] \\
&\text{end step 2:} \\
&\quad \text{set } t = t^* \\
&\text{else if } \text{false then} \\
&\quad \text{process all } f \ni G_D(f) = 0, G_A(f), G(f) = 0 \\
&\quad \text{update:} \\
&\quad \bar{x}_a(h) \forall h > h^* \text{ \textbf{and} } \forall a \in A \\
&\quad \text{according to the revised } AD(f), AA(f) \\
&\text{end if} \\
end AND}$
2.1.5 Performance Measures of Interest

Performance measures of interest can be classified as “local” and “network”. Local performance measures are those related to delays incurred at each individual airport during a particular visit by an aircraft using that airport, given the (possibly modified) demand rates over the course of the day.

Network performance measures account for the delays that are observed at individual airports, but cannot be attributed to local congestion because they have been incurred earlier in the day at airports previously visited by an aircraft. We refer to this type of delay as “upstream” or “propagated” delay, while in Europe the term “reactionary” delay is used. AND estimates the upstream delay associated with every arrival and departure at all the airports in the network throughout the day, as described in section 2.1.3 and by equations 2.7—2.10. Given these estimates, more composite performance measures (such as the total amount of upstream delay observed at a specific airport during the course of a day) can be easily computed. AND also routinely computes another network performance measure of practical interest, the “fraction of arrivals with expected delay greater than 15 minutes” at each airport and system-wide. This measure is analogous to the well-known statistic published by the US Department of Transportation, which is used in ranking the on-time performance of airlines in the United States. The difference is that the measure computed by AND refers to airport performance, not to airlines. However, airline-specific on-time statistics can also be computed by AND, if desired, since aircraft itinerary data include airline identification.

We present below a list of local and network performance measures that are calculated by the AND model.

Local performance measures include average delay (for specified parts of the day, e.g., for each hour, or for the entire day), maximum expected delay during the day, and the time-averaged number of aircraft waiting to land or takeoff. Let \( I \) be the number of epochs at which the probability distribution is recalculated over the course of the day, and \( I_h \) the epoch corresponding to the beginning of sub-period \( h \). Then
for every airport $a \in A$:

- Time-averaged daily local delay:
  \[
  \frac{\sum_{l=1}^{L} \lambda_a(t_l) W_{a,q}(t_l)(t_l - t_{l-1})}{\sum_{l=1}^{L} \lambda_a(t_l)(t_l - t_{l-1})}
  \]  \hspace{1cm} (2.15)

- Time-averaged local delay in a specified sub-period $h$:
  \[
  \frac{\sum_{l=i_h}^{l_{h+1}} \lambda_a(t_l) W_{a,q}(t_l)(t_l - t_{l-1})}{\sum_{l=i_h}^{l_{h+1}} \lambda_a(t_l)(t_l - t_{l-1})}
  \]  \hspace{1cm} (2.16)

- Maximum expected local delay for the entire period of interest:
  \[
  \max_{t_l, d \in \{1, \ldots, I\}} [W_{a,q}(t_l)]
  \]  \hspace{1cm} (2.17)

- Time-averaged number in queue:
  \[
  \frac{\sum_{l=1}^{L} \lambda_a(t_l) L_{a,q}(t_l)(t_l - t_{l-1})}{\sum_{l=1}^{L} \lambda_a(t_l)(t_l - t_{l-1})}
  \]  \hspace{1cm} (2.18)

Network performance measures include:

- Expected upstream delay: a measure of how much of the delay observed at an airport can be attributed to congestion incurred earlier in the day in the network. For airport $a \in A$, the total expected upstream delay during a day is given by:
  \[
  \sum_{\{f \in F|d(f) = a\}} [AA(f) - SA(f)]
  \]  \hspace{1cm} (2.19)

- Fraction of arrivals with expected delay greater than 15 minutes (or any other amount); for every airport $a \in A$ this is calculated in AND through the expression:
  \[
  \frac{\sum_{\{f \in F|d(f) = a\}} \delta_a(f)}{|\{f \in F|d(f) = a\}|}
  \]  \hspace{1cm} (2.20)
where

\[
\delta_a(f) = \begin{cases} 
1 & \text{if } AA(f) - SA(f) + \text{landing delay of } f > 15 \text{ minutes} \\
0 & \text{otherwise}
\end{cases}
\]  

(2.21)

This measure can also be calculated for individual airlines.

- Adjusted (or actual) demand rates at airports after running AND, \( \lambda^{\text{adj}}_a(h) \): this is a measure useful for comparisons with the original scheduled demand rates input into the model, \( \lambda^{\text{sch}}_a(h) \).

- Flight arrival delay: the total delay on arrival for each flight, defined as the sum of the upstream delay and the local delay of that flight. For a flight \( f \) arriving at airport \( d(f) \), this delay is given by:

\[
AA(f) - SA(f) + W_{d(f),q}(AA(f))
\]  

(2.22)

- The breakdown of delay between local and upstream sources: This is estimated from the quantities in equations 2.15 and 2.19.

### 2.1.6 Discussion of the Main Assumptions

We turn next to a review of the principal assumptions underlying the basic AND model described in this chapter. We begin with six assumptions which are fundamental to the dynamic and stochastic queuing engine described in Section 2.1.2:

**Assumption 1**

Demands at the runway threshold (for departures) and at the terminal airspace fixes (for arrivals) are modeled as a non-stationary Poisson process, i.e., the times between successive demands ("inter-arrival times" in queuing terminology) follow an exponential pdf with time-varying mean value.

**Assumption 2**
The service times follow a $k^{th}$ - order Erlang distribution with time-varying mean value.

**Assumption 3**

Each runway system, or runway configuration, is modeled by either one or two servers depending on the airport runway characteristics. In the two-server case it is assumed that one server is used by arrivals only and the other by departures, while in the single-server case arrivals and departures share the same resource (the runway capacity).

**Assumption 4**

Aircraft are served according to a First-Come, First-Served (FCFS) discipline.

**Assumption 5**

There is infinite waiting space for both the departing and arriving aircraft to queue up, so that no demands are lost.

**Assumption 6**

Each queuing system (i.e., each individual airport of the network) starts out empty at the beginning of each AND day.

Furthermore, the following three assumptions are fundamental to the overall modeling approach of AND:

**Assumption 7**

The en route airspace system is un-capacitated, i.e. does not cause any bottlenecks.

**Assumption 8**

Expected waiting times provide an adequate basis for updating schedules and downstream arrival rates.

**Assumption 9**

The slack in the ground turnaround time is known for every aircraft at every airport, as is the “padding” in the scheduled block time of every flight.
We address next the validity of these assumptions and discuss cases when these are violated.

Assumption 1 is a standard approximation in airport queue modeling which was first used by Koopman in 1972[29]. Its reasonableness for arrivals (defined to be contact with the terminal airspace control facility—TRACON in the United States—to join the landing queue) at most busy airports is supported by empirical data, as in Daniels[11] and Willemain[57]. Most operations at major airports are, of course, scheduled in advance. However, there is (i) considerable variability from day to day in the number of operations that actually take place due to general aviation flights, unscheduled operations, flight cancellations, etc, and (ii) large deviations from the scheduled times of the operations due to delays at “upstream” airports, variability of flight times due to weather and winds aloft, gate delays for departures, different “pushback” times, etc. The combined effect of these two types of uncertainty is to “randomize” the schedule sufficiently to render the non-stationary Poisson process a reasonable approximation of the actual demand.

A second related argument regarding Assumption 1 is that, at the network level, it is difficult in practice to find any major airport that receives more than 10% of its arrivals from any other airport during an hour of appreciable traffic. Figure 2-5, for example, shows how weakly Chicago O’Hare Airport (ORD) is connected to the rest of the network. We may observe that, with the exception of a few off-peak night hours, the arrivals from any given airport of origin do not exceed 6% of the total. Thus, the ORD arrivals stream is the sum of many small streams of traffic from other airports. This ensures the “weak connectivity” of the network (i.e., arrivals at each queue do not depend heavily on the output of any particular “upstream” airport), making it plausible to assume that demands at each airport in the network are non-stationary Poisson, without having to worry about the exact stochastic characteristics of the individual streams of traffic that contribute to each airport’s total demand. Malone [34] reports some interesting computational experiments demonstrating this point.

It should be noted that the assumption of non-stationary Poisson arrivals is most plausible at busy airports, where the arguments mentioned above apply best.
This is also true of the relationship between arrivals and departures which, at any given airport, are not independent: the number of departures in any hour is often strongly related to the number of arrivals in the preceding hours. However, at busy airports, when there are many operations in each hour, there is significant randomization of the sequence and the timing of specific departures because different aircraft and flights may spend greatly different amounts of time at the gates. Hence, demands by arrivals and departures at the queuing system (the runways) will appear to be independent of each other. This argument may break down when considering hours with few operations but, at busy airports, this applies to only a small fraction of the total number of operations and even then it is of little importance, since these flights suffer little delay anyway.

Regarding Assumption 2, Odoni and Hengsbach [39] were the first to propose an $M(t)/E_k(t)/1$ queueing system as a model that could be used by itself to compute approximate queueing statistics for airports. The Erlang ($E_k$) family of random variables captures a broad range of service time distributions, from the negative
exponential service times ($M$), obtained by setting $k = 1$, to constant service times ($D$) by setting $k$ to infinity. More important, the Erlang distribution is very useful for approximating more general distributions, such as those for service times at airports, by selecting an appropriate value of $k$, as can be seen in Figure 2-6.

In Chapter 3 we present results that demonstrate the validity of the Erlang distribution assumption for the departure service times at Newark and Boston airports. As we shall see in Chapter 3, an $M(t)/E_k(t)/1$ system approximated the current departure operations extremely well when tested in Boston and in Newark.

Assumption 3 is important for modeling airport operations. As will be discussed in Section 3.1.1, depending on the runway layout of the airport and the operational procedures there, arrivals and departures can be modeled either as the same queuing system or as two separate ones. In the second case, both servers are assumed to be $M(t)/E_k(t)/1$ queuing systems. The service characteristics of each server may be different but are not independent, as very often the departure service rate depends
on the arrival service rate and vice versa. We make further assumptions about this dependence in Section 3.1.1 and provide an analysis on the topic in Section 3.2.3.

A First-Come, First-Served (FCFS) queuing discipline (Assumption 4) is consistent with FAA procedures. Once an aircraft enters the departure queue or the arrival queue, it is hard (physically or due to ATC workload considerations) to overtake other aircraft already waiting in the queue. On the other hand, during highly congested periods, arrivals are often accorded priority over departures in practice, mainly for safety reasons. When, however, the departures queue becomes long in such cases, landings may be interrupted, so that a string of departures can take place, or the controllers may start alternating arrivals and departures on the runway. The FCFS assumption is thus adhered to only in an approximate sense in such circumstances. It can then be stated that, overall, the FCFS assumption is occasionally violated in the presence of heavy congestion. AND's estimates of waiting times, using the FCFS assumption, may present a somewhat distorted picture on such occasions, probably overestimating arrival delays and underestimating those for departures. However, for purposes of macroscopic approximate modeling this is only a second-order effect. We have therefore used a FCFS discipline independently of how much congestion is present. We make an exception only for the most severe cases of congestion, namely when Ground Delay Programs (GDP) go in effect. We have implemented for such cases a GDP algorithm, which effectively assigns to a subset of arriving flights specific time slots for use of the arrival runways, independently of what else happens at the airport(s) that is/are "running" a GDP. The GDP algorithm is described in detail in Chapter 4.

Assumption 5 is based on the fact that there is always sufficient physical space at airports for departing aircraft to queue up so that these aircraft are never rejected for service even when long queues are present. Physically, the departure queue might consist of aircraft waiting at the runway threshold, on the taxiways, or even at the gate. In cases of extreme congestion, the entire airport surface can act as waiting space for departing aircraft. Similarly, there is always sufficient capacity for the arriving aircraft to queue up in different parts of the airspace, anywhere between the
origin and the destination of each flight, or even on the ground at the airport of a flight’s origin when GDPs are in effect.

Assumption 6, stating that the queuing system starts out empty at the beginning of each 24-hour period, is based on the fact that most airports have no or minimal activity during nighttime hours (typically between 1 a.m. and 6 a.m.) while peak airport activity typically ends by about 10 p.m. Therefore, with the exception of major snowstorms, there is almost always enough time (and capacity slack) available at the end of each day to ensure that the system returns to rest, i.e., that any queues and aftereffects from the previous day will have dissipated by early in the morning of the following day. Thus, by setting 5 a.m. EST as the beginning of each day, we can model demands and queues on successive days as independent events.

Turning next to the three more general assumptions, the AND model does not include a representation of airspace sectors, assuming that airports constitute by far the most constraining bottleneck in the ATM system (Assumption 7). In view of the macroscopic scope of AND, this assumption is certainly valid for the ATM system in the United States. It is less valid in the context of the current European ATM system where, for a variety of technical and institutional reasons, en route airspace also plays a significant role as a generator of air traffic delays. However, even in Europe, the importance of en route airspace as a system bottleneck has been diminishing over the past decade and is likely to become even smaller in the future, under the Single European Sky program.

Assumption 8 states that the AND model uses expected waiting time in queue, calculated from the probability distribution for the number in queue, to compute point estimates of delays. Peterson et al (1995b) performed extensive computations to determine delays in networks of queues. They compared the results of using expected values against distributions of waiting times (using simulations) and found the results of the two approaches to be close. Therefore, for the purposes of macroscopic modeling, we believe that using expected waiting times to estimate delays and to propagate delays is an adequate approximation.
Finally, it is assumed that the amount of slack in the ground turnaround time in every aircraft’s schedule as well as the amount of slack in the scheduled block time of every flight ("padding") are known (Assumption 9). These values are an input to the AND model. As will be shown in Chapter 3, we have performed extensive statistical analysis to obtain estimates of the minimum turnaround time of different types of aircraft, operated by different airlines and separated according to hub and non-hub operations. Similarly, we rely on results from Skaltsas [50] for the estimates of the "padding" in the scheduled block time of US domestic flights. We shall explore this topic more extensively in Chapter 3.

2.1.7 Summary

We presented in this chapter the Airport Network Delays model, a stochastic and dynamic queuing network representation of air transportation networks. AND is a macroscopic tool aimed to support the study of the propagation of delays within a network of airports, as well as to assess the network-wide impacts of policy changes and of local or regional infrastructure improvements.

AND applies a decomposition approach, by iterating between a queuing engine — delay estimator — and a delay propagation algorithm. The queuing engine utilizes an approximation to the $M(t)/E_k(t)/1$ queuing system to estimate how delays occur at individual airports due to changes in the demand and/or the capacity at each airport during the course of a day. The delay propagation algorithm tracks the aircraft through the network in small time steps — typically of 15 minutes length. At every time step the algorithm assigns improved delay estimates to every flight that has completed an operation (take-off and landing) by that time, and updates the schedules of flights for the rest of the day. The updated schedules are used by the queuing engine to re-estimate the delays for the remainder of the day.

The main inputs to the model are the scheduled itineraries of all aircraft and the expected capacities of the airports, as well as the slack in the ground turnaround times of aircraft and in the block times (gate-out to gate-in) of all flights. The outputs of the AND model are the local and upstream delays as a function of time at each
airport, in addition to the estimated flown timetable of every aircraft that includes estimates of the delays that each flight has encountered.

We have also discussed the validity of all the principal assumptions of the Airport Network Delays model. The most important among them are that: a) arrivals at the airport queues follow a time-varying Poisson distribution, b) airport service times follow a time-varying Erlang distribution, c) each runway system is modeled as a single server or as two servers, depending on local conditions, d) the airspace system is uncapacitated and e) expected waiting times provide an adequate basis for updating flight schedules.
Chapter 3

Calibration, Validation and Testing of the AND model

In this chapter we focus on the development of various components of the Airport Network Delays model, on the validation of the model and on some of the insights obtained by testing several traffic and capacity scenarios with AND. We begin (Section 3.1) by describing briefly the development of AND and its application to a large network consisting of the 34 busiest commercial airports in the continental United States. In Sections 3.1.1-3.1.3, we revisit some of the assumptions listed and discussed in Section 2.1.6 to provide further details. More specifically, we explain why it is necessary to separate the arrival and departure queues at some airports and we address the issues that arise when arrival and departure service rates are interdependent. We then present a statistical analysis of aircraft ground turnaround times and discuss through a literature review the importance of padding of scheduled block times. We continue, in Section 3.2, by modeling the departure process at two major airports as an $M(t)/E_k(t)/1$ queuing system, using the DELAYS approximation described in Section 2.1.2. This serves as a validation of the DELAYS model with real data. We then expand our validation efforts by studying several days of operations in 2007 in AND and comparing the system-wide results from AND with real observations of delays. In Section 3.4, we discuss several important insights on network phenomena obtained while testing the AND model. Lastly, we describe the development of a Eu-
ropean AND model and perform a preliminary comparison between some performance characteristics of the US and EU networks.

3.1 Development of AND

The AND model has been programmed in Java and has been implemented for a network consisting of the 34 continental US airports listed in the FAA’s Operational Evolution Partnership (OEP) [18]. (This means all OEP airports, with the exception of Honolulu.) Figure 3-1 shows these 34 airports. In addition, a 35th “virtual airport” acts as an un-capacitated source and sink of traffic. The purpose of the virtual airport is to include in the AND network all flights of any aircraft that at some point in the day passes through any of the 34 airports: any flight between an airport in the 34-node network and some other (“external”) airport is taken into account by having the relevant aircraft fly to/from the virtual airport. Moreover, flights between two external airports are included in AND by having the aircraft fly from the virtual airport back into the virtual airport. Hence, we account for aircraft that may leave the network at some point during the day, visit one or multiple locations outside AND’s network and return to the network at a later time before the end of the day. Figure 3-2 shows an example of the flight sequence of an aircraft that exits the network at the end of its first flight and returns to the network at the end of its third flight of the day. When visiting the virtual airport, it is assumed that an aircraft will not experience any further delays. However, any delay incurred at any of the 34 airports will still be propagated through the virtual airport by adjusting the delay to account for the ground slack that corresponds to the aircraft type and airline operating the flight—more details on ground slack are given in Section 3.1.2. Hence the delay propagation equations presented in section 2.1.3, equations 2.11 and 2.12, for the virtual airport are revised as follows:
For a flight, $f$, arriving at the virtual airport:

$$AA(f) = \max[SA(f), AD(f) + (\text{scheduled block time of } f) - (\text{block time padding of } f)]$$

For a flight, $f$, departing from the virtual airport:

$$AD(f) = \max[SD(f), SD(f) + (AA(f') - SA(f)) - \text{slack}(f', f)]$$

Figure 3-1: Airport map of the US AND model.

Figure 3-2: Use of the virtual airport in AND.

On a typical day, approximately 35,000 flights are included in AND; roughly 12,000
of these flights (34%) are flown between some pair of the 34 airports, while 19,000 (54%) take place between one of the 34 airports and the virtual airport (as Flights 1 and 3 in Figure 3-2) and only 4,000 flights operate entirely outside the network (i.e., beginning and ending at the virtual airport) as part of aircraft schedules that at some point in the day visit the AND network (e.g., Flight 2 in Figure 3-2).

Four types of inputs are critical to the AND model:

1. **Aircraft itineraries:** For every aircraft performing commercial service within the network, this input indicates the detailed sequence of flights to be performed, including scheduled arrival and departure times for the aircraft at each airport. Aircraft itineraries are obtained from the FAA’s Aviation System Performance Metrics (ASPM) Individual Flights database [19]. The database includes all domestic and international flights operated by US and international carriers, as well as cargo flights. For every flight, ASPM reports the aircraft tail number, the aircraft type, the airline, the origin and destination airports and the scheduled departure and scheduled arrival times, all of which are used as inputs to the AND model.

2. **Full daily demand schedules** at each airport: In AND, the day is subdivided into 96 periods of equal length (15 minutes) and the expected number of scheduled demands (arrivals and departures) in each period is indicated for each of the 34 airports in the US implementation of AND. This input is obtained from the aforementioned aircraft itineraries and complemented by an estimate of charter and general aviation flights that are expected in each sub-period at every airport. Estimates for the latter may be obtained reliably from historical records for each airport and hour of the day from the FAA’s ASPM Airport Efficiency database [19].

3. **Expected service rates** at each airport: The “service rates” indicate the throughput capacity of each airport’s runway system in terms of the expected number of arrivals and departures that can be served per sub-period. The Queuing Engine of AND requires as an input this throughput capacity, i.e., the
expected value of the airport throughput under continuous demand conditions. Continuous demand conditions guarantee that the rate at which customers exit the queue is equal to the expected service rate of the system. For this reason we obtained the expected capacities of each of the airports in the model as described in Simaiakis and Pyrgiotis [49], or estimated by Morisset [37] and Donaldson [13]. In Section 3.2 we present our work on the estimation of the expected service rate for two of the airports in AND. For airports that are not included in the analysis of the aforementioned literature we have used the capacity estimates found in the FAA's Airport Capacity Benchmark Report 2004 [36]. In practice, as well as in the AND model, service rates may vary with time across sub-periods, reflecting potential changes in weather, noise-related restrictions during certain hours of the day, changes of runway configurations etc.

4. Slack in the airlines’ schedule: As described earlier, two crucial parameters for the AND Delay Propagation Algorithm are the slack included in the scheduled ground time between two flights of the same aircraft and the slack in the block time of each flight, also known as schedule padding. For the schedule padding inputs, we have used results from extensive analyses performed by El Alj [14] and Skaltsas [50], in which padding is estimated for flights between several pairs of airports and which we briefly describe in Section 3.1.3. In Section 3.1.2 we present our own analysis that led to the estimation of the slack in the ground turnaround times.

3.1.1 Separating arrival and departure queues

Depending on the runway system’s layout, the runway configuration in use, and the operational procedures in use, we may distinguish between two types of airports:

- Airports with entirely independent arrival and departure operations. This would require independent sets of runways, where each set is used either for arrivals or for departures.
• Airports where arrivals and departures are sharing the same runway(s), or where arrivals and departures are utilizing separate, but intersecting runways so that arrivals interact with departures due to ATC separation requirements.

In the first case, two entirely separate queuing systems are utilized by AND to compute delays to arrivals and to departures. For airports in this category, we have represented in AND each airport as consisting of two separate servers, one for arrivals and the other for departures, each with independent demand and service rates. For each of the two servers we run the Queuing Engine separately at every AND iteration. A flight that is scheduled to operate at an airport with these characteristics will arrive at the arrival server of that airport and will be transferred to take off (when ready) from the departure server of the airport. [Refer to Figure 2-2a in Chapter 2.]

In the second case, the arrival throughput of the airport depends on the departure throughput and vice versa, since arrivals and departures share the same resources. Even in this case, however, it is necessary to draw some distinctions between the departure and the arrival queues. The reasoning is as follows: In any given period, the airport can handle a certain total number, say c, of operations (arrivals and departures). Additionally, however the airport cannot handle more than \( c_A \) arrivals and \( c_D \) departures, where \( c_A \leq c \) and \( c_D \leq c \). Modeling the airport as consisting simply of one server with capacity \( c \) (i.e. without also considering the limits \( c_A \) and \( c_D \) on arrivals and departures, respectively), would mean that, when the demand for arrivals \( d_A \), is \( c_A < d_A \leq c \) and the demand for departures \( d_D \) is such that \( d_A + d_D \leq c \), the arrivals would experience minimal delays according to the Queuing Engine, while in reality the arrival delays may be large.

For airports in this category, we therefore use one server for arrivals and one for departures, but with interdependent server capacities. Typically, arrivals are given priority over departures in practice. Hence, in our model, the arrival service rate is a function only of the runway configuration and weather conditions, while the departure service rate is a function of the runway configuration, the weather conditions and the
arrival demand in each period. Hence we can write:

\[ \mu_{\text{arr}}(t) = f(RC(t), MC(t)) \]
\[ \mu_{\text{dep}}(t) = g(RC(t), MC(t), D_{\text{arr}}(t)) \]

where \( RC(t), MC(t) \) and \( D_{\text{arr}}(t) \) are the runway configuration, the meteorological conditions and the demand for arrivals, respectively, as a function of the time of the day. We provide more details on the topic of the estimation of the departure service rate as a function of arrivals in Section 3.2.3.

### 3.1.2 Estimation of Aircraft Turnaround Times

During the flight scheduling process, airlines make sure that their aircraft are scheduled to spend more time on the ground between flights than the minimum time required to turn their aircraft around and get them ready for their next flight. Typically, the turn time includes provision for operations like passenger disembarkation, cabin cleaning, refueling, pre-flight inspection of the aircraft, passenger boarding, luggage and cargo loading, and programming of the Flight Management System. Many of these operations happen in parallel and define the earliest time that an aircraft is ready to depart for the next flight. Typically, the turn time increases with the size of the aircraft, due to the longer amount of time required to carry-out some these pre-flight operations for large aircraft.

The purpose of much of the additional time assigned to in the scheduled turn times is to account for various scheduling requirements, such as providing sufficient time for passenger and crew connections and for flight banking operations, especially at hubs, as described by Belobaba et al. [6]. However, in order to produce a more robust schedule, airlines also add some slack to the schedule of an aircraft in order to absorb part of, or, the entire delay of earlier flights. This practice mitigates delay propagation between flights and increases the flexibility of recovery actions, when an airline’s schedule is disrupted by congestion. For these reasons, both the scheduled and the actual turn times would be expected to be higher at hub airports for “hubbing”
Low Cost Carriers (LCCs) are generally more efficient in turning aircraft around but also operate tighter schedules with less slack in order to increase aircraft utilization. Gittell [22], for example, estimated that Southwest Airlines (the largest and oldest LCC in the United States) is able to turn its aircraft around in 20-30 minutes on average. Belobaba et al. [6] point out that LCCs fly, on average, shorter flights than Network Legacy Carriers (NLCs), which means that LCCs fly more legs per day, hence, on average, aircraft of LCCs are more times on the ground per day than aircraft of NLCs. At the same time LCCs utilize their aircraft, on average, almost 2 hours more per day than NLCs. Then we may assume that NLCs, on average, spend longer periods of time on the ground between flights.

As described in section 2.1.3, the time in which airlines can turn around an aircraft on the ground and prepare it for the next flight is a critical parameter of AND. The time when a flight, \( f \), is ready to depart depends on the arrival delay of the predecessor flight, \( f' \), of the same aircraft and the time scheduled to be spent on the ground between the departure of \( f \) and the arrival of \( f' \). The expression that propagates the delay of an arriving aircraft to the subsequent departure of the aircraft utilizes an estimate of slack in ground turnaround times, which is calculated from the following equation:

\[
slack(f'; f) = turn(f', f) - minturn(f', f) \tag{3.1}
\]

It is thus clear that the correct calibration of the minimum turnaround time in the AND model is important. Of great interest, as well, is an analysis of the sensitivity of the delay estimates of AND to the minimum turnaround time. For this purpose, we present in this section a statistical estimation of the minimum turnaround times in today’s system. In order to perform this work, we have used the Individual Flights database of ASPM for 2007 [19]. The time spent on the ground between two consecutive flights of the same aircraft is referred to as an “aircraft connection”. In our analysis we classify aircraft connections according to a combination of attributes that include the operating airline, the aircraft type and the airport type where the
connection takes place. We define two airport types for each airline: hubs and non-
hubs.

Since our goal is to provide estimates of the minimum turnaround times, we have 
filtered the available data on aircraft connections and included in our analysis only 
the aircraft connections that satisfy the following two criteria:

1. The observed turnaround time between two consecutive flights of the same 
aircraft is less than 150 minutes; in this way we exclude overnight connections, 
maintenance checks, or airline banks at a hub airport which are spaced far apart in time.

2. The aircraft arrived with a delay of more than 30 minutes and departed late 
for its next flight, but the departure delay is smaller than the arrival delay. In 
other words:

\[
(\text{actual gate-in time} - \text{schedule gate-in time}) \geq 30 \text{ minutes} \geq \\
(\text{actual gate-out time} - \text{scheduled gate-out time}) > 0
\]

The rationale behind the second criterion is that, for the estimating the minimum turn time, we only consider connections that have tried to utilize any slack in their scheduled ground time to compensate for some portion of the arrival delay. We chose 30 minutes as the threshold of the arrival delay in order to increase the probability that the airline tried to take advantage of the slack in the scheduled turn time to compensate for the delay. At the same time, 30 minutes is a value sufficiently small to permit obtaining enough data points to perform a statistically significant analysis.

The box-plots in Figure 3-3 show the observed turn times achieved by two airlines—
a LCC, Southwest Airlines (WN), and a NLC, Continental Airlines (CO)—for their 
B-737 family aircraft, at their respective hubs and non-hubs. The red line indicates 
the median for each sample, the edges of the box are the 25th and 75th percentiles 
and the whiskers extend to the most extreme data points not considered outliers. The
takeaways from Figure 3-3 are the following:

- **WN** manages to turn around all the types of B-737 aircraft it operates much faster than **CO**. For example, WN turns the B737-300 20 minutes faster, on average, than CO at the respective hubs of the two airlines, and 15 minutes faster at all the other airports. The variability of the turn times is also much smaller for WN than CO for all B737 aircraft.

- **CO** turns its aircraft around faster outside its hubs. For example CO turns the B737-700, on average, in 50 minutes at its main hubs in Newark and Houston and in 39 minutes at the other airports. An intuitive explanation for this observation is that many departing flights at the hubs have to wait for passenger and crew connections, which is not the case at other airports.

- The variability of turn times is greater at the hubs of CO than at the non-hub airports.

- In contrast to CO, WN turns its aircraft around faster at its hub at Chicago Midway. This might be explained by the fact that WN does not operate a hub-and-spoke network like CO, so that even at MDW there are few passenger connections. Thus, it is plausible to assume that WN operations at MDW are more efficient, as they can be customized to the airline's needs, due to their scale, turning the aircraft around faster.

To determine whether the turn times of different aircraft types that belong to the same aircraft family could be samples from the same probability distribution, we performed the non-parametric Kruskal-Wallis test. The Kruskal-Wallis test assumes no particular distribution form and tests the null hypothesis that the probability distributions underlying different sample groups are identical (details in [44]). The test results lead to a rejection of the null hypothesis, both for CO's and WN's fleets of aircraft belonging to the B737 family. We therefore conclude that the turn times for the B737-300, the B737-500 and the B737-700 have different probability distributions.
In a similar analysis, we compare the turn times of United Airlines, UA, and JetBlue, B6, A320 aircraft. The box plots of Figure 3-4 show the distribution of turn times at the hubs and at the non-hub airports of the two airlines. Surprisingly, we observe that, in this case the Network carrier, UA, turns the A320 faster than the Low Cost carrier, B6, both at their respective hubs, and at the other airports. As was
the case with CO, we observe again that airlines tend to turn their aircraft around faster outside their hubs and with smaller variability. This observation was made for all airlines in our analysis, with the exception of WN, as shown in the estimates presented in Appendix A.

![Box-plots of the turn time of the A320 aircraft for UA and B6 at their hubs and at all other airports.](image)

Figure 3-4: Box-plots of the turn time of the A320 aircraft for UA and B6 at their hubs and at all other airports.

Furthermore, Figure 3-5 presents the histograms of both the scheduled turn times and the actual turn times for the A320 aircraft of UA and B6. The dashed lines show the average values for each of the histograms. Observing these figures we may conclude that B6 operates on a slightly tighter schedule than UA, i.e. with less slack in the schedule turnaround times, as also suggested in the literature. UA on average actually turns the A320 in 38 minutes, while the average scheduled turn time for the same flights is 58 minutes, resulting in 20 minutes of estimated slack. In contrast, B6 actually turns the A320 in 43 minutes on average, while the average scheduled turn time is 61 minutes, thus giving an estimated slack of 18 minutes. One more thing to observe is that the distribution of turn times is approximately normal. This was further demonstrated by performing a Pearson's goodness-of-fit test (details in [44]) on the data, which showed that the data samples fit a normal distribution with 90% confidence.
Figure 3-5: Histograms of the actual and the scheduled turn times of the A320 aircraft for UA and B6 outside their hubs.

In order to understand the effect of aircraft size, Figure 3-6 depicts the box-plots of six aircraft types in DL’s fleet. Clearly, as the size of the aircraft increases, the turn time increases as well. On average, DL turns the MD-88 in 36 minutes, the B737-800 in 43 minutes, the B757-200 in 47 minutes, the B767-300 in 55 minutes, the B767-400 in 63 minutes and the B777-200 in 60 minutes.

Figure 3-6: Box-plots of the actual turn times of 6 aircraft types in DL’s fleet.

Finally, it is important to perform a more rigorous test of the significance of the difference between the average values of different data sets, especially of the ones
whose averages lie close to each other. Hence, we performed a cluster of hypothesis tests based on the assumption that observations are independent and normally distributed. We could then use simple t-tests to obtain the confidence interval for the difference of two sample averages \( \bar{X} \) and \( \bar{Y} \) as follows:

\[
t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}
\]

where, \( s_p \) is the weighted average of the sample variances of the \( X \)'s and the \( Y \)'s and \( n, m \) are the sample size of \( X \) and \( Y \) respectively.

We tested the null hypothesis that there is no systematic difference between the distributions of the \( X \)'s and the \( Y \)'s, \( H_0 : \mu_X = \mu_Y \), against the two-sided alternative that the true mean of \( X \) is different from the true mean of \( Y \), \( H_1 : \mu_X \neq \mu_Y \). In our analysis we distinguish among three types of pairs of samples:

1. \( X \) are the turn time observations for a specific aircraft type of an airline at its hub and \( Y \) for the same aircraft type at every other airport,

2. \( X \) are the turn time observations for a specific aircraft type of an airline at its hub and \( Y \) for the same aircraft type of another airline at its hub.

3. \( X \) are the turn time observations for a specific aircraft type of an airline at non-hub airports and \( Y \) for the same aircraft type of another airline at non-hub airports.

Table 3.1 presents the results of a few of the comparisons that we have performed. The critical value of \( t \) above which we reject the null hypothesis—and assume that the samples have different means—with 99% confidence is \( t_{\infty}(0.01) = 2.576 \). It may be observed in Table 3.1 that with 99% confidence we can accept that there is no systematic difference between the mean turn time of DL and CO at their hubs, either for the B737-800 or for the B757-200. Moreover, we may assume that, on average, F9, a LCC, and CO, a NLC, turn around their B737 aircraft in the same amount of time at their respective hubs (F9 at Atlanta and CO at Newark and Houston), while
UA turns its B767-300 aircraft on average in the same time at its hub and at all other airports.

Table 3.1: Turn time statistics.

<table>
<thead>
<tr>
<th>A/C Type</th>
<th>Airprot</th>
<th>Airline</th>
<th>Average</th>
<th>St. Dev.</th>
<th># samples</th>
<th>t-test results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A320</td>
<td>Non-hub</td>
<td>B6</td>
<td>42.8</td>
<td>12.7</td>
<td>6518</td>
<td>19.343, reject H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UA</td>
<td>38.4</td>
<td>12.7</td>
<td>6087</td>
<td></td>
</tr>
<tr>
<td>B738</td>
<td>Hub</td>
<td>DL</td>
<td>59.3</td>
<td>21.8</td>
<td>502</td>
<td>0.746, accept H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO</td>
<td>58.6</td>
<td>17.4</td>
<td>2615</td>
<td></td>
</tr>
<tr>
<td>B738</td>
<td>Non-hub</td>
<td>DL</td>
<td>43.0</td>
<td>14.0</td>
<td>1621</td>
<td>3.606, reject H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO</td>
<td>44.4</td>
<td>11.3</td>
<td>3394</td>
<td></td>
</tr>
<tr>
<td>B752</td>
<td>Hub</td>
<td>DL</td>
<td>62.0</td>
<td>21.0</td>
<td>2981</td>
<td>2.434, accept H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO</td>
<td>65.2</td>
<td>17.6</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td>B752</td>
<td>Non-hub</td>
<td>DL</td>
<td>46.6</td>
<td>14.8</td>
<td>3046</td>
<td>6.537, reject H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO</td>
<td>50.0</td>
<td>8.6</td>
<td>870</td>
<td></td>
</tr>
<tr>
<td>B737</td>
<td>Hub</td>
<td>F9</td>
<td>51.3</td>
<td>24.3</td>
<td>1498</td>
<td>1.960, accept H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO</td>
<td>49.5</td>
<td>17.6</td>
<td>961</td>
<td></td>
</tr>
<tr>
<td>B763</td>
<td>Hub</td>
<td>UA</td>
<td>64.0</td>
<td>24.5</td>
<td>269</td>
<td>0.917, accept H&lt;sub&gt;0&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>Non-hub</td>
<td>UA</td>
<td>62.6</td>
<td>19.6</td>
<td>523</td>
<td></td>
</tr>
</tbody>
</table>

From this analysis we obtain the minimum turn times per airline per aircraft type and per type of airport (hub or non-hub) that are used in AND. More specifically, we use in AND the mean of each of the sample groups as the value of the minimum turnaround time. For the cases where there is no systematic difference between the sample means, we use the pooled mean of the samples. In Appendix A we present a snapshot of the minimum turn times that are used by AND to estimate the delay propagation between flights.

**Sensitivity analysis**

In order to perform a sensitivity analysis of the effect of the minimum turnaround time on the results of AND, we have run the model for a set of different airport capacities and different minimum turnaround times. The tests were the following:

1. All airports are operating under optimum conditions for a full day of operations and the minimum turnaround times used are the ones obtained from the analysis described in the previous section.
2. All airports are operating under optimum conditions for a full day of operations, but the minimum turnaround times for every airline and aircraft type are reduced from the mean, \( \mu \), of the distribution to \( \mu - \sigma/2 \). For example, the minimum turnaround time of DLs B737-800 in Atlanta, which was originally set to 59 minutes, is now reduced to \( 59 - 11 = 48 \) minutes.

3. Same as Test 1, but with a high level of congestion at Atlanta caused by low IFR conditions (i.e., with a capacity drop of 15% at ATL).

4. Same as Test 2, but with a high level of congestion at Atlanta caused by low IFR conditions there.

We shall observe the response of the following statistics to the reduction of the minimum turn time:

- Network upstream delay. (Upstream delay is defined as the portion of the total arrival delay experienced by a flight, which can be attributed to earlier delays in the airport network and not to the local delay at the airport of arrival.)

- Total departure delay at ATL

- Total departure delay for two hub airlines at ATL:
  - Delta (DL), a NLC
  - AirTran (F9), a LCC

In Figure 3-7 we show the sensitivity of the aforementioned statistics to the reduction of the minimum turnaround time under both low and high congestion at ATL. In Table 3.2 one can see the effect of increasing the ground slack on the average delays of different groups in the system. By reducing the minimum turn time by \( \sigma/2 \) we are essentially decreasing, throughout the network, the minimum turn times by approximately 20% on average. This results in a 9% reduction in the total network upstream delay in the system under Test 1, and to a 13% reduction when ATL operates in low IFR. Clearly, the higher the congestion in the system, the higher the delay savings.
are when the minimum turn time is reduced, since any available slack is fully utilized only when there is enough delay to compensate for.

It can also be seen that F9 is more sensitive than DL to the reduction in minimum turnaround times. This can be explained by the fact that F9 operates on a much tighter schedule than DL. In particular, the average scheduled turnaround time for the B737 fleet of F9 at ATL was 51 minutes in 2007, leaving on average only 3 minutes of slack, while for the B737-800 of DL it was 71 minutes, leaving 10 minutes of slack.

![Figure 3-7: Sensitivity analysis results.](image)

3.1.3 Padding of Scheduled Block Times

In 1987 the U.S. Department of Transportation implemented the On-Time Disclosure Rule. Under this rule, all US carriers performing scheduled commercial service
Table 3.2: Sensitivity analysis results.

<table>
<thead>
<tr>
<th></th>
<th>Test 1</th>
<th>Test 2</th>
<th>% diff.</th>
<th>Test 3</th>
<th>Test 4</th>
<th>% diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of arrivals delayed &gt; 15mins</td>
<td>3.3%</td>
<td>3.2%</td>
<td>-0.1</td>
<td>4.2%</td>
<td>4.0%</td>
<td>-0.2</td>
</tr>
<tr>
<td>Avg departure delay at ATL</td>
<td>2.7</td>
<td>2.2</td>
<td>-18%</td>
<td>8.2</td>
<td>7.2</td>
<td>-1.2%</td>
</tr>
<tr>
<td>Avg departure delay of DL</td>
<td>7.9</td>
<td>7.0</td>
<td>-12%</td>
<td>14.1</td>
<td>11.9</td>
<td>-15%</td>
</tr>
<tr>
<td>Avg departure delay of F9</td>
<td>4.7</td>
<td>3.7</td>
<td>-25%</td>
<td>11.4</td>
<td>8.3</td>
<td>-27%</td>
</tr>
</tbody>
</table>

are required to submit monthly Airline Service Quality Performance Reports to the Bureau of Transportation Statistics (BTS). BTS releases publicly available monthly reports with performance statistics for each carrier. The most widely used performance metric is the on-time arrival rate, according to which a flight is considered to be “on time” if it arrives no more than 15 minutes later than its scheduled arrival time. The main goal of publishing the statistic is to incentivize carriers to enhance the efficiency of their operations and reduce delays.

As a response to the On-Time Disclosure Rule, US carriers started increasing the scheduled travel time from gate departure to gate arrival of their flights (scheduled block time) in order to improve their published on-time performance. Over time, as increasing congestion led to higher delays, airlines also increased their scheduled block times to take into account the delay that each flight was expected to experience. As in the case of ground turnaround times, a flight’s scheduled block time includes by now a significant amount of slack to compensate for any delay the flight is expected to sustain. This slack is also referred to as the padding of the scheduled block time. The amount of schedule padding depends on several factors, such as the itinerary of the flight and the time of the day when the flight is scheduled to operate, since different airports face different levels of congestion that vary by time of day.

El Alj in 2003 [14] developed a new metric to estimate delays due to congestion at airports and in the airspace and to other system inefficiencies. This metric was defined as the difference between the actual block time as measured from air traffic statistics and a baseline time for each origin-destination pair. The baseline time is route-specific and is measured as the 15th percentile of the actual block time distribution for any given month for each route. Note that the baseline time is essentially an estimate...
of the delay-free travel time between a pair of airports, i.e., of the true travel time without any “padding”. El Alj estimated the baseline time for 618 markets and found that the true arrival delays, based on her metric, were about 40% to 60% greater than the reported arrival delays relative to schedule in 2000. In other words, according to her estimates, block time padding “hid” about 40-60% of the true delays suffered by flights in her 618 markets in 2000.

Skaltsas in 2011 [50] took a similar approach and estimated the schedule padding in more than 2000 routes in the US in 2009. He assumed that the nominal airborne time of a flight in a specific route corresponds to the 10th percentile of the observed data and obtained estimates of the taxi-in and taxi-out times at every airport from the ASPM database [19]. He was thus able to estimate the nominal block time for each of the 2000 markets as the sum of the unimpeded travel times associated with each of the components of a flight: taxi-in, airborne and taxi-out. In our calibration of AND we used the results of Skaltsas to obtain values for nominal flight times and schedule padding. In Appendix B we include a snapshot of the data obtained from Skaltsas for a representative subset of the 2000 routes.
3.2 Validation of the $M(t)/E_k(t)/1$ queuing system at the airport departure process

In this section we begin our validation efforts by focusing on individual airports at which we apply the approximate $M(t)/E_k(t)/1$ queuing model, DELAYS, which is used as the Queuing Engine of AND. We examine the accuracy of the estimates of delays generated by this model by comparing the delay estimates it generates with field data. Specifically, we use the Queuing Engine as described in Section 2.1.2, to approximate the departure process at two of the busiest airports in the United States and compare the results of the model to data obtained from field observations. The research presented here is joint work with Ioannis Simaiakis [49] and complements his earlier work on models of the departure process at congested airports; for details refer to [47, 46].

Operations on the airport’s surface include those at the gate areas/aprons, the taxiway system and the runway systems, and are strongly influenced by terminal-airspace operations. Taxi-out time is defined as the time between the actual pushback time of an aircraft and its takeoff time, i.e., it is the amount of time that the aircraft spends on the airport surface with engines on, and includes the time spent on the taxiway system and in the runway queues.

We begin our analysis by developing a queuing model of the departure process, in order to describe quantitatively how queues form on the airport’s surface. Then, we validate this model in terms of its ability to predict taxi-out times at two airports with very different runway characteristics: Boston Logan International Airport (BOS) and Newark Liberty International Airport (EWR).

3.2.1 Model structure

The departure process at an airport is depicted in Figure 3-8 and can be conceptually described as follows [49]: Aircraft push back from their gates according to a (possibly modified) departures schedule. They enter the ramp and then the taxiway system
on which they travel to the departure queue that is formed at the threshold of the departure runway(s). During this traveling phase, aircraft interact with each other. For example, aircraft queue to cross an active runway, or to enter a taxiway segment on which another aircraft is taxiing. We refer to these cumulative spatially distributed interactions and the delays which occur while aircraft traverse the airport surface from their gates towards the departure queue as ramp and taxiway interactions. After traveling on the taxiway system the aircraft line up awaiting takeoff, thus forming a departure queue at the runway threshold. We model the runway service process as a server, with the departure runways processing the departing aircraft in a First-Come-First-Served (FCFS) manner.

By modeling the departure process in the manner described above, the taxi-out time $\tau$ of each departing aircraft can be expressed as

$$\tau = \tau_{\text{unimped}} + \tau_{\text{taxiway}} + \tau_{\text{dep.queue}}$$  \hspace{1cm} (3.2)

The first term of Equation (3.2), $\tau_{\text{unimped}}$, reflects the unimpeded taxi-out time of the flight. This is the time that the aircraft would spend on the taxiway system, up to the beginning of its takeoff roll, if it were the only aircraft on the ground. The second term, $\tau_{\text{taxiway}}$, reflects the delay due to aircraft interactions on the ramp and the taxiways. The magnitude of this delay depends on the level of congestion on the taxiways. The third term, $\tau_{\text{dep.queue}}$, is the time the aircraft spends in the departure queue. The duration of this time depends on the number of aircraft in the departure queue and the runway service characteristics.

The inputs to the model are:
• The pushback schedule, \( PS \).

• The gate location of the departing flight, \( GL \).

• The “segment” in use, \((MC; RC)\), expressed as a combination of the general meteorological conditions, \( MC \), and the runway configuration, \( RC \).

The desired outputs of the model include:

• The level of congestion on the airport surface.

• Statistics about taxi-out times.

• The predicted taxi-out time of each departing flight.

The following two sections focus on the estimation of the three terms of Equation 3.2, given a set of the explanatory variables \((RC, MC, GL, PS)\). In Section 3.2.2 we briefly describe how Simaiakis estimates the first two terms of Equation 3.2 \((\tau_{\text{unimped}} \text{ and } \tau_{\text{taxiway}})\). In Section 3.2.3 we provide a detailed discussion on how to estimate the third term of the Equation \((\tau_{\text{dep.queue}})\), which is obtained from the DELAYS Queuing Engine of AND.

### 3.2.2 Unimpeded taxi-out times and Ramp and taxiway interactions

The travel time of an aircraft from its gate to the runway queue depends on its unimpeded taxi-out time and on the amount of traffic on the ramps and the taxiway system at the time:

\[
\tau_{\text{travel}} = \tau_{\text{unimped}} + \tau_{\text{taxiway}}
\]  

(3.3)

Simaiakis [47] describes a method to estimate the unimpeded taxi-out times of each airline on each airport segment as a normal random variable \( U \) with mean value \( \bar{\tau}_{\text{unimped}} \) and variance \( s^2_{\text{unimped}} \):

\[
U \sim N(\bar{\tau}_{\text{unimped}}, s^2_{\text{unimped}})
\]  

(3.4)
(Note that the identity of the airline provides information about the approximate location of the gate from which each flight commences.)

As also shown by Simaiakis [47], the assumption that the unimpeded taxi time is equal to its estimated mean value does not compromise the validity of this model. Thus, we assume that the unimpeded taxi time is simply equal to its expected value:

\[ \tau_{\text{unimped}} = \bar{\tau}_{\text{unimped}} \] (3.5)

Furthermore, to estimate \( \tau_{\text{taxiway}} \) we use the same approach as presented in Simaiakis [47], in which a linear relationship is assumed between \( \tau_{\text{taxiway}} \) and the number of departing aircraft on the ramp and the taxiways that have not yet reached the departure queue. The constant of proportionality depends on the airport and the runway configuration.

### 3.2.3 The runway service process

In order to estimate the last component of Equation 3.2, \( \tau_{\text{dep.queue}} \), we use the queuing model described in Section 2.1.2 and more specifically in Equations 2.2-2.5. We model the runway service process as an \( M(t)/E_k(t)/1 \) queueing system with infinite queue capacity. The inputs to this model are the departure demand rate, \( \lambda(t) \), which is defined as the expected number of aircraft per unit of time that appear at the threshold of each departure runway ready to take-off, as well as the runway server characteristics, \( \mu(t) \) and \( k \), defined as the mean number of departing aircraft that can take-off per unit of time under continuous demand conditions and the Erlang order of the service time distribution, respectively. The output of the model is \( W_q(t) \), as estimated through Equation 2.5, which is defined as the expected waiting time in the departure queue. Hence, we may re-write Equation 3.2 for the estimation of the total taxi-out time as:

\[ \tau = \tau_{\text{travel}} + W_q(t) \] (3.6)
The expected departure demand rate, $\lambda(t)$, is calculated using the estimates of $\tau_{travel}$ for each flight. It is assumed that the expected time when an aircraft will arrive at the departure queue is at time $\tau_{travel}$ after the pushback event. Then, given a pushback time from the gate for each flight (the pushback schedule), we may use the estimates for $\tau_{travel}$ for each departing aircraft to obtain the arrival time of each flight at the departure queue. Thus, $\lambda(t)$ is determined by counting the flights that are estimated to arrive at the departure queue during a period $[t, t + \Delta t]$, where $\Delta t$ is the unit of time used in the approximation of the queuing system.

In order to estimate the runway service process characteristics, $\mu(t)$ and $k$, we perform the following two-step analysis:

**Identification of throughput saturation points**

As a first step, we observe the inter-departure times during heavy loading. Under such conditions the runways operate at their capacity. By observing the output of the departure process (i.e., the inter-departure times), the statistical properties of the server (the runways) may be inferred. However, the regimes in which the runway process is saturated and the runway operates at capacity, first need to be identified.

Following the approach of Simaiakis [47], we use the number of departing aircraft on the ground, $N(t)$, as an indicator of the loading of the departure runway. As $N(t)$ increases, the takeoff rate initially increases, but reaches a saturation limit at a critical value $N^*$. Beyond $N^*$, the runway becomes the defining capacity constraint. Increasing further the number of departing aircraft on the ground does not increase the throughput of the airport.

Figure 3-9 shows the average takeoff rate as a function of $N(t)$ for the segment (VMC; 22L | 22R) at EWR for 2007. (This notation means that the airport operates under Visual Meteorological Conditions with runway 22L used for arrivals and 22R for departures.) The error bars depict the standard deviation of the takeoff rate. The saturation point is also denoted. We note that the takeoff rate initially increases as $N(t)$ increases, but saturates at about 0.67 departures/min or 40 departures/hour.

However, due to the sharing nature of resources (i.e. of runways and taxiways),
by both arriving and departing aircraft, there is a tradeoff between the arrival and departure capacity of airports like EWR and BOS (as described in Section 3.1.1). We therefore assume that the departure service rate is also a function of the arrival demand. Hence, in order to accurately model the departure process, we need to estimate the departure service rate under different levels of arrival traffic. For simplicity, we make use of the observation that, at the two airports in our analysis, BOS and EWR, arriving traffic is higher in the afternoons than in the mornings [8] and thus we calculate the expected service time and its standard deviation for only two sub-periods, before noon and in the afternoon. A finer subdivision of time can be used, if desired.

The departure service rate during the course of the day can be visualized by plotting the takeoff rate as a function of $N(t)$ in Figure 3-10, and by separating morning and afternoon traffic. In Figure 3-10 it may be observed that, the departure throughput rate in saturation is higher before noon than in the afternoon at EWR. Specifically, high arrival traffic in the afternoon can reduce the departure service rate by as much as 6 departures per hour.

Figure 3-9: Takeoff rate as function of $N(t)$ [47]
The derivation of the runway service time distribution

Having identified the regime of operations when the runway loading is high, it is now possible to model the runway service process itself. First, we observe the takeoff rate when \( N(t) \) is larger than \( N^* \) and draw a histogram of inter-departure times under this condition. Such a histogram is shown in Figure 3-11 where we plot the inter-departure times of aircraft at EWR under configuration 22L | 22R in the morning hours during high loads \( (N(t) > N^* = 27) \). Given that in the presence of saturation there are always aircraft in the departure queue and the server is fully loaded, we may assume that the inter-departure time histogram approximates accurately the service rate histogram of the runway server. Hence the mean and standard deviation of the inter-departure histogram are essentially estimates of the mean service time \( (E[S]) \) and the standard deviation \( (\sigma_{st}) \) of the runway service times, respectively. From here on, we shall refer to the inter-departure histogram (distribution) as the service time histogram (distribution). Furthermore, from \( E[S] \) we can calculate the departures service rate as \( \mu = \frac{1}{E[S]} \times 60(AC/hour) \).

The estimation of the throughput in saturation as a function of both the departing
Figure 3-11: Histogram of the inter-departure times in saturation and Gamma and Erlang distribution fits

traffic on the ground and of the time of day yields the time-dependent estimates of the departure time characteristics listed in Table 3.3.

Table 3.3: Runway service time characteristics for two frequently-used runway configurations at BOS and at EWR.

<table>
<thead>
<tr>
<th>Airport, Configuration</th>
<th>Time</th>
<th>$E[S]$ (min)</th>
<th>$\sigma_{st}$ (min)</th>
<th>$\mu$ (AC/hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOS, (4L, 4R</td>
<td>4L, 4R, 9)</td>
<td>before noon</td>
<td>1.1654</td>
<td>0.8007</td>
</tr>
<tr>
<td>BOS, (4L, 4R</td>
<td>4L, 4R, 9)</td>
<td>after noon</td>
<td>1.3042</td>
<td>0.8695</td>
</tr>
<tr>
<td>EWR, (22L</td>
<td>22R)</td>
<td>before noon</td>
<td>1.3392</td>
<td>0.7324</td>
</tr>
<tr>
<td>EWR, (22L</td>
<td>22R)</td>
<td>after noon</td>
<td>1.5547</td>
<td>1.0522</td>
</tr>
</tbody>
</table>

One of the assumptions of the runway service time model is that the service times follow a time-dependent $k^{th}$ - order Erlang distribution, as implied by the adoption of a $M(t)/E_k(t)/1$ queuing model. As a first step to estimating $k$, we use the Method of Moments to fit a gamma distribution to the observed service times histogram. The first two moments of the gamma distribution with shape parameter $k_{\text{gamma}}$ and scale parameter $\theta_{\text{gamma}}$ yield estimates $\hat{k}_{\text{gamma}}$ and scale parameter $\hat{\theta}_{\text{gamma}}$:

$$\hat{k}_{\text{gamma}} = \frac{(E[S])^2}{\sigma_{st}^2}$$  \hspace{1cm} (3.7)

$$\hat{\theta}_{\text{gamma}} = \frac{\sigma_{st}^2}{E[S]}$$  \hspace{1cm} (3.8)
As can be observed in Figure 3-11, the gamma distribution, with estimated parameters from Equations 3.7 and 3.8, fits the actual service time distribution reasonably well. As a next step, we constrain \( \hat{k} \) to be an integer, in order to transform the Gamma distribution to an Erlang distribution. We seek to find the Erlang distribution that has the same mean as the observed service time distribution and the Erlang order \( (k) \) that will result in a standard deviation as close as possible to the observed one. Hence we may write that:

\[
\hat{k}_{\text{erlang}} = \lfloor \hat{k}_{\gamma} + 0.5 \rfloor = \lfloor \frac{(E[S])^2}{\sigma_{st}^2} + 0.5 \rfloor \tag{3.9}
\]

\[
\hat{\theta}_{\text{erlang}} = \frac{E[S]}{\hat{k}_{\text{erlang}}} = \frac{\lfloor \frac{(E[S])^2}{\sigma_{st}^2} + 0.5 \rfloor}{E[S]} \tag{3.10}
\]

The resulting Erlang distribution will have a mean \( E[S] \) and variance \( \sigma_{\text{erlang}}^2 \):

\[
\sigma_{\text{erlang}}^2 = \frac{\hat{k}_{\text{erlang}}}{\hat{\theta}_{\text{erlang}}} = \frac{(E[S])^2}{\lfloor \frac{(E[S])^2}{\sigma_{st}^2} + 0.5 \rfloor} \approx \sigma_{st}^2 \tag{3.11}
\]

Figure 3-11 also shows the Erlang distribution fit that results from applying Equations 3.9 and 3.10 to the service time distribution in segment (VMC ; 22L | 22R), before noon.

In Table 3.4, we list the service rates \( \mu(t) \) (in \( \text{AC/min} \)) and Erlang orders \( (k) \) that result from applying Equations 3.9 and 3.10 to the four service time distributions, which we examine here. We also list the corresponding \( \sigma_{\text{erlang}} \), as calculated by Equation 3.11. Comparing \( \sigma_{\text{erlang}} \) to \( \sigma_{st} \), listed in Table 3.3, it can be observed that the standard deviation of the best-fit Erlang distribution is slightly greater than the observed one.

### 3.2.4 Predictions at EWR and BOS

We ran the model for all periods in 2007 during which each of the segments of BOS and EWR, presented in Table 3.4, was realized. We computed the taxi out time of all flights that were observed to pushback and take-off during these periods. By
Table 3.4: Time-dependent Erlang distribution for the runway service time for two frequently-used runway configurations at BOS and at EWR

<table>
<thead>
<tr>
<th>Airport, Configuration</th>
<th>Time</th>
<th>(\mu(t)) (AC/min)</th>
<th>(k)</th>
<th>(\sigma_{\text{erlang}}) (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOS, (4L, 4R</td>
<td>4L, 4R, 9)</td>
<td>before noon</td>
<td>0.8581</td>
<td>2</td>
</tr>
<tr>
<td>BOS, (4L, 4R</td>
<td>4L, 4R, 9)</td>
<td>after noon</td>
<td>0.7667</td>
<td>2</td>
</tr>
<tr>
<td>EWR, (22L</td>
<td>22R)</td>
<td>before noon</td>
<td>0.7467</td>
<td>3</td>
</tr>
<tr>
<td>EWR, (22L</td>
<td>22R)</td>
<td>after noon</td>
<td>0.6432</td>
<td>2</td>
</tr>
</tbody>
</table>

solving Equation 3.6 iteratively as the day progresses, using the pushback schedule as an input, we generated estimates of the individual taxi times \(\tau\) of all flights, and, subsequently, also obtained estimates of the congestion state of the airport, \(N(t)\), and of the length of the departure queue, \(Q(t)\). Since all the terms of the model \((\tau_{\text{travel}} \text{ and } W_q(t))\) are determined analytically, there is no need to run a Monte Carlo simulation.

**Airport Throughput Prediction**

Figure 3-12 shows the predicted throughput at EWR and BOS as a function of the number of departing aircraft on the ground. As can be observed, the model predicts very accurately the throughput and its standard deviation (the error bars) for all traffic conditions up to \(N(t) = 18\) for BOS and \(N(t) = 25\) for EWR. For higher values of \(N(t)\) the model predicts correctly the average takeoff rate, but with a different variance than the observed one.

**Congestion Prediction**

Figure 3-13 shows the frequency of different congestion states, as observed in practice (blue line) and as estimated by the model (red bars), for BOS and EWR. The x-axis represents the number of aircraft on the airport's surface, while the y-axis shows the number of periods in 2007 (15-minute intervals) during which each number was observed. Clearly, the model predicts the ground traffic at both BOS and EWR very accurately.
Figure 3-12: Actual and modeled takeoff rate as a function of $N(t)$

**Taxi times Prediction**

Tables 3.5 and 3.6 list the mean actual taxi-out time and the mean estimated taxi-out time for the BOS segment (VMC; 4L, 4R | 4L, 4R, 9) and the EWR segment (VMC; 22L | 22R) in 2007, respectively. They also contain actual and estimated values of more detailed statistics about the number of aircraft and the taxi times at different congestion levels. These statistics were obtained by running the model just a single time. Clearly, the model predicts accurately both the taxi times and the congestion states at each airport. The model’s estimates of average taxi-times over different ranges of $N(t)$ are also consistently close to the observed ones.

In addition to the aggregate comparisons presented so far, it is interesting to observe how the model performs in predicting the taxi times of individual flights. For
this purpose, for each flight operating in the BOS segment (VMC; 4L, 4R | 4L, 4R, 9) and the EWR segment (VMC; 22L | 22R), we compare the individual taxi times as estimated by the model with the observed ones. Table 3.7 shows the error statistics in the prediction of individual flight taxi-out times. We may observe that the model predicts taxi times of individual flights at BOS slightly better than it does at EWR, as the mean absolute percent error is 27% for BOS and 33% for EWR. Furthermore, for BOS 70% of the predictions are within 5 minutes of their real values.

**Daily taxi-out estimations**

In contrast to the results presented so far, which correspond to the model being run for the entire 2007, we present in this section the results of model tests involving only
Table 3.5: Aggregate taxi time predictions for BOS segment (VMC; 4L, 4R | 4L, 4R, 9)

<table>
<thead>
<tr>
<th>Congestion level</th>
<th># of flights modeled</th>
<th>Act. avg. taxi time</th>
<th>Estimated avg. taxi time</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>27,326</td>
<td>18.21</td>
<td>17.93</td>
</tr>
<tr>
<td>(N ≤ 8)</td>
<td>11,876</td>
<td>15.72</td>
<td>15.22</td>
</tr>
<tr>
<td>(9 &lt; N ≤ 16)</td>
<td>13,716</td>
<td>19.12</td>
<td>18.91</td>
</tr>
<tr>
<td>(N &gt; 17)</td>
<td>1,731</td>
<td>25.50</td>
<td>25.06</td>
</tr>
</tbody>
</table>

Table 3.6: Aggregate taxi time predictions for EWR segment (VMC; 22L | 22R)

<table>
<thead>
<tr>
<th>Congestion level</th>
<th># of flights modeled</th>
<th>Act. avg. taxi time</th>
<th>Modeled avg. taxi time</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>49,598</td>
<td>25.62</td>
<td>25.37</td>
</tr>
<tr>
<td>(N ≤ 8)</td>
<td>16,215</td>
<td>17.14</td>
<td>16.95</td>
</tr>
<tr>
<td>(9 &lt; N ≤ 16)</td>
<td>18,281</td>
<td>22.29</td>
<td>22.01</td>
</tr>
<tr>
<td>(N &gt; 17)</td>
<td>15,102</td>
<td>37.95</td>
<td>38.48</td>
</tr>
</tbody>
</table>

a single day of operations compared against observed data. We chose July 22, 2007 to make predictions about the departure process at BOS using our model. Figure 3-14 shows the results using as an input the pushback schedule for a 10-hour period (10:00am-8:00pm). It should be noted again that our estimates are obtained through only a single run of the model, as the model is an analytical one, not a simulation. The top subplot shows the observed and the predicted number of departures in every 15-minute interval, the middle one the average taxi-out times of the flights that depart in the corresponding 15-minute interval, and the one at the bottom the average predicted departure queue size for each 15-minute interval.

We note that the model predictions match the observations very well. This is confirmed by Table 3.8 that presents the statistics concerning the differences between the observed and the model-predicted values.

In summary, we have shown that the DELAYS approximation of the $M(t)/E_k(t)/1$ queuing system can model very accurately the departure process at a busy airport. The correct calibration of the model in terms of its server characteristics, $\mu(t)$ and $k$, is essential and, for this reason, we described in detail the overall methodology and
Table 3.7: Individual taxi time predictions for BOS segment (VMC; 4L, 4R | 4L, 4R, 9) and EWR segment (VMC; 22L | 22R)

<table>
<thead>
<tr>
<th></th>
<th>BOS</th>
<th>EWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute error</td>
<td>5.04 min</td>
<td>8.36 min</td>
</tr>
<tr>
<td>Mean absolute percent error</td>
<td>0.27</td>
<td>0.33</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>8.50 min</td>
<td>13.66 min</td>
</tr>
<tr>
<td>Root mean square percent error</td>
<td>0.36</td>
<td>0.46</td>
</tr>
<tr>
<td>Prediction accuracy within ± 5 minutes</td>
<td>70.0</td>
<td>51.6</td>
</tr>
<tr>
<td>Prediction accuracy within ± 10 minutes</td>
<td>91.7</td>
<td>76.7</td>
</tr>
</tbody>
</table>

Figure 3-14: Prediction of departure throughput, average taxi-out times and departure queue lengths in each 15-min interval over a 10-hour period on July 22, 2007.

an approach for estimating these two parameters. The above analysis provides strong evidence that the DELAYS model is adequate for describing queues within an airport network model, such as AND. To our knowledge, as indicated in Section 1.2.1, the airport modules of other analytical airport network models, such as LMINET [32], have not been validated at a similar level of detail. For example, LMINET assumes an Erlang order of 22 for the probability distribution of the departure service times at an airport, which, according to our findings, is unjustifiable. We believe that AND is the first airport network model with a validated queuing engine for estimating delays at individual airports.
Table 3.8: Evaluation of model predictions for July 22, 2007 at BOS.

<table>
<thead>
<tr>
<th></th>
<th>RMS Error (in minutes)</th>
<th>RMS % Error</th>
<th>Mean Error (in minutes)</th>
<th>Mean % Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxi-out time</td>
<td>2.510</td>
<td>18.8</td>
<td>1.918</td>
<td>13.6</td>
</tr>
</tbody>
</table>

3.3 Validation of AND

In the previous section, it was shown that, the Queuing Engine of AND is capable, after careful calibration, of estimating very accurately delays at individual airports. In this section we extend our validation efforts to the entire AND model in order to investigate the extent to which AND generates realistic delay estimates on a network-wide scale. It should be remembered that, as noted earlier, AND does not capture several important features of an air transportation network. A non-exhaustive list of such missing features includes:

- Airline reactions to congestion. AND does not account for flight cancellations, spare aircraft at hubs and swapping of aircraft assignments to flights during periods of irregular operations.

- Crew schedule restrictions and delayed departures to accommodate passenger connections: these may lead to propagation of delays between flights not operated by the same aircraft.

- En-route congestion.

Ideally, since AND is an analytical model that generates aggregate statistics of delays, the validation of AND should compare the results of the model for a day of operations under a specific set of capacity conditions at the 34 airports with the delay statistics assembled by processing observed data from many days when exactly the same capacity conditions existed in practice. For example, we would like to compare the results of AND when ORD is operating in IFR weather for an entire day and all other airports operate under optimum conditions, with field data from all days when these same conditions were observed. However, it is almost impossible to find
even two days in a single year when the exact same capacity conditions existed at all the US airports. We were therefore forced to validate the AND results against field data from single days of operations in the US air transportation network. For this purpose, we chose one day in 2007 when the three New York airports (EWR, JFK, LGA), Minneapolis/St. Paul (MSP) and Philadelphia (PHL) were operating under IFR conditions (05/10/2007) for most of the day and another day when only Chicago O’Hare (ORD) was operating under IFR conditions (04/01/2007). All other airports were operating under optimum conditions on those two days. The Airport Efficiency database of ASPM was used to obtain the weather conditions and runway configuration during all 15-minute intervals of each day. This information was, in turn, necessary to estimate the capacity of each of the 34 airports in AND during the two days of interest.

We further limited our comparisons to flight arrival delays (as defined in Section 2.1.5), at the 34 airports. We cannot compare statistics regarding upstream delays and departure delays, which are reported by AND but are not available directly from the ASPM database. For example, even though ASPM reports the difference between the actual and the scheduled gate-out time for every flight, it is impossible to distinguish the part of that delay which can be attributed to local congestion from the part that is due to the late arrival of the previous flight of the same aircraft. Finally, as AND is a macroscopic model of airport delays, not meant to predict individual flight delays, we shall focus our analysis on average delays by airport and by hour of the day.

Figure 3-15 shows the average arrival delay as a function of the time of day at ORD as estimated by AND and as reported in the ASPM database for the day 04/01/2007. The estimates provided by AND follow a pattern very similar to the observed data. The maximum difference between the two graphs shown is 23 minutes (or about -18% of the ASPM-reported delay) for the hour 17:00-17:59. Overall, the average arrival delay for the entire day as estimate by AND is 49.0 minutes, while the observed ASPM arrival delay was 56.7 minutes (-13%).
Figure 3-16a shows the average arrival delay at all the other airports in the network for all those flights that are in the itinerary of aircraft that passed through ORD at least once on that day. Again we compare the AND results with the delays reported by ASPM. Figure 3-16b shows the absolute difference between the average daily arrival delay as computed by AND and as obtained by ASPM for every airport in the model. The difference ranges from 1% (at MIA) to 77% (at DCA). The average absolute error across all airports is 27%. As already mentioned, we are comparing the results of an analytical model against observed data for just one day of operations in the NAS. The statistics obtained for a single day of actual operations may not be reliable due to an insufficient number of data points. In the case of Tampa (TPA), for example, where the difference between the AND results and the field data is 42%, there are only 17 aircraft that at some point in the day passed through both ORD and TPA. Moreover, on any given day, events not modeled in AND affect significantly the reported delays. We also speculate that the large overestimation of the arrival delay at LGA and DCA is the result of the fact that most flights from ORD arrive at these two airports during times of high congestion and hence suffer an additional arrival delay as computed by AND. In reality, flights subjected to a high departure delay due to congestion at ORD and also expected to arrive during a period of congestion at their destinations, may be assigned a fixed departure time earlier at ORD, if a Ground Delay Program, is in effect thus avoiding a double penalty.

Figure 3-17 shows the average arrival delay suffered by all aircraft on the final arrival of that aircraft at ORD. The aircraft are classified according to the total number of visits they make to ORD that day. For example, the bars with a value of 2 on the x-axis correspond to the average arrival delay of the second flight into ORD experienced by all aircraft that passed through ORD exactly twice that day. Note that AND consistently overestimates the delay by a small amount and that clearly the error increases in absolute terms, as aircraft perform more visits to ORD. This is a result of the lack of a model within AND that captures airline reactions to congestion. In practice aircraft that visit ORD multiple times in a day and belong to one of the two “hubbing” airlines or to their regional partners would probably be
replaced by other available aircraft as they fall far behind schedule during days of high congestion, such as the one tested here.

We present next the delay statistics for the day 05/10/2007 as computed by AND. As can be seen in Figure 3-18 AND is, once again, able to closely replicate the general delay profile patterns observed at the three NY airports. Overall the delay profiles produced by AND are smoother than those obtained from the ASPM data throughout the day, as in some cases they fail to follow the local peaks and valleys of the arrival delay profiles in the observed data. The average arrival delay for each of the three NY airports is shown in Table 3.9. Clearly, AND predicts well, with very small errors, the average arrival delays at the three airports.

Finally, Figure 3-19 shows the absolute difference between the average arrival delay of all the flights that are flown by aircraft that pass through one of the NY airports at some point during the day, as estimated by AND and as reported by ASPM, for each of the 34 airports in the network. This difference ranges from 1% (at PHX) to 65% (at ORD). The average absolute error across all airports is 27% and for 22 out of the 34 airports the prediction error is less than 25%.

From the comparison of the AND results to real data for the two congested days in 2007 presented above, we may conclude that there is no clear bias in AND's system-
Figure 3-16: Average arrival delay by airport on 04/01/2007.

Figure 3-17: Average arrival delay of the final arrival of an aircraft at ORD as a function of the number of total visits to ORD on 04/01/2007.
Figure 3-18: Average arrival delay per hour at the three New York airports on 05/10/2007.
Figure 3-19: Average arrival delay by airport on 05/10/2007.
Table 3.9: Average arrival delay (in minutes) at EWR, JFK and LGA on 05/10/2007.

<table>
<thead>
<tr>
<th></th>
<th>ASPM (in mins)</th>
<th>AND (in mins)</th>
<th>% error</th>
</tr>
</thead>
<tbody>
<tr>
<td>EWR</td>
<td>28.5</td>
<td>27.8</td>
<td>-2.5%</td>
</tr>
<tr>
<td>JFK</td>
<td>35.3</td>
<td>34.9</td>
<td>-1.1%</td>
</tr>
<tr>
<td>LGA</td>
<td>34.6</td>
<td>37.3</td>
<td>7.8%</td>
</tr>
</tbody>
</table>

Wide delay estimates, nor in the local delay estimates at airports. AND will tend to overestimate delays related to local congestion at an airport during very congested days, since AND ignores airline reactions to congestion. On the other hand, as AND does not capture other sources of delays (such as en-route congestion) total delays generated by AND will tend to be underestimated. In the statistical validation of AND there is no indication on whether any of the two biases (under- or over-estimation of delays) prevails.

Overall, the tests in this and the previous section (as well as in several subsequent examples in this thesis) indicate that AND is an appropriate tool for exploring at a “macroscopic”, approximate level the types of issues that it was designed for, such as studying the impacts on delays of future changes to air traffic management systems that increase capacity to varying extents, or of the imposition of slot limits at certain busy airports.

### 3.4 A Detailed Example and Insights from the AND Model

In this section we present a detailed example of AND results based on one of the families of tests conducted with the model. The intent is to illustrate the types of inter-airport interactions that can be observed through AND, with emphasis on the insights that the model provides. The particular example described involves the 34 airport network in the continental United States described in Section 3.1.

Three scenarios will be considered here (Table 3.10), with each scenario representing different operating conditions at one or several airports. The same day of
operations, 8:00am GMT 10/15/2007 to 7:59am GMT 10/16/2007, was used for all three scenarios. In Scenario 1, all airports in the network operate under VFR conditions at optimum capacity levels for the entire day. In Scenario 2, the capacity at ORD is affected by bad weather for the entire day of operations and is reduced to the low IFR level of 36 movements per 15 minutes, while all other airports operate at optimum capacity. In this case, the scheduled demand at ORD exceeds the airport’s capacity for most of the day. ORD, which is already quite congested under optimum operating conditions, faces a large decrease in capacity when moving from optimum to IFR operating conditions. Finally, in Scenario 3, six airports in the Northeast—BOS, EWR, JFK, LGA, DCA and PHL—are affected by a storm, which decreases the capacity of all six airports to IFR levels for the entire 24-hour period.

Table 3.10: The three scenarios tested in AND.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity levels</td>
<td>All airports in optimum capacities</td>
<td>ORD in low IFR</td>
</tr>
</tbody>
</table>

3.4.1 Overall Network Behavior

When one or more airports operate under low IFR conditions one would expect to observe some increase in upstream delays at all the airports of the network. This is indeed the case when the expected upstream delays estimated by AND for Scenarios 2 and 3 are compared to those for Scenario 1, as shown in Table 3.11. Similarly, as expected, the fraction of flights that arrive with more than 15 minutes of delay undergoes an (often large) increase, either because of local congestion at the airport of arrival or due to upstream delay incurred at earlier destinations. Clearly, under Scenario 1, the three New York airports create the largest upstream delays even though the capacity at them was assumed to be optimum. Most of the upstream delay to the 34 airports under Scenario 1 is due to congestion at EWR, JFK and LGA. Note, as well, how upstream delays increase by very large amounts in Scenarios
2 and 3, when some critical airports operate in less-than-ideal conditions.

Table 3.11: Statistics for the 24-hour period for each of the airports in AND under each of the three scenarios.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upstream Delay</strong></td>
<td>Avg Local Flight Arr. Delay</td>
<td>% of Upstream Arr. delays &gt;15min delay</td>
</tr>
<tr>
<td>ATL</td>
<td>364</td>
<td>3.4</td>
</tr>
<tr>
<td>BOS</td>
<td>370</td>
<td>2.2</td>
</tr>
<tr>
<td>BWI</td>
<td>68</td>
<td>0.3</td>
</tr>
<tr>
<td>CLT</td>
<td>190</td>
<td>1</td>
</tr>
<tr>
<td>CLE</td>
<td>181</td>
<td>0.6</td>
</tr>
<tr>
<td>CVG</td>
<td>155</td>
<td>0.9</td>
</tr>
<tr>
<td>DCA</td>
<td>291</td>
<td>2.6</td>
</tr>
<tr>
<td>DEN</td>
<td>81</td>
<td>0.3</td>
</tr>
<tr>
<td>DFW</td>
<td>199</td>
<td>0.1</td>
</tr>
<tr>
<td>DTW</td>
<td>158</td>
<td>0.2</td>
</tr>
<tr>
<td>EWR</td>
<td>175</td>
<td>8.6</td>
</tr>
<tr>
<td>FLL</td>
<td>159</td>
<td>1.5</td>
</tr>
<tr>
<td>IAD</td>
<td>143</td>
<td>0.6</td>
</tr>
<tr>
<td>IAH</td>
<td>154</td>
<td>0.8</td>
</tr>
<tr>
<td>JFK</td>
<td>42</td>
<td>6.6</td>
</tr>
<tr>
<td>LAS</td>
<td>64</td>
<td>0.8</td>
</tr>
<tr>
<td>LAX</td>
<td>149</td>
<td>0.8</td>
</tr>
<tr>
<td>LGA</td>
<td>83</td>
<td>10.2</td>
</tr>
<tr>
<td>MCO</td>
<td>160</td>
<td>0.2</td>
</tr>
<tr>
<td>MDW</td>
<td>50</td>
<td>1.3</td>
</tr>
<tr>
<td>MEM</td>
<td>53</td>
<td>0.2</td>
</tr>
<tr>
<td>MIA</td>
<td>129</td>
<td>0.2</td>
</tr>
<tr>
<td>MSP</td>
<td>74</td>
<td>1.7</td>
</tr>
<tr>
<td>ORD</td>
<td>392</td>
<td>5</td>
</tr>
<tr>
<td>PDX</td>
<td>36</td>
<td>0.2</td>
</tr>
<tr>
<td>PHL</td>
<td>61</td>
<td>2.4</td>
</tr>
<tr>
<td>PHX</td>
<td>48</td>
<td>0.9</td>
</tr>
<tr>
<td>PIT</td>
<td>171</td>
<td>0.1</td>
</tr>
<tr>
<td>SAN</td>
<td>65</td>
<td>1.5</td>
</tr>
<tr>
<td>SEA</td>
<td>46</td>
<td>1.2</td>
</tr>
<tr>
<td>SFO</td>
<td>109</td>
<td>0.6</td>
</tr>
<tr>
<td>SLC</td>
<td>38</td>
<td>0.4</td>
</tr>
<tr>
<td>STL</td>
<td>105</td>
<td>0.3</td>
</tr>
<tr>
<td>TPA</td>
<td>79</td>
<td>0.2</td>
</tr>
</tbody>
</table>

A second observation concerns the large amount of expected upstream delay that ORD experiences in Scenario 2, when ORD operates in low IFR, while all other airports operate at their highest capacity. While this may appear counter-intuitive at first, it is perfectly logical when one considers the planned aircraft itineraries of the two airlines, American and United, which use ORD as a hub. Many of the aircraft of these two airlines have daily itineraries of the type "ORD—other airport(s)—ORD—other airport(s)—ORD—..., i.e., involve repeated visits to ORD during the course of the day. This means that, if an aircraft suffers a significant delay on one or more of its visits to ORD (e.g., as a result of poor weather, as in the case of Scenario 2), this delay may affect the entire subsequent schedule of that aircraft and may propagate through part or all of the day. In fact, the more times the aircraft lands at and takes off from ORD under adverse conditions, the more susceptible it becomes to upstream delay that may affect its subsequent visits to ORD. Essentially, the local delay suffered
at ORD on each visit to the airport "feeds on itself" and becomes upstream delay as far as the remaining scheduled flight legs for the day are concerned. In a nutshell, ORD causes high upstream delays onto ORD itself!

This phenomenon is much less intense at non-hub airports, as suggested by Table 3.11, since the number of aircraft that visit non-hubs, like LGA, several times during the course of a day is much smaller than in the case of hubs, like ORD. In fact, as can be seen from Table 3.11, although the local arrival delay at ORD in Scenario 2 is approximately 1.5 times greater than that at LGA in Scenario 3 (38.9mins vs 24.6mins), the average upstream delay at ORD is more than 4 times greater than that at LGA (30.2mins vs. 7.2mins).

Looking at the statistics for LGA in Scenarios 1 and 2 in Table 3.11 we may observe that, even though the capacity conditions at LGA are the same in both, the average local arrival delay there drops by 1.5 minutes (-15%) from Scenario 1 to Scenario 2. This is a direct effect of the extreme congestion at ORD in Scenario 2. Many flights from ORD are originally scheduled to arrive during peak hours at LGA. However, due to congestion at ORD their arrival time at LGA is pushed later, in periods were there is less arrival traffic. This "relieves" the peak hours of LGA and consequently reduces the local arrival delay there. This observation for airport interactions also holds between ORD and other congested airports (like EWR and JFK). We also note that under Scenario 2 approximately 30% of the average flight arrival delay at LGA is incurred by congestion at ORD (8.7mins of average local arrival delay, 4.0mins of average upstream delay and 12.7mins of average flight arrival delay).

### 3.4.2 Local versus upstream delays

As defined in Chapter 2, the local arrival delay is any delay incurred due to congestion at the airport of arrival, while the total flight arrival delay includes the local arrival delay and any upstream delay incurred by a flight at earlier destinations. It is interesting to point out, with reference to Table 3.11, that the more congestion increases at one or more airports (Scenarios 2 and 3) the more the average flight delay deviates from the average local delay at all other airports in the network. For
example, the average local arrival delay at ATL remains the same, at 3.4 minutes, under all three scenarios, but as more airports become congested, the flight delay increases from 4 minutes to 4.8 minutes in Scenario 2 and 5.3 minutes in scenario 3. This means that in Scenario 3 approximately 40% of the delay at ATL is caused by congestion elsewhere in the system.

Figure 3-20 shows how the total arrival delay suffered by all aircraft during the day breaks down between local and upstream delay at the 34 airports in the network under each of the three different scenarios. It is clear that when all airports operate under optimum conditions (Scenario 1) the local delay accounts for most of the delay at each airport, as for 22 out of the 34 airports the upstream delay accounts for less than 30% of the total arrival delay. When ORD operates in low IFR (Scenario 2), the upstream delay at all other airports becomes much more significant, ranging from 13% (at MDW) to 98% (at PIT) of the total delay at each airport. The upstream delay at 26 airports accounts for more than 50% of the total flight delay. Similarly, under Scenario 3, the upstream delays at every airport account for a much greater fraction of the total delays, compared to Scenario 1. A further interesting observation is that the upstream delay under Scenario 3 is higher than under Scenario 2 for only 6 out of the 34 airports shown, although there are six airports operating in low IFR under Scenario 3 and only one (ORD) under Scenario 2. This underscores the central role that ORD played in 2007, as a major and often-congested hub, in generating delays and spreading these delays throughout the national air transportation system.

### 3.4.3 The effect of delay propagation

**Shift in demand profiles and smoothing of the expected local delay**

In order to perform a more detailed and airport-specific analysis and to understand better the effects of the propagation of delays on individual airports, we now concentrate on just two congested airports, ORD and LGA—a hub and a non-hub, respectively. For Scenario 2, Figure 3-21a shows (i) the total number of landings and of take-offs that are initially scheduled in every 15-minute interval of a day of operations.
at ORD ("scheduled", in blue) and (ii) how this demand profile changes as a result of the propagation of delays ("adjusted", in red). Please note that the "adjusted" and the "scheduled" demand profiles coincide during the early part of the day, until roughly 10:30. After that time, however, the adjusted demand shifts very noticeably to the right by comparison to the scheduled demand. This is a consequence of the same phenomenon noted earlier: when congestion occurs at ORD, aircraft that must fly through the airport several times a day, suffer delays earlier in the day, which propagate to their subsequent "visits" to ORD, i.e., generate "upstream" delays for these aircraft. This moves the instants when they demand access to ORD to later times than in the original schedule. The net effect is to spread demand more evenly than originally scheduled. The propagation of delays thus effectively creates a "smoother" demand profile at ORD later in the day and especially towards the end of the day when, as can be seen in Figure 3-21a, most of the peaks are "evened out" and the demand level approximately "tracks" the available capacity.

In an analogous way, Figure 3-22a shows the changes in the demand profile at LGA under Scenario 3. Although many of the peaks of the scheduled profile are again evened out, the smoothing effect in this case is not nearly as strong as that observed at ORD. As previously explained, the reason is that LGA, in contrast to ORD, is not a hub: there are many more aircraft which fly through ORD more than
Figure 3-21: Demand and delay profiles at ORD in Scenario 2.

The shift in the demand profile of ORD under Scenario 2 and resulting smoothing and spreading of demand over more hours of the day leads to a much lower level of expected local delay and a much less “peaked” profile of the delay by time-of-day, as shown in Figure 3-21b. Note that the estimated local delay, based on the originally scheduled demand profile at ORD (shown in Figure 3-21a in blue) peaks at almost 160 minutes (Figure 3-21b) of expected delay for a movement scheduled at 21:45. But, with the “adjusted” demand profile (Figure 3-21a in red), the local delay peaks once during a 24-hour day than through LGA.
at approximately only 60 minutes (Figure 3-21b). Thus, a flight arriving at ORD around 21:30 on that particular day would be expected to experience a local delay of about 60 minutes upon arrival, as opposed to about 2.5 hours with the original demand profile. The explanation is that, as shown in Figure 3-21a, delays earlier in the day cause many flights, which were originally scheduled to arrive or depart around 21:30, to request service later in the day, thus making the time interval around 21:30 less congested than originally. The corresponding effect at LGA (Figure 3-22b) is similar, but not as strong as at ORD, because LGA is not a hub and fewer aircraft visit it more than once in a typical day.

**Total Delay**

Figure 3-23a shows the expected total arrival delay suffered by every flight that operates at ORD under Scenario 2 during the day examined. Note that the expected total delay is the sum of the expected upstream delay that each aircraft experienced at locations that it previously visited during the day and of the expected local delay suffered upon arrival at ORD. Clearly, as also shown in Figure 3-23b, the expected total delay increases as the day progresses and as aircraft experience delays at various points in the network that cannot be absorbed by the slack in their scheduled itineraries. Moreover, observe in Figure 3-23a that not only does the expected total delay increase on average later in the day (until about 22:00 hours), but its variability increases greatly as well. This is also shown in Figure 3-23b through the increase later in the day of the standard deviation of the total delay. Hence, as the day progresses, the expected deviation from schedule, i.e. the expected total delay, increases, while the reliability of the schedule deteriorates i.e., the uncertainty about flight arrival and departure times increases. This is a phenomenon familiar to experienced US air travelers.

In Figure 3-23a the graph labelled “expected local delay based on the adjusted schedule” is the same as the graph showing the “adjusted” expected local delay in Figure 3-21b. Note that in Figure 3-23a the points corresponding to many flights fall on top of this graph, i.e., the expected total arrival delay experienced by many flights
Figure 3-22: Demand and delay profiles at LGA in Scenario 3.

is equal to the “expected delay based on the adjusted schedule”. The flights that belong to this category are those performed by aircraft that operate for the first time during the day at ORD. As long as they have not incurred any upstream delay at other airports before coming to ORD, the expected total arrival delay for these flights consists only of the expected local delay upon arrival at ORD. It is evident that these flights benefit from the smoothing of the demand profile at ORD as they experience smaller delays than they would have suffered under the original schedule in ORD. This underscores the usefulness of models like AND: had we not accounted for delay
propagation and the resulting shift of the demand profile at ORD, we would have
estimated that flights by aircraft operating for the first time at ORD would suffer an
expected delay equal to the (much higher) delay shown by the “expected local delay
based on the initial schedule” graph of Figure 3-23a.

Figure 3-23a also shows the breakdown of expected total delay for those flights
performed by aircraft not visiting ORD for the first time. For each of the points
lying above the “expected local delay based on the adjusted schedule” graph, the
corresponding flight suffers an expected local delay equal to the value of the “expected
local delay based on the adjusted schedule” plus an upstream delay equal to the
difference between its expected total flight delay, as shown in Figure 3-23a, and its
expected local delay. For example, if we track aircraft N869AE scheduled to visit
ORD 5 different times that day, we observe that for the first two arrivals of the
aircraft at ORD its total expected arrival delay is the same as the local arrival delay
at ORD upon arrival. However, at the next three arrivals at ORD, aircraft N869AE
carries a large and increasing amount of upstream delay due to its previous visits at
ORD. Hence, the total delay it suffers is significantly greater than the local delay on
all three of these occasions.
3.5 AND European Air Transportation Network

The US and European air transportation systems are currently the most congested ones in the world. At the same time, there exists a fundamental difference in the way they operate. Practically all major airports in Europe operate with slot limitations,
which means that demand at every airport is controlled so that the delays rarely become excessive. By contrast, no such restrictions exist on the number of flights that can be scheduled at all but a very few US airports. Moreover, European airports operate under Instrument Flight Rules (IFR) under all weather conditions, in contrast to the United States where—weather permitting—visual separations may be used.

In order to investigate the implications for delays of these differences between the two systems, we have applied the AND model to a set of the 19 busiest European airports, as shown in Figure 3-24. The data for the capacities of each of the European airports, as well as the aircraft itineraries that were input into AND have been provided by the Central Office of Delay Analysis (CODA) of the Eurocontrol [15]. However, these itineraries correspond to only approximately 60% of the daily flights operating at European airports. In order to complete the demand profiles of arrivals and departures at the 19 airports modeled by AND, we used the online database of FlightStats [9], that includes all the actual and scheduled arrival and departure times of all scheduled flights in the world.

Figure 3-24: European AND airport map.
3.5.1 A Comparison Between the US and EU Airport Hubs

In this section we compare two of the main hubs in each system, namely Chicago O’Hare (ORD) in the US and Frankfurt International (FRA) in Europe.

The statistics shown in Figure 3-25 provide some background on the characteristics of operations at each airport. Figures 3-25a and 3-25b show the breakdown of traffic by airline at each airport. The main hub carrier in FRA is Lufthansa with its regional partners, while at ORD the dominant carriers are United and American Airlines. Figure 3-25c shows the percentage of aircraft passing one or more times from each hub. Note that 36% of the aircraft passing through ORD visit the airport more than once, while only 22% do so at FRA. We can then hypothesize that delay propagation from early in the day to late hours will be much more intense at ORD, since more flights will accumulate delay each time they pass through the airport.

![Figure 3-25: Flight share by airline at FRA and ORD.](image-url)
The comparison we perform is for a day when there is high congestion at both airports due to poor weather conditions. The capacity reduction from optimum levels is 20% for ORD and 9% for FRA.

![FRA Demand Profiles](image1)

![ORD Demand Profiles](image2)

(a) FRA profiles  
(b) ORD profiles

![Average arrival delay per 15-minutes at FRA](image3)

![Average arrival delay per 15-minutes at ORD](image4)

(c) FRA delays  
(d) ORD delays

Figure 3-26: Analysis at FRA and ORD.

Figures 3-26a and 3-26b show (i) the total number of landings and of take-offs that are initially scheduled in every 15-minute interval of a day of operations (scheduled, in blue) and (ii) how this demand profile changes as a result of the propagation of delays (adjusted, in red), at FRA and at ORD, respectively. As described in Section 3.4, due to propagation of delay the adjusted demand after the morning hours shifts very noticeably to the right by comparison to the scheduled demand, while most of the high peak demand instances are made less sharp. In contrast to ORD we observe (Figure 3-26a) that at FRA the shift to the right of the demand profile is not as strong and the scheduled and the adjusted demand profiles coincide for most of the day.

Figures 3-26c and 3-26d present the average arrival delay per 15 minutes at FRA
and ORD, respectively, broken down into local and upstream delay. Looking at these figures, it is evident that, at ORD, the upstream delay becomes almost as important a contributor to the total arrival delay as the local delay as the day progresses, accounting for almost 50% of the arrival delay after 5pm. This is completely different from what happens at FRA. Overall, the average local delay throughout the day is 24 minutes at FRA and 31 minutes at ORD, while the average upstream delay is 3 and 13 minutes, respectively. Thus, the local delay at ORD is only 22% greater than the local delay at FRA, while the upstream delay is 77% greater. This difference in upstream delay is a result of the fact that there are many more flights that visit the airport more than once during the day at ORD as explained earlier and shown in Figure 3-25. We have performed an analysis of turn times for European airlines similar to the one presented in Section 3.1.2. We concluded from that analysis that the turnaround time slack is about the same in both networks, hence not contributing to the major differences in propagated delay observed here. The difference can indeed be attributed for the most part to the different structures of the itineraries of aircraft passing through ORD and FRA, as described above.
Chapter 4

Extensions to the Airport Network Delays Model

The Airport Network Delays model has a flexible structure, as described in Chapter 2, that allows the addition of various features. This chapter focuses on two of the most important extensions implemented into the AND model, so far. We describe, first, how we model Ground Delay Programs, an important strategic practice of the Air Traffic Management system. Ground Delay Programs are used to mitigate congestion at airports when demand exceeds capacity by holding aircraft at their airport of origin. We also introduce an alternative queuing engine, which is non-stationary and assumes deterministic demand and service rates. We compare the results of this deterministic queuing engine to those of the stochastic one, described in Chapter 2. By using the two queuing engines, we also provide an approximate analysis of the potential benefits stemming from one of the features of the Next Generation Air Transportation System.

4.1 Modeling Ground Delay Programs

In order to provide a more accurate representation of the National Airspace System, it is necessary for AND to incorporate a model of one of the most fundamental operational tools of the Air Traffic Management system, the Ground Delay Program (GDP). The main idea behind a GDP is that its safer and less expensive to experience
the same amount of delay on the ground than in the air. Hence during periods of extreme congestion at an airport, which typically occur due to inclement weather conditions (but also, at some airports such as LaGuardia, due to excessive volume of traffic even during good weather) it is better to hold aircraft on the ground at the airport of origin rather than queue them up in the airspace. When a GDP is initiated, flights are assigned a ground hold according to their expected landing time at the congested airport to which they are headed.

Ball [2] and de Neufville and Odoni [12] provide extensive descriptions of the operation of a GDP. In practice, a GDP is initiated when long delays are expected at an airport due to the imbalance between the demand and the expected service rate. The procedure is executed in three steps. In the first step, FAA assigns arrival slots to flights according to the expected arrival acceptance rate (AAR) of the affected airport by using a First-Scheduled-First-Served scheme—otherwise known as the Ration-By-Schedule method. In the second step, the airlines may plan recovery actions that attempt to minimize flight disruptions. Within this context airlines are allowed to swap arrival slots between their own flights and/or cancel some of them. In the final step, the FAA modifies the initial slot assignments to incorporate the changes proposed by the airlines and compresses the arrival schedule at the affected airport to account for any cancelled flights. The last two steps of the process are also referred to as Collaborative Decision Making (CDM).

Ground Delay Programs have been the primary tool of the Federal Aviation Administration (FAA) for Traffic Flow Management (TFM) for the last three decades. According to data provided by Metron Aviation [35] there was a GDP in effect in at least one airport in the United States in a total of 347 days in 2007. In addition, the FAA was issuing GDPs for more than 5 airports daily on average, while the average GDP length in 2007 exceeded 8.5 hours. As described by Barnhart et al. [3] delays during GDPs account for 20% of all delays in the US air transportation system.

Modeling Ground Delay Programs within an airport network model, such as AND, is important since they greatly affect delays and the propagation of delays in networks of airports. As an example consider a flight scheduled to take off from New York’s
La Guardia Airport (LGA) with Chicago O’Hare Airport (ORD) as its destination, during a period when the latter is operating a GDP. This flight will then depart from LGA at a time that depends on its arrival slot at ORD issued by the GDP. Without the GDP model the aircraft would join the departure queue at LGA at a time that might be totally different from the departure time issued by the GDP. Because of the GDP, the aircraft may therefore affect the queue of a different set of aircraft on departure from LGA. Moreover, the GDP model can also be used to distinguish and handle differently flights that are exempt from a GDP and which, on average may experience lower delays than the non-exempt flights.

Long et al. [33] have, also, identified the need for an airport network model to incorporate Ground Delay Programs and have included one such model in their LMINET2 model. However, as explained below, that model differs significantly from the one developed for AND.

LMINET’s GDP model is initiated only if there is bad weather at an airport and works in the following way: for the first time epoch of 15 minutes during the GDP, if the demand at the subject airport exceeds capacity, the model’s algorithm selects randomly a set of excess flights that have not yet departed and delays them by 15 minutes. In the next 15-minute time epoch, those delayed flights have priority to depart if they can fit within the available capacity. This iterative scheme will run until no more flights are delayed. Clearly all flights that are affected by the GDP will experience a delay that is always an integer multiple of 15 minutes. In contrast AND’s GDP algorithm is initiated whenever there is increased congestion, even on good weather days, and will assign flights to slots on a First-Scheduled-First-Served basis and consistently with the airport acceptance rate (AAR). In addition the algorithm can perform slot swaps and schedule compression, as will be described in the next section.

In order to provide some background on the Ground Delay Program model we provide here a brief statistical analysis regarding flights affected by GDPs. We used the ASPM “Individual Flights” database [19] to look at delay statistics for all flights in 2007 with destination Chicago O’Hare (ORD) that arrived while ORD was operating
Figure 4-1: Average ground hold time as a function of flight time at ORD.

with a Ground Delay Program. For this purpose, we also used a database provided by Metron Aviation [35] that includes all Ground Delay Programs, to obtain the exact start and end time and date, of each GDP period in 2007 at ORD. Combining the two databases we were able to determine all flights that landed at ORD while a GDP was being executed. Furthermore, the ASPM database includes per flight data for the ground hold time, the arrival delay compared to schedule and the difference between the expected arrival time of the flight according to the GDP plan and the actual arrival time.

Figure 4-1 clearly shows that flights with more than four and a half hours of flight time receive a smaller amount of ground hold at the departure airport than shorter flights. The average ground hold for affected flights with flight time less than 4.5 hours is 44 minutes while for flights with more than 4.5 hours of flight time, it is only 9 minutes. Moreover only 40% of flights of more than 4.5 hours received a ground hold while almost 80% of flights of less than 4.5 hours of flight time that arrive during a GDP received a ground hold. In Figure 4-1 we can also see that flights that receive ground holds actually arrive very close to the expected arrival time assigned by the
Table 4.1: Average ground hold with respect to total GDP length, of flights with more than 4.5 hours flight time destined to ORD.

<table>
<thead>
<tr>
<th>GDP Length (hrs)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Grnd Hold (mins)</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

GDP. We will use this observation in the GDP algorithm, as will be explained in detail in section 4.1.5, to make a necessary assumption about the operation of GDPs. Table 4.1 clearly shows that, even for very long Ground Delay Programs at ORD, flights that are longer than 4.5 hours are essentially exempt from the respective GDP. This is consistent with the FAA’s typical practice of exempting from GDPs flights longer than four hours, due to the fact that weather forecasts are imprecise for such long time horizons.

### 4.1.1 Ground Delay Program Algorithm

The GDP algorithm developed for AND follows the rules of the GDP procedure that is currently in practice in the United States. To summarize, a GDP procedure is executed as follows: a) arrival slots are assigned to flights according to a First-Scheduled-First-Served procedure, b) flights that have already departed from their destination when the GDP is planned are exempt, c) under the Collaborative Decision Making (CDM) program each airline is allowed to swap arrival slots between its affected flights and/or cancel flights, d) a final compression of the schedule takes place to account for unused slots and e) all flights with an assigned arrival slot are given an Estimated Departure Control Time (EDCT), which is basically a departure slot such that each flight will reach its destination at the time of its assigned arrival slot. The GDP algorithm in AND implements the first two and the final step exactly and steps c and d approximately—as we discuss in detail in this section.

The GDP algorithm developed for AND is broken down into 3 main parts:

1. GDP initiation

2. Slot assignment
(a) Assign Expected Arrival Control Times (EACT) and EDCT to all affected flights
(b) Perform slot swaps between flights of the same airline
(c) Compress schedule
(d) Reassess EDCTs.

3. Decision on whether to continue or terminate the GDP

4.1.2 GDP initiation

At the end of every time period \( h \), the GDP module looks at the estimates of the queuing engine of the future arrival delay \( W_{\alpha,q}(t) \) (using the notation from Chapter 2, Equation 2.5) at every airport \( \alpha \in K \) in order to determine if a GDP should be initiated.

We define the "planning horizon of a GDP" as the time that elapses between the planning of the GDP and the time when the first flight that was assigned an arrival slot by the GDP lands. We set the length of the GDP to \( k\Delta h \) where \( k \in I^+ \). In practice the planning horizon is typically three hours, which means that at any point in time, \( t \), all flights scheduled to arrive at the congested airport 3 hours or more later than \( t \) may be affected by the GDP.

In order to initiate a GDP at one of the airports in the AND network one or both of the following two criteria has to be fulfilled. If the current time period is \( h \), we initiate a GDP if:

1. The arrival service rate at an airport, starting at period \( h + k\Delta h \), drops by a certain percent (user-specified input) below its VFR value for at least 30 minutes continuously.

2. The expected delay at an airport increases above a specified threshold (typically specified as 20 minutes in our tests) at period \( h + k\Delta h \) and is expected to stay above that threshold for at least 30 minutes.
These two conditions allow a GDP to be initiated not only if an airport is affected by bad weather, so that its capacity deteriorates, but also if the demand at the airport is sufficiently higher than the airport’s VFR capacity to create significant delays. The latter is often observed, for example, at LaGuardia and Newark airports in the late afternoon when the arrival demand exceeds the respective arrival throughput for a few consecutive hours.

Clearly, at the moment when a GDP is initiated the arrival queue at the congested airport must be much greater than zero. To account for that, the GDP algorithm obtains from AND the expected length of the arrival queue at that moment and shifts the first assigned slot by an amount of time equal to the expected arrival queuing time at the moment of the GDP initiation.

4.1.3 Slot assignment

When the GDP is initiated, at period \(h_0\), the algorithm looks for all flights scheduled to arrive in the range \([h_0 + k\Delta h, h_0 + k\Delta h + \frac{60\text{mins}}{\Delta h}]\) that have not yet departed (by period \(h_0\)) and inserts them into a virtual (priority) queue modeled in the GDP algorithm by a heap. These flights are not exempt from the GDP. All flights that have already taken-off by period \(h\) and are expected to land within the same range \([h_0 + k\Delta h, h_0 + k\Delta h + \frac{60\text{mins}}{\Delta h}]\) join a separate virtual queue that keeps track of the flights exempted from the GDP. The expected time of arrival for a non-exempt flight is given by \(AA(f)\) (the Adjusted Arrival time of \(f\)) as estimated by AND.

The algorithm also creates arrival slots that are equally spaced in time for every period in \([h_0 + k\Delta h, h_0 + k\Delta h + \frac{60\text{mins}}{\Delta h}]\) according to the Arrival Acceptance Rate (AAR) of the issuing airport. A portion of the slots in each time period is dedicated to the exempt flights that are expected to arrive in that period. Hence, if we let \(N_e\) be the count of exempt flights expected to arrive in a period and \(N_s\) the number of slots available in that period:

- The first \(N_s - N_e\) slots in the period are assigned to the non-exempt flights.
- The last \(N_e\) slots are assigned to the exempt flights.
Clearly this rule guarantees that no exempt flight will be delayed more than $\Delta h$ minutes.

The heap structure that is used to store flights in the algorithm is basically a priority queue in the form of a binary tree with two properties:

- All nodes of the tree store flights with scheduled arrival times that are earlier than or concurrent with those of their descendants. Thus, the top node of the tree holds the flight with the earliest arrival time, as shown in Figure 4-2.

- It is almost complete, which means that all levels of the tree are complete except for the bottom level.

The heap structure makes the operation of adding a flight in the queue or removing the earliest scheduled flight from the queue very efficient. In particular the complexity of each of these operations is $O(\log n)$ where $n$ is the number of flights in the queue. Figure 4-2 shows an example of the representation of a virtual queue in the GDP algorithm.

Slots are assigned by removing the top node from the tree and assigning it to the next available slot. This process is repeated until all slots in the range $[h_0 + k\Delta h, h_0 + k\Delta h + \frac{60 \text{ mins}}{\Delta h}]$ have been assigned. In the example of Figure 4-2 we assume that the GDP is planned at 13:00 for the period 15:00-15:15. Hence all flights that have departed before 13:00 are exempt from the GDP. The AAR at the arrival airport is 5 flights for the period 15:00-15:15, i.e. $N_e = 5$. The heap of Figure 4-2b shows that two exempt flights are expected to arrive during the period 15:00-15:15 (AA32 and AA777) hence the GDP algorithm assigns them the last two slots of the period ($N_e = 2$). The remaining 3 slots are assigned to flights from the non-exempt heap shown in Figure 4-2a. Flights UA532, UA621 and UA789 are assigned EACTs at 15:00, 15:03 and 15:06 respectively and the exempt flights AA32 and AA777 are assigned arrival slots 15:09 and 15:12 respectively.

The remaining flights (exempt and non-exempt, as shown in Figures 4-2c and 4-2d respectively) will be assigned slots in the following time periods. Assuming 5 arrival slots are available in the next 15-minute period (15:15-15:30) the GDP will assign the
Figure 4-2: Heap representation of the arrival queue before and after the slot assignment in the period 15:00-15:15. (a) shows 10 non-exempt flights scheduled to arrive between 15:00 and 15:20; (b) shows the 5 exempt flights and their respective departure times in parenthesis. (c) and (d) show the arrival priority queue of the non-exempt and exempt flights, respectively, after the slot assignment in 15:00-15:15.

First two arrival slots to the non-exempt flights UA643 and DL52 and the last three slots to the exempt flights UA39, CO673 and CO674.

Similarly, the remaining flights (4 non-exempt flights left) will be assigned slots after 15:30.

Working backwards the GDP algorithm assigns \textit{EDCT}s to the non-exempt flights by subtracting the flight time from \textit{EACT} to determine the \textit{EDCT}. However, there are cases when the \textit{EDCT} of a flight is earlier than the time when the aircraft becomes available for departure. This might happen, for example, in the case of aircraft that visit busy airports early in the day, hence carrying an upstream delay throughout much of their daily schedule. If such an aircraft is scheduled to visit an airport issuing GDPs towards the end of the day, it might not make its assigned \textit{EDCT}. 
This is especially true in airports issuing a GDP that lasts for a long period of time, with some aircraft scheduled to visit the airport multiple times during the GDP.

In order to account for such cases the GDP algorithm looks for flights $f$ with $EDCT(f) < AD(f)$, where $AD(f)$ is the adjusted departure time of flight $f$ as estimated by AND before the initiation of the GDP (see Section 2.1.3 and Equation 2.12). If such a flight $f^*$ is found, the algorithm searches through the rest of the flights affected by the GDP for a flight $f^{**}$ operated by the same airline as $f^*$ with $EACT(f^*) < EACT(f^{**})$ that can make the assigned arrival slot of $f^*$. If no such flight is found then $f^*$ is given an $EACT(f^*) = AD(f^*) + \text{flight time}(f^*)$ and all flights that were assigned a slot between the original $EACT(f^*)$ and the new $EACT(f^*)$ are moved earlier by one slot. If a flight $f^{**}$ of the same airline is found, then this flight is moved up in the schedule to occupy the slot originally assigned to $f^*$. Finally, working backwards again, the GDP algorithm assigns new $EDCT$s according to the changes in the assigned slots. This process will continue until all non-exempt flights receive a feasible $EDCT$.

4.1.4 Continuation and Termination of the GDP

The GDP algorithm reassesses the situation at the airports running GDPs every 60 minutes and decides whether to continue or terminate the GDP in each case. The GDP may, of course, not end until much later than the time when the capacity at the GDP airport(s) has been restored, i.e., until the arrival queue has reached a tolerable level.

As long as in every hour after the initiation of the GDP the heap with the arrival queue has non-zero size, the GDP will continue in the manner described thus far. That is, for the next hour $[h_0 + k\Delta h + \frac{60\text{mins}}{\Delta h}, h_0 + k\Delta h + 2\frac{60\text{mins}}{\Delta h}]$ new slots will be created, all flights that are scheduled to arrive in that time range are added to the arrival heap, and slots are assigned by removing flights from the heap. Finally, slot swaps and schedule compression lead to the final $EDCT$s for the affected flights.

During the assignment of slots, the GDP algorithm will terminate if the $EACT$ of a flight is less than a threshold amount (typically 10-15 minutes) later than its
scheduled arrival time. The GDP algorithm is terminated by removing all remaining flights from the heap of the arrival queue. After termination flights will not be assigned arrival slots but may continue to experience delays according to the delay estimates from AND.

Overall, the complexity of the GDP algorithm is $O(n \log n)$ since it scans through $n$ flights in linear time, adds them to the heap in $O(\log n)$ and removes them in $O(\log n)$.

### 4.1.5 Link with AND

After running the GDP algorithm for one hour, AND restarts and executes schedules according to the EDCT and EACT for the flights that are affected by the GDP. For all these flights AND ensures that they will not experience any additional delay due to local traffic delays at the airport of departure. This means that flights with an EDCT will overtake other aircraft in the departure sequence at their airports of origin in order to adhere to their EDCT. Thus, for every flight to which an EDCT has been assigned, AND computes the expected length of the departure queue at the departure airport of that flight at the moment of the EDCT, in order to obtain an estimate of the number of aircraft that it will have to overtake. Moreover, it is assumed that each flight will reach its destination exactly at the EACT assigned by the GDP, an assumption that relies on the findings presented in Figure 4-1 and discussed in Section 4.1.

The process described above can be applied to more than one airports executing Ground Delay Programs simultaneously. In fact, the advantages, in terms of computational efficiency and structural flexibility, of the heap structure implemented in this algorithm become most apparent when more than one airports issue Ground Delays for a long period of time, e.g., when the New York Metroplex is operating in bad weather for an entire day.
In order to test and validate the GDP algorithm of AND we chose a day in 2007 when only one major airport was operating a Ground Delay Program. According to Metron’s GDP database [35] on 04/01/2007 ORD was experiencing bad weather and thus operated a GDP for the greatest part of the day. The only other airport that operated a Ground Delay Program that day was Newark Liberty International Airport for approximately 2 hours. This is the same day as the one used in the validation of AND presented in Chapter 3. Metron’s database includes values for the number of slots issued per 15 minutes for every GDP. We have used these values to adjust the AAR at ORD in AND, while the GDP was running. For the rest of the day we used the service rate suggested by the ASPM database that depends on the runway configuration and weather conditions at ORD. Figure 4-3 summarizes the service rate and scheduled demand rate at ORD for 04/01/2007.

For this test case, the first GDP slot was assigned by the GDP algorithm of AND at 12:15pm and the last one at 00:20am. According to the ASPM database the first flight with a recorded EDCT with destination ORD was assigned a slot at 12:10pm and the last one at 00:33. Figure 4-4 shows the distribution of the average
Arrival delay throughout the day at ORD as estimated by AND without the GDP algorithm (AND), AND with the GDP algorithm (AND/GDP), and as reported in the ASPM database. The results obtained from AND and AND/GDP are close to each other. However, AND/GDP improves slightly on the estimation of delays when compared to the ASPM data.

Figure 4-5a shows the average arrival delay at all airports in the network for all those flights that are in the itinerary of an aircraft that passes through ORD at least once that day. Again we compare results from AND, AND/GDP and the reported delays from ASPM. Figure 4-5b shows the absolute percent difference between the AND and AND/GDP estimates and the ASPM data. Observing the two figures no definite conclusion can be drawn regarding any improvement of the delay estimates when using AND/GDP instead of AND. Over all airports the average absolute error is 27% and 28% for AND and AND/GDP respectively.

Figure 4-6 shows the average arrival delay for all aircraft at their final arrival at ORD classified by the total number of visits of each aircraft at ORD. For example,
the data with a value 2 on the x-axis corresponds to the average arrival delay of the second flight into ORD of all aircraft that passed through ORD exactly twice that day. Observing this figure we can see that AND/GDP improves significantly the results of AND, especially for aircraft that pass through ORD three or four times daily. This happens primarily due to two reasons, which we explain below with two simple examples.

First, if a flight is assigned an arrival slot, it will not experience any additional departure delay at the airport of origin. Consider, for example, a Comair flight
scheduled to depart JFK airport at 17:40 with ORD as its destination and scheduled arrival time 19:55 (all times GMT). This flight carries no upstream delay so it is ready to depart from JFK at 17:40. However, due to the GDP at ORD, this flight is assigned an arrival slot at 21:57 and a corresponding EDCT at 19:52 (2hrs and 12mins of ground hold). The departure delay at JFK was estimated by AND to be approximately 15 minutes at 17:40 that day. When running AND without the GDP this flight would then be assigned a departure delay of 15 minutes and upon arrival at ORD an additional arrival delay of approximately 2 hours and 10 minutes. When running AND/GDP the Comair flight will not experience any departure delay at 19:52. Thus, for such cases AND/GDP will produce smaller delays than AND.

The second reason is that flights with an assigned EDCT can compensate for some of the upstream delay that they might be carrying due to earlier congestion. As an example, consider American Airlines aircraft N289AA flying the schedule shown in Table 4.2. The fourth flight of this aircraft carries 25 minutes of upstream delay, which is shown by the difference between the adjusted and scheduled departure time of that flight, due to earlier congestion at ORD. This means that the flight cannot depart before 21:55. However, this flight is given an EDCT at 23:23 according to
Table 4.2: Schedule and delays of American Airlines' N289AA aircraft flying through ORD (All times in GMT). “ZZZ” signifies an out-of-the-network airport.

<table>
<thead>
<tr>
<th>Flight Order</th>
<th>Origin</th>
<th>Dest.</th>
<th>Sch Dep Time</th>
<th>Adj Dep Time</th>
<th>EDCT Delay (mins)</th>
<th>Loc Dep Sch</th>
<th>Adj Arr Time</th>
<th>Arr Slot Delay (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ORD</td>
<td>LGA</td>
<td>13:35</td>
<td>13:35</td>
<td>4</td>
<td>15:45</td>
<td>15:45</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>LGA</td>
<td>ORD</td>
<td>16:29</td>
<td>17:22</td>
<td>10</td>
<td>20:50</td>
<td>21:25</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>ORD</td>
<td>ZZZ</td>
<td>19:50</td>
<td></td>
<td></td>
<td>22:40</td>
<td>23:05</td>
<td>00:23</td>
</tr>
</tbody>
</table>

its scheduled arrival time at ORD, which means that the upstream delay does not affect any longer the flight’s schedule. If AND was run without the GDP this flight would continue carrying the upstream delay, depart from ZZZ at 21:55, and assigned an additional delay of approximately 2 hours upon arrival at ORD.

As expected, the arrival delay of flights arriving for the third and fourth time at ORD in Figure 4-6 still remains higher than observed in ASPM, since AND does not capture airline responses to congestion, such as aircraft swaps and utilization of spare aircraft at hubs. However, there is a clear improvement from AND to AND/GDP, as one might expect.

One more feature that the GDP algorithm adds to the AND model is that it captures the effect of aircraft that are subjected to ground holding on the departure queues at their respective airports of origin. Using, again, the example of Table 4.2 and specifically the second flight of that aircraft, we notice that, in the absence of the GDP at ORD, the flight would join the departure queue at 16:29. The departure demand for that 15-minute period is 8 aircraft, as given by AND’s output of the adjusted schedule at LGA. With the GDP at ORD the flight will depart at 17:22 increasing the departure demand for that 15-minute period from 8 to 9 aircraft and reducing the demand at 16:15-16:30 to 7 aircraft. This translates to an increase of expected departure delay during the 17:15-17:30 period from 5 to 6 minutes, a 20% increase. According to the estimates of the queuing engine, the expected queue length at 17:22 at LGA will be 6 aircraft. These 6 aircraft will be overtaken by N289AA in the departure sequence so that N289AA may depart at a time consistent with its EDCT.
4.2 Alternative Queuing Engine: Deterministic Queues

As described in Chapter 2, the structure of AND allows the use of any queuing model to estimate the local delay statistics at each of the network’s airports. With reference to Figure 2-1 and the description in Chapter 2, AND is comprised of two main components, the Queuing Engine and the Delay Propagation Algorithm, that are distinct and replaceable. As explained, the DPA utilizes the delay estimates from the QE in order to assign these delays to all flights and update the schedules of airports—Figure 2-4.

An alternative queuing model that has been implemented and tested in AND assumes dynamic deterministic demand, dynamic deterministic service rate, one server and infinite queue capacity, denoted as a $D(t)/D(t)/1$ system in queuing theory. In this section we shall refer to the $D(t)/D(t)/1$ system as the “deterministic” queuing engine and the $M(t)/E_k(t)/1$ system, already described in Chapter 2, as the “stochastic” one. Correspondingly we refer to the “deterministic AND” and “stochastic AND” as the two versions of AND that utilize these two queuing engines, respectively. In fundamental contrast to a stochastic queuing system, a deterministic one will produce delays only when the demand exceeds the capacity at an airport. The deterministic model described below, thus provides lower bounds on delays under given demand and capacity rates.

The value of implementing a deterministic queuing engine within a network model is twofold. First, although there have been many studies in which each of the two queuing models has been used to obtain estimates of flight delays at individual airports (see Simaiakis [49], Long [32] and Malone [34] for stochastic results and Hansen [24] for deterministic airport models), and a few studies where results from the two models have been compared (Hansen et al. [25]), there has never been a comparison of the results within the context of a network model. This is accomplished in this work in which we model the US airport network under the same conditions (flight schedules and airport capacities) using both the deterministic and the stochastic AND
for various scenarios.

The second goal of this AND extension is to provide estimates of one type of delay savings that may be potentially feasible with the implementation of the Next Generation Air Transportation System (NextGen). One prominent NextGen feature is the implementation of 4D trajectories, which may lead to reduced and more precise separations between aircraft; for more details refer to [38]. But a more direct benefit of 4D trajectories is that increased precision of flight paths and better predictability of times when demands will occur at airports will lead to more efficient operations during congested periods. One way to model this efficiency increase is by reducing the stochasticity of the queuing models of airports. Assuming that the current terminal airspace operations are reasonably accurately modeled by the stochastic queuing engine, as discussed Chapter 3, then high precision operations, as envisioned in NextGen, may be better modeled through the deterministic queuing engine. In other words, by comparing the difference in the delay estimates generated by the stochastic and deterministic models, we can estimate an upper bound for the delay savings that will arise just from the reduction in stochasticity, without taking into account any changes to airport capacities.

### 4.2.1 Deterministic Queuing Engine

As is the case with the stochastic queuing engine, the deterministic estimates delays by using aggregate data concerning, dynamic (time-varying) demand. If the day is divided into sub-periods \( h = \{1...H\} \) with length \( \Delta h \), we define \( \lambda(h) \) and \( \mu(h) \) as the aggregate demand rate and service rate, respectively, during period \( h \) in terms of number of aircraft movements per period. A simple way to visualize and estimate queue statistics obtained through this deterministic queuing engine is through the cumulative—over time—diagram of demand and airport throughput [12]. We define as airport throughput the number of aircraft that are served in each period. It is then clear that the cumulative throughput at the end of period \( h \), \( F_T(h) \), cannot exceed the cumulative demand \( F_\lambda(h) \). Then we have that:
\[ F_\lambda(1) = \lambda(1) \]  
\[ F_\lambda(h) = F_\lambda(h - 1) + \lambda(h) \]  
\[ F_T(1) = \min[\mu(1), \lambda(1)] \]  
\[ F_T(h) = \min[F_\lambda(h), F_T(h - 1) + \mu(h)] \]

Figures 4-7 and 4-8 show an example of how a deterministic queuing system can be represented with cumulative diagrams. The horizontal distance between the cumulative demand and throughput curves in Figure 4-8 corresponds to the delay that aircraft any, \( y \), will face as long as the queue discipline is First-Come-First-Served, while the vertical distance between the two curves corresponds to the number of aircraft waiting to be served, i.e. the length of the queue, at any instance. Thus:

\[ \text{Queue Length at time } t = L(t) = F_\lambda(t) - F_T(t) \]

\[ \text{Delay of aircraft } y = W^D(y) = F_T^{-1}(y) - F_\lambda^{-1}(y) \]

where \( F^{-1} \) is the inverse function of \( F \). In addition,

\[ \text{Average Delay per period } h = \frac{\sum_{y: (h-1)\Delta t < \tau_y \leq h\star\Delta t} W^D(y)}{\sum_{y: (h-1)\Delta t < \tau_y \leq h\star\Delta t}} \]

where \( \tau_y \) is the time when aircraft \( y \) requests service for take-off or landing.

In this deterministic model, it is assumed that the demand is evenly distributed over every time period \( h \), so that the inter-arrival times at the server are constant within \( h \). Quantity \( W^D(y) \) is used by the Delay Propagation Algorithm to assign delays to every arrival and departure of a flight at every airport in AND and subsequently to update all arrival and departure demand profiles. Hence, with reference to the AND pseudocode given in section 2.1.4 of Chapter 2, quantity \( W^D(y) \) replaces the stochastic delay estimates denoted by \( W_q(AD(f)) \) and \( W_q(AA(f)) \).
Figure 4-7: Demand and Capacity per period at a random airport.

Figure 4-8: Cumulative demand and throughput and the corresponding queue length and delay for the example shown in Figure 4-7.
Table 4.3: Four scenarios for the comparison of the Stochastic and the Deterministic AND.

<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 34 airports operate in maximum capacities</td>
<td>Northeast airports in marginal IFR capacities: BOS, EWR, JFK, LGA, PHL, DCA, IAD</td>
<td>Northeast airports in low IFR capacities</td>
<td>ORD in low IFR capacity</td>
</tr>
</tbody>
</table>

4.2.2 Comparison of the Stochastic and the Deterministic Airport Network Model

In order to compare the delays produced by the stochastic and the deterministic versions of AND, we use identical flight schedules and perform a comparison over four different capacity scenarios, as summarized in Table 4.3. Under Scenario 1 all airports operate under optimum conditions, hence at maximum capacity. For Scenario 2, six airports operate in Marginal Instrument Flight Rules (MIFR) weather, while in Scenario 3 the same six airports operate in Low IFR (LIFR) conditions. Finally, in Scenario 4, one main hub operates under extreme weather conditions and hence under LIFR and low capacity. Furthermore, we use the same values for the expected service rate of the stochastic model as for the service rate of the deterministic model for all airports in the network. The slack in the turnaround times of aircraft and the schedule padding, as discussed in Chapter 3, also remain the same for the deterministic model. The comparison is performed in terms of both network statistics and local performance measures, as defined in 2.1.5.

4.2.3 Local Comparisons

Figure 4-9 shows the average daily local arrival delay per airport in each of the four scenarios. Each point corresponds to each of the 34 airports included in the AND United States model, a description of which can be found in Chapter 3. Each airport is plotted based on the stochastic AND delay estimates on the x-axis and the deterministic on the y-axis. Each graph contains a dotted line representing the
\( y = x \) threshold, as well as a linear regression line through the 34 points (airports). To avoid confusion, it should be noted that each graph has a different scale, since each scenario creates very different delay lengths.

As expected, the delay estimates from AND for all airports fall under the \( y = x \) line, which means that the deterministic model produces lower delays than the stochastic one in all cases. In each of the four graphs in Figure 4-9, we group the airports into two categories based on the local congestion levels at each airport. For example in Figure 4-9a EWR, JFK, LGA, ATL and ORD are shown in red as the most congested airports in the network. By plotting a different linear regression line for each group of airports we notice that the higher the delay, the closer are the predictions of the deterministic and the stochastic model. In Scenarios 1 through 3 the regression line for the congested airports is always steeper than the one for the least congested airports. Table 4.4 also summarizes the gradients of all regression lines. It is also interesting to note that, as congestion increases from Scenario 1 to Scenarios 2 and 3 there is also a significant increase in gradient of the congested group from 34\% to 44\%, 70\% and 86\%, respectively, while for low congested airports it remains approximately the same.

Table 4.4: Gradient of the regression lines from Figure 4-9.

<table>
<thead>
<tr>
<th>Local Arrival Delay</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Congestion Airports</td>
<td>0.34</td>
<td>0.44</td>
<td>0.70</td>
<td>0.86</td>
</tr>
<tr>
<td>Low Congestion Airports</td>
<td>0.21</td>
<td>0.25</td>
<td>0.26</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Next we consider the daily local arrival delay profile at LGA and compare the estimates from the deterministic and the stochastic queueing engine at the end of the AND day when delay propagation and all the changes in the demand profile at LGA have been taken into account. Figure 4-10 shows the average local arrival delay per 15-minute period at LGA for each of the three scenarios. The deterministic model clearly produces lower delays than the stochastic one. However, as congestion increases the gap between the two lines becomes smaller, again, on a percentage basis.

Furthermore, the deterministic model always recovers from a peak queue faster than the stochastic. This has the following implication: if between two peaks the
queue returns to zero in the deterministic model but not in the stochastic one, at the second peak the stochastic queuing engine will start from a non-zero queue—in contrast to the deterministic one—which will result in an increase in the difference between the deterministic and stochastic estimates. This can be seen in Figure 4-10a as the gap increases from 5 minutes at the peak at 08:00 to 12 minutes at the peak of 11:00; the corresponding increase in the percentage difference between the deterministic and the stochastic estimates is from 38% to a 55%.
Figure 4-10: Average local arrival delay by 15-minute period for LGA under each of the three scenarios.
Figure 4-11: Ratio of the deterministic over the stochastic AND local arrival delays at LGA.

Generally, as shown in Figure 4-10, the pattern of delays over time produced by the deterministic and the stochastic estimations are very similar with the exception of the evening hours for Scenario 3, for which the estimates from the deterministic AND lead those from the stochastic one (Figure 4-10c). This is due to the fact that the stochastic model produces higher delays earlier in the day at LGA and therefore the propagation of delays is much more pronounced in the stochastic than in the deterministic model. This causes many flights which were initially scheduled to arrive between 18:00 and 20:00 to arrive between 21:00 and 23:00 in the stochastic model. In contrast, in the deterministic AND delay propagation is much lower and hence more aircraft fly closer to their original schedule. That leads to a lower demand during the period 18:00-20:00 and higher during the period 21:00-23:00 for the stochastic AND. As a consequence of this discrepancy, the delay gap between the stochastic and the deterministic AND becomes very small between 18:00-20:00 but increases greatly after 21:00.

In order to quantify the difference between the deterministic and the stochastic model estimates, we show in Figure 4-11 the ratio of the deterministic over the stochastic average arrival delay as a function of the day for Scenarios 1 to 3 at LGA.
The smallest difference between the two models is in Scenario 3 while the largest in Scenario 1. Actually, in Scenario 3 the deterministic delay is consistently greater than 40% of the stochastic delay and, as explained earlier, in the evening the difference drops to less than 10%.

In order to further test the observation that the deterministic model always recovers from a peak queue faster than the stochastic, we perform an additional set of tests. We run AND by replacing the two queues at LGA—one for arrivals and one for departures—with a single queue shared by arrivals and departures. This change leads to a more persistent demand, due to the fact that there is often an imbalance in demand between arrivals and departures at LGA, so that, when there are many scheduled arrivals, there are few departures and vice versa. As a result, when considering each queue separately, there are many “peaks” followed by “valleys” in the demand profiles. This permits the queue to dissipate in the deterministic case. In contrast, when modeling arrivals and departures with one shared queue the demand is more persistent, which does not leave time for the queue to dissipate completely under the deterministic model. Figure 4-12 presents the outcome of this change tested under scenario 2. What we notice now is that the queue never dissipates significantly and hence the estimates of the deterministic and the stochastic models fall much closer to each other. When changing from two queues to a single queue, the average delay throughout the day as estimated by the stochastic model falls from 21 minutes to 18 minutes, while the average delay as estimated by the deterministic model increases from 8 minutes to 13 minutes. Thus, the difference in the average delay estimates of the two models is reduced from 13 minutes in the case of two queues to only 5 minutes with the single queue.

4.2.4 Network-wide comparison

Figure 4-13 shows the average daily arrival delay per airport in each of the four scenarios. It should be noted that by “arrival delay” we refer here to the total flight arrival delay that includes the local arrival delay—due to local congestion—and the upstream arrival delay—due to delays generated elsewhere in the system. Figure 4-
Figure 4-12: Comparison of the deterministic to the stochastic and local arrival delay estimates at LGA with a single queue shared by arrivals and departures.

14 presents the total upstream delay that each airport faces throughout the day for each of the four scenarios. All graphs show the amount of delay as estimated by the stochastic and the deterministic and in the x-axis and the deterministic and in the y-axis. Again, each point on these eight figures corresponds to each of the 34 airports included in the United States model. On Figures 4-13b, 4-13c, 4-13d, 4-14b, 4-14c and 4-14d the blue points show the airports operating in optimum conditions, while the red points represent the airports operating in IFR conditions.

As with the local arrival delay per airport, the network delay estimates from for all airports also fall under the $y = x$ line which means that, as expected, the deterministic model produces lower delays than the stochastic one in all cases. Furthermore, as the level of congestion in the system increases, the difference between the delays from the stochastic and the deterministic is reduced. This becomes clear when observe the increase of the slope of the regression lines from Figure 4-13b to 4-13c and then to 4-13d as well as from figure 4-14b to 4-14c and then 4-14d. Table 4.5 shows the gradients for each of the regression lines in Figures 4-13 and 4-14.

In Scenarios 2, 3 and 4 the large delays created by the few congested airports (airports in IFR conditions) are propagated to the rest of the network. These up-
stream delays become the primary source of arrival delay at all other airports with low congestion—and low local delays—in AND's network. As already described, the difference between the local delay estimates of the deterministic and the stochastic queuing engines is small at the congested airports. Then, for Scenarios 2, 3 and 4 we should also expect that, the difference between the estimates of the deterministic and the stochastic AND of the average total arrival delay at the low-congested airports should also be small, since most of the total arrival delay at these airports is attributed to the congested ones. This is shown in Figures 4-13 and 4-14 where, in each of the Scenarios, a single regression line fits all the data points well (in contrast to Figure 4-9 where it was necessary to separate airports according to their local congestion level and plot separate regression lines.)

Table 4.5: Gradient of the regression lines from Figures 4-13 and 4-14.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Arrival Delay</th>
<th>Scenario</th>
<th>Arrival Delay</th>
<th>Scenario</th>
<th>Arrival Delay</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.31</td>
<td>2</td>
<td>0.44</td>
<td>3</td>
<td>0.62</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>0.16</td>
<td></td>
<td></td>
<td>3</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

Concluding, we observed that there is a significant reduction in expected delays as one moves from a stochastic National Airspace System to a completely deterministic one. By increasing the predictability of the system through the implementation of 4D trajectories, we have shown examples in which the savings in delays due to airport congestion may be as high as 70% at optimum weather conditions across the country and more than 30% when several airports operate under heavy congestion. These estimates ignore any capacity benefits that may result from the 4D trajectory program, but, at the same time, assume that demand is evenly spread within each subperiod.

4.3 Summary

In this chapter we described two important extensions implemented in the AND model. First, we presented in detail an algorithm which was implemented in AND in order to model Ground Delay Programs (GDP), a central coordination tool of the
Air Traffic Management System. The GDP algorithm developed for AND consists of three main parts: a) GDP initiation, b) slot assignment and c) deciding whether to terminate or continue the GDP. The slot assignment part includes the assignment of Expected Arrival Control Times and Expected Departure Control Times to all affected flights, slot swapping between flights of the same airline and schedule compression. The algorithm contains a priority queue in order to store flights, in the form of a binary tree. This heap structure makes the operation of adding a flight in the
Figure 4-14: Total upstream delay by airport for each of the four scenarios, including a linear regression line.

queue or removing the earliest scheduled flight from the queue very efficient.

We have validated the GDP algorithm against observed data obtained by ASPM [19] for a day when ORD was operating under low IFR conditions and was issuing arrival slots for most of the day. In addition to expanding the AND’s capabilities, the GDP algorithm improves the accuracy of AND’s delay estimates in relevant cases.

The second extension presented in this chapter was an alternative deterministic queuing engine to estimate delays. We implemented a queuing engine with a non-
stationary and deterministic demand and service rates \((D(t)/D(t)/1)\). We compared the results of the deterministic to the stochastic, \(M(t)/E_k(t)/1\), queuing models at both the local and the network level, using AND. We showed that as delays increase, the estimates of the queuing models get closer. We also provided insights into how delay propagation affects the comparison between the deterministic and stochastic AND models. Finally, we provided an example of how the “stochastic” and the “deterministic” versions of AND may be used to provide approximate estimates of the benefits obtainable from increased predictability of aircraft trajectories and processing times—both of which count among the principal objectives of the NextGen.
Chapter 5

On the Effect of Slot Control at Busy US Airports

As described in Chapter 2, AND is a macroscopic and fast, stochastic and dynamic model of a network of major airports, designed as a tool for assessing the network-wide impacts of policy changes and local or regional infrastructure improvements. One important set of relevant policies is concerned with potential demand management measures, often referred to as slot controls. Slot controls and slot allocation schemes have been introduced over the years at busy airports around the world in order to mitigate and control congestion. In this chapter we start by summarizing current practices in this respect. We then perform an analysis of the relationship between the scheduled demand at Newark’s Liberty International Airport (EWR) and the airport’s capacity under different runway configurations and weather conditions. This analysis underlines the importance of potentially reducing the number of slots available to airlines at EWR. We present next an Integer optimization model that generates a new flight schedule when the number of available slots at one or multiple airports is reduced, while respecting aircraft connectivity and passenger connectivity constraints. Finally, we use the AND model and the optimization model to test the effect of several flight schedules, produced under various slot limits, on local and network-wide delays for several different capacity conditions.
5.1 The Slot Control System

Slot control is one of the most common international practices in air transportation, for mitigating and controlling congestion at airports. Most major airports in Europe and Asia place limits on the number of scheduled operations per unit of time. Each slot-controlled airport, in coordination with the local aviation authorities allocates slots to airlines according to the airport’s declared capacity. de Neufville and Odoni [12] define declared capacity “as the number of aircraft movements per hour that an airport can accommodate at a reasonable Level-of-Service”, where delay is the measure of level-of-service. Typically, European airports, set slot limits for arrivals and departures separately, for every hour of a day, and often for every 15-, 10- or 5-minute period, as well. Since ATC procedures at European airports rely on Instrument Flight Rules (IFR) even on good weather days, the slot limits are usually set near the IFR capacities of the airports.

The International Air Transport Association (IATA) has developed a slot allocation system that is used worldwide to coordinate airports and airlines during the slot assignment procedure. Slot allocation takes place twice a year, once for the summer and once for the winter schedules of airports. The IATA system applies an extensive set of rules that guide the assignment of slots. These are designed to ensure that an airline can operate services in the long term out of an airport. At the same time, they make slots available to “new entrant” airlines, so that barriers to entry may be reduced.

In contrast to the rest of world, most airports in the United States do not utilize slot controls. Airlines are essentially allowed to schedule arrivals and departures at any time they wish at practically all the airports. Until 2007, the Federal Aviation Administration (FAA) exercised slot controls at only four busy US airports under the High Density Rule (HDR), namely ORD, JFK, LGA and DCA. After the expiration in 2007 of the HDR (with the exception of DCA), under Congressional mandate, the FAA issued in the summer of 2008 new rules imposing constraints on the number of the slots available at JFK, LGA, and EWR airports. However, the enforcement of
the slot controls at these three airports is still quite loose, primarily because traffic there has dropped since 2008, following the overall US trend.

5.2 Analysis of demand and capacity at Newark Liberty Int’l Airport

5.2.1 Newark Liberty Int’l Airport

EWR is one of the most congested airports in the United States and has been operating under limited slot coordination since the summer of 2008. According to FAA’s operational rules [17] the number of slots at EWR is limited to 81 in any 60-minute period and 44 in any 30-minute period for scheduled operations, while 2 additional operations per hour are allowed for non-scheduled flights before noon, and 1 per hour in the afternoon. EWR, as we will show later, has a very unbalanced schedule when it comes to the number of arrivals and the number of departures scheduled by hour of day. The FAA’s current rules do not specify separate limits for the number of arrivals and departures that can be scheduled at EWR.

Figure 5-1 shows the runway layout of EWR, which consists of a pair of long closely-spaced parallel runways (denoted as 22R/4L and 22L/4R) and one intersecting short runway (11/29). As indicated in Table 5.1 EWR operates primarily in 4 configurations (87% of the time) when the weather conditions allow Visual Flight Rules (VFR) and also has 2 predominant configurations when operating under Instrument Flight Rules (IFR)—78% of the time. When operating under configurations 22L|22R and 4R|4L, runway 22L/4R is used for departures and 22R/4L for arrivals. Under configuration 11, 22L|22R runways 11, 22L are used for arrivals and 22R for departures, hence providing a higher arrival throughput.

Figure 5-2 shows the percentage utilization of the four most frequent runway configurations of EWR by hour of the day. Until 7:00am the predominant configuration is 22L|22R while from 7:00am to 12:00pm 22L|22R and 4R|4L are used for roughly the same amount of time. In the early afternoon, when there is a large number of
Table 5.1: Runway configuration frequency of utilization at EWR under VFR and IFR conditions.

<table>
<thead>
<tr>
<th>Runway Configuration</th>
<th>VFR Frequency</th>
<th>VFR Cumulative Frequency</th>
<th>IFR Frequency</th>
<th>IFR Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>22L/22R</td>
<td>37%</td>
<td>37%</td>
<td>38%</td>
<td>38%</td>
</tr>
<tr>
<td>4R/4L</td>
<td>21%</td>
<td>58%</td>
<td>40%</td>
<td>78%</td>
</tr>
<tr>
<td>11, 22L/22R</td>
<td>19%</td>
<td>77%</td>
<td>8%</td>
<td>87%</td>
</tr>
<tr>
<td>4R, 11/4L</td>
<td>11%</td>
<td>87%</td>
<td>8%</td>
<td>95%</td>
</tr>
</tbody>
</table>

scheduled arrivals, EWR utilizes configuration 11, 22L/22R in order to increase the arrival throughput.

Figure 5-3 shows the average number of scheduled operations per hour at EWR in 2007. We also plot the average number of scheduled movements for weekdays only, as
Figure 5-2: Utilization of the main runway configurations by hour of day at EWR.

well as for Thursdays only, the busiest day of the week at EWR. The bold line shows the slot limits as introduced in 2008 at the level of 81 operations per hour. We can see that, on average, on weekdays the limit of 81 is exceeded only during three hours – from 16:00 to 19:00. Furthermore, the hourly slot limit was exceeded during at most 5 hours per day, if we consider the moving average number of scheduled operations for Thursdays only.

Figure 5-3: Hourly scheduled operations at EWR in 2007; the bold line shows the hourly slot limit introduced in 2008. [note: TR, S and SN stand for Thursday, Saturday and Sunday respectively]
5.2.2 Capacity and Schedule at EWR

On the Topic of Capacity Estimation

In this section we perform a detailed comparison of the scheduled demand for arrivals and departures at EWR with the capacity of the airport, in order to better understand the problem of congestion at that airport. We first introduce two metrics of airport capacity that we will use throughout this chapter. Both of these metrics, which refer to the capacity of the runway system, are two-dimensional vectors that consist of a value for arrival movements and a value for departure movement. They are measured in terms of aircraft per unit of time—typically chosen to be 15 minutes or one hour long:

Expected capacity is defined as the expected (average) number of movements—a combination of arrivals and departures—that an airport can perform in a given runway configuration, conditioned on the fact that there is persistent demand by aircraft requiring service. Persistent demand exists during periods of time in which the queue of aircraft waiting to land or to takeoff is continually nonzero and the runways have to operate at their full potential. If the service rate of the airport under the aforementioned conditions has a mean value \( \mu \) then the expected capacity is given by \( E[C] = \frac{1}{\mu} \).

Maximum feasible capacity is defined as the maximum number of movements that can be served during some given interval of time by an airport with a given runway configuration, under favorable conditions. The maximum feasible capacity may not be sustainable over time.

Obviously, the maximum feasible capacity is always at least as large as the expected capacity of the airport. Note that each of these two capacity metrics is associated with a specific runway configuration and set of weather conditions, which in turn affect the ATC operating procedures in use.

The most common and simple method for estimating the maximum feasible capacity of an airport is through the capacity envelope of the airport, first introduced by Gilbo in 1993 [21]. Let us define the throughput vector as consisting of two elements:
the count of arrivals and the count of departures that an airport served in a given period of time (typically 15 minutes or 60 minutes). If we draw all observations of such vectors over a period of time (typically a year or a month) on a two-dimensional graph, with the number of arrivals on the x-axis and of departures on the y-axis, then the convex hull of all such observations forms the capacity envelope of the airport. The airport has sufficient capacity to serve any combination of arrival and departure demands that falls within the capacity envelope and cannot serve any combination that falls outside that envelope. The maximum feasible capacity, as defined in the previous paragraph is given by this airport capacity envelope. Usually, in order to account for noise in the data that may skew significantly the capacity envelope, outlier points are identified according to their low frequency of occurrence and are ignored. Figure 5-4a shows all the observed vectors when EWR was operating in configuration 4R|4L under VFR conditions in 2007 and the green line shows the corresponding capacity envelope.

The problem that arises from this approach for estimating the capacity of an airport is that it provides just the “optimum” airport capacity under a given runway configuration and weather conditions. Typically this is only achievable under very specific conditions of aircraft mix, mix of arrivals and departures on different runways, optimum sequencing of movements, favorable wind direction and wind speed and other such factors that affect the airport throughput, but are not explicit at this level of aggregation.

Overall, the maximum feasible capacity gives an overly optimistic picture of the “true” capacity of an airport. From a strategic planning perspective the use of maximum feasible capacity, which is a widely used capacity metric in the air transportation sector, could have undesired effects. For example, if the number of arrival and departure slots that are made available to the airlines per hour or 15 minutes is determined using the maximum feasible capacity of the airport, then, delays than are much higher than can be tolerated may result.

The expected capacity metric leads to more realistic estimates of the sustainable airport throughput and is thus more useful for such strategic decisions. Simaiakis
[47], [49] has proposed a way to account for persistent demand in the departure side of the capacity estimation. The relevant methodology was described in Section 3.2. In summary, as a first step, the departure throughput is estimated as a function of the number of aircraft on the ground. The threshold number of aircraft on the ground, above which the departure throughput rate remains constant, is then determined. This threshold number guarantees that there is continuous departure demand at the runway threshold. The airport departure capacity is then determined by calculating the departure throughput as a function of the arrival throughput. To do so, Simaiakis retains in the analysis only the throughput vectors that meet the continuous demand condition. Finally, a least square function is fitted to the plot of departure throughput versus the arrival throughput.

**EWR Analysis**

We start our analysis by providing capacity estimates for EWR under different configurations and weather conditions. In Figure 5-4 we present a scatter plot of all pairs of observed counts of arrivals and departures in 15 minutes for two of the most common configurations of EWR under VFR conditions and the most frequent configuration under IFR conditions in 2007. Each circle in the scatter plots corresponds to an observed count of arrivals and departures (throughput vector) on a 15-minute period, while the radius of each circle is proportional to the frequency of each observation. Each graph also shows the expected capacity for that given runway configuration and weather conditions as estimated by Simaiakis [48] in the manner described in the previous section. Furthermore, for the sake of comparison of the two capacity metrics—the maximum feasible capacity and the expected capacity—, we plot the capacity envelope of EWR for configuration 4R|4L (Figure 5-4a). Clearly, the maximum feasible capacity produces a much higher capacity metric than the expected capacity metric, the difference being of the order of 2 additional departures per 15 minutes. This overestimation of capacity may lead to undesirable levels of delay if it is used for planning purposes of a slot allocation policy. Comparing Figures 5-4a and 5-4b we also observe that the expected capacity of EWR drops by almost one
departure and one arrival per 15 minutes when the weather deteriorates from VMC to IMC.

![Scatter plots of observed arrival and departure counts per 15 minutes at EWR in 2007; red line indicates the expected capacity.](image)

Figure 5-4: Scatter plots of observed arrival and departure counts per 15 minutes at EWR in 2007; red line indicates the expected capacity [48].

Figure 5-5 shows the counts of scheduled arrivals and departures during the periods when the two most common runway configurations—4R|4L and 11, 22L|22R—are used (in contrast to Figure 5-4 where we presented the observed throughput of the airport in terms of the actual count of arrivals and departures). Clearly, the scatter plots for the scheduled operations at EWR are much more widely spread than the
throughput plots (Figure 5-4), reaching as high as 21 departures and 22 arrivals per quarter. Furthermore, by juxtaposing the scatter plots of the scheduled operations versus the expected capacity plots of the two configurations, one can immediately see that EWR is frequently over-scheduled. In fact, Figure 5-5a shows that for 30% of the time, scheduled operations fall outside the expected capacity line for VFR conditions, while in Figure 5-5b scheduled operations fall outside the expected capacity line 39% of the time. In total, for 33% of the time when one of these two runway configurations is used EWR is expecting more operations than it can actually handle. When comparing the scheduled operations to the IFR capacity of the airport, the percent of the time when the number of scheduled operations exceeds the capacity rises to 43%. Finally, in Figure 5-5 we notice that, on average, configuration 11, 22L|22R is operated when more arrivals than departures are scheduled since it permits a higher arrival throughput, with two runways used for arrivals. Actually, when configuration 4R|4L is used there are on average 7.1 scheduled arrivals per quarter, while under configuration 11, 22L|22R there are on average 8.8 scheduled arrivals per quarter.

Figure 5-6 shows the cumulative frequency of scheduled operations per hour at EWR in 2007. It is observed that when aggregating the schedule on an hourly basis, the scheduled operations do not exceed the expected capacity or the current slot limits as often as when aggregating on a 15-minute basis or when considering arrivals and departures separately. In fact, only 15% of the time is the number of scheduled operations greater than the limit on total slots (> 81 operations per hour) and only 10% of the time greater than the expected hourly capacity (> 84 operations per hour).

From the analysis presented so far we have shown that EWR was over-scheduled for at least 30% of the year, depending on the runway configuration in operation and the weather conditions. In order to determine the level of congestion at EWR during different times of the day, we extend our analysis to the daily profiles of the schedule of movements. Figure 5-7 shows the yearly average scheduled departure and arrival daily profile, in 15-minute intervals. In Figure 5-7a we also plot the expected departure capacity based on the scheduled arrival demand at each 15-minute
Figure 5-5: Scatter plots of scheduled arrival and departure counts per 15 minutes at EWR in 2007; lines indicate the expected capacity [48].

Figure 5-6: Cumulative frequency of scheduled operations counts per hour at EWR in 2007; lines indicate the expected capacity and the current slot limit.

period. The departure capacity is obtained from Figure 5-7a as a function of arrivals: under the assumption that arrivals have priority over departures, we input to the expected departure capacity function of Figure 5-4a the number of scheduled arrivals per quarter from Figure 5-7b, to obtain a value for the departure capacity for every
15 minutes as shown in Figure 5-7a.

The shaded area in Figure 5-7b represents the range of arrival capacities that is expected at EWR depending on the weather conditions and departure traffic—the lowest arrival capacity occurs at IFR conditions and high departure demand and the highest at VFR conditions and low departure demand. Observing these plots it becomes evident that EWR, very often during the day, has a scheduled demand that exceeds the capacity by as much as 50%. For example, at approximately 08:30, the yearly average number of scheduled departures is almost 17, while the expected capacity is 11 departures per quarter. From 06:00 to 09:00 the scheduled departure demand is constantly above the VFR capacity of the airport. Moreover, from 16:30 to 20:15 there are many more scheduled departures than the airport can handle, especially when considering that during the same time period there is always a high arrival demand, which results in a reduced departure capacity. Similarly, depending on the conditions (weather and departure demand), at EWR the scheduled arrivals exceed the expected capacity very often during the day. Actually there exist periods in the early afternoon when the capacity is exceeded, on average, every day even under favorable conditions.

**Slot limits at EWR**

The main takeaway from the analysis presented so far is twofold. First, in order to have a meaningful slot control system at EWR slot limits should be set on a 15-minute basis rather than the existing hourly slot limits. As shown in Figure 5-3 the aggregate scheduled demand rarely exceeds the capacity of the airport on an hourly basis. In contrast, as shown in Figures 5-5 and 5-7 the capacity is very often exceeded on a 15-minute basis. In addition, there should be separate slot limits for arrivals and separate for departures, rather than just limits on the total number of operations which are currently being enforced. This is due to the fact that the schedule at EWR is very unbalanced, in the sense that, in the morning, there are periods with many departures and few arrivals, and the departure demand is almost constantly above the capacity, even though the aggregate capacity is not exceeded. An analogous
Figure 5-7: Average departure and arrival schedule daily profile in 2007 at EWR. [note: TR stands for Thursday]

observation applies to the arrival demand in the early afternoon.

The second takeaway is that the expected capacity metric is more adequate for planning purposes than the maximum feasible capacity metric. It was shown that the maximum feasible capacity at EWR can be as high as 13 departures and 12 arrivals per quarter. However, this number can only be achieved under very favorable conditions and is rarely observed even though there is often sufficient demand to match this number. Thus, setting slot limits with reference to the maximum feasible capacity may not result in any reduction of delays at EWR.

We shall use these takeaways in Sections 5.3.7 and 5.5 where the application of slot control at EWR is tested.
5.3 The Demand Smoothing Model

In this section we introduce an integer optimization model in order to test the effect of applying slot constraints at some busy airports in the United States. The main idea behind the model is the following: given an initial schedule that has been created as a result of the airlines’ scheduling choices in the absence of any slot constraints, the model generates a feasible schedule of arrivals and departures that obeys the slot limits that are imposed at an airport. In other words, starting from the current OAG schedule of flights at an airport, which, as we saw for EWR, may lead to over-scheduling in many hours during the day, the optimization model will modify the original schedule to produce a “smoother” one without reducing total aircraft and passenger flows. With such a schedule in hand it is then possible to test the effects of applying different levels of slot constraints at congested airports on both local delays and network-wide delays using the Airport Network Model (AND) model. As an example, it will then be possible to test the effect of applying slot limits consistent with the IFR capacity of an airport versus limits consistent with the VFR capacity.

The main assumption in our analysis is that the network structure of each airline’s flights will not change dramatically due the introduction of slot constraints at a few airports. Especially from a macroscopic point of view, the assumption that aircraft will fly the same itineraries as before but at slightly different times is realistic given that our goal is to estimate aggregate delay savings per airport and average delays per hour of day, and not to focus on delays to individual flights.

Apart from testing the effect of slot limits on delays, the model to be presented is useful for observing the complex interactions in an airport network when the schedule is smoothed, particularly at hub airports. In theory, we expect that a smooth demand will lead to lower local delays due to the non-linear relationship between delays and the demand-to-capacity ratio (for details refer to Chapter 2). However, due to the delay propagation taking place at hub airports, as discussed in Chapter 3, the realized demand, especially in the late afternoon, might not remain smooth in the least. Delay propagation on busy days at hub airports may result in “smoother” demand profiles.
towards the end of the day, as shown for Chicago O'Hare airport in Chapter 3 (see Figure 3-21a). Hence, it is also possible that the reverse may take place, i.e., that smoothing the demand in advance will lead to a more uneven realized demand than the one under the original schedule.

Our demand-smoothing optimization model differs from the tactical intervention into airline schedules that typically occurs when Ground Delay Programs are in effect. Under a GDP, as described in Chapter 4, an airport will issue arrival slots tactically according to a First-Scheduled-First-Served discipline just a few hours before the departure of each flight. The difference between the demand-smoothing model to the tactical intervention is then primarily twofold. First, the model does not assign slots on a First-Scheduled-First-Served basis, but rather optimizes the schedule to minimize overall schedule displacement, while respecting passenger and aircraft connections. This means that if the first flight of an aircraft is displaced then all its subsequent flights will be displaced as well. Second the model may reschedule flights to times earlier than in the original schedule.

We proceed by first introducing the inputs, the parameters and the variables of the optimization model. We, then, describe the objectives and the constraints and finally we present the model development, the computation of the solution and the experimental setup for EWR.

5.3.1 Input

To begin we determine the set of airports \( A := \{1, \ldots, K + 1\} \) where \( K \) is the total number of airports at which we wish to apply slot constraints for arrivals and departures. The \( (K + 1)^{th} \) airport corresponds to all airports where we do not apply any slot constraints. Furthermore, we divide the day into \( N \) periods of constant length \( h \) (typically \( h = 15 \text{ minutes} \)) so that the set of all periods is \( P := \{1, \ldots, N\} \). The model requires as an input the set of aircraft \( AC \) that fly at least once during the period \( \{1, \ldots, N\} \) through at least one of the airports in \( \{1, \ldots, K\} \). The set of flights \( F := \{1, \ldots, f\} \) is determined by the aircraft set \( AC \). Every flight operated on that day by each of the aircraft in \( AC \) is an element of \( F \). Hence a subset of
F contains all the flights arriving at or departing from any of the K airports while another subset contains flights that do not operate at any of the K airports. Clearly the two subsets are mutually exclusive and their union defines F. As an example consider an aircraft, flying the itinerary A → B → C → A and suppose we apply slot constrains at airport A. Flights A → B and C → A are directly involved with airport A but flight B → C is also included in the model. Finally, the earliest departure in F defines the first period in N, while the last arrival in F defines the last period in N.

5.3.2 Parameters

Every flight i ∈ F is associated with four parameters: the departure period and the departure airport, and the arrival period and arrival airport, which we map into two sets of binary parameters. Hence we define ∀i ∈ F, k ∈ A, n ∈ P:

$S_{ink}^{dep} = \begin{cases} 1, & \text{iff } i \text{ is scheduled to depart from airport } k \text{ during period } n \\ 0, & \text{otherwise} \end{cases}$

$S_{ink}^{arr} = \begin{cases} 1, & \text{iff } i \text{ is scheduled to arrive at airport } k \text{ during period } n \\ 0, & \text{otherwise} \end{cases}$

Furthermore for every airport k ∈ {1, . . . , K} we predefine the number of arrival slots, of departure slots and the total number of slots per period n ∈ P:

$C_{nk}^{dep} = \text{departure slots at } k \text{ in period } n, \forall k \in \{1, \ldots, K\}, n \in P$

$C_{nk}^{arr} = \text{arrival slots at } k \text{ in period } n, \forall k \in \{1, \ldots, K\}, n \in P$

$C_{nk}^{T} = \text{total number of slots at } k \text{ in period } n, \forall k \in \{1, \ldots, K\}, n \in P$

We, also, define parameters that determine if pairs of sequential flights are flown by the same aircraft. We input the minimum connection time between such pair of flights using parameter $\beta_{ij}$ ("minimum turnaround time" as defined in Chapter 3).
Hence, we define \( \forall i, j \in F, i \neq j: \)

\[

z_{ij} = \begin{cases} 
1, & \text{iff } i, j \text{ are flown by the same aircraft and } j \text{ is the immediate successor flight of } i \\
0, & \text{otherwise}
\end{cases}
\]

\[

\beta_{ij} = \begin{cases} 
\geq 1, & \text{iff } z_{ij} = 1 \\
0, & \text{otherwise}
\end{cases}
\]

Clearly, \( z_{ij} \) is a binary parameter, while \( \beta_{ij} \in \mathbb{I}^+ \) so that the minimum turnaround time between two flights \( i, j \) is given by \( \beta_{ij} \times h. \)

Similarly we use parameters to define if there are passengers connecting between flights, so that \( \forall i, j \in F, i \neq j: \)

\[

pax_{ij} = \begin{cases} 
1, & \text{iff there is at least 1 passenger connecting from flight } i \text{ to flight } j \\
0, & \text{otherwise}
\end{cases}
\]

Finally, parameter \( \alpha_k \forall k \in \{1 \ldots K\} \) is the minimum time required by any connecting passenger to transfer from the first to the second flight in his itinerary. \( \alpha_k \) is airport-dependent as transfer times depend on the size and layout of each airport. The International Air Transport Association (IATA) provides standards for the minimum connecting time at every airport [26].
5.3.3 Variables

In the optimization model we use two binary variables \( x_{ink}, y_{ink} \in [0, 1], \forall i \in F, n \in P, k \in A \) and one integer variable \( u_i \in \mathbb{I}^+, \forall i \in F \):

\[
x_{ink} = \begin{cases} 
1, & \text{if flight } i \text{ is given a departure time during period } n \text{ from airport } k \\
0, & \text{otherwise}
\end{cases}
\]

\[
y_{ink} = \begin{cases} 
1, & \text{if flight } i \text{ is given an arrival time during period } n \text{ at airport } k \\
0, & \text{otherwise}
\end{cases}
\]

\( u_i = \) number of periods that flight \( i \) was displaced by,

5.3.4 Objective

The objective in the demand-smoothing optimization model is to minimize the maximum displacement that any given flight will sustain in the process of satisfying the slot constraints. However, there may be multiple feasible solutions under this objective, so we introduce a second objective, to minimize the total displacement over all flights \( i \in F \).

Hence the objective function is to minimize:

\[
\lambda \max_{i \in F} |u_i| + \sum_{i \in F} |u_i|
\]

where parameter \( \lambda >> 1 \) in order to assign more weight to the maximum displacement than to the total displacement across all flights. If \( \lambda \) is chosen large enough—e.g. \( \lambda > \sum_{i \in F} |u_i| \)—then the objective is equivalent to solving two optimization problems sequentially: in the first we minimize only the maximum displacement, \( \max_{i \in F} |u_i| \), and in the second we minimize the total displacement, \( \sum_{i \in F} |u_i| \), while constraining the displacement of each flight to be less than or equal to the first objective function's solution.
5.3.5 Constraints

Flight constraints

The first set of constraints provides the relationships between the three variables. We constrain the flight time between the origin and the destination of each flight to remain the same. In other words, the departure time and the arrival time of any given flight are always displaced by the same amount $u_i$. Hence we write that:

$$\sum_{k \in A} \sum_{n \in P} x_{inkn} = \sum_{k \in A} \sum_{n \in P} S_{inkn}^{\text{dep}} + u_i, \quad \forall i \in F$$

$$\sum_{k \in A} \sum_{n \in P} y_{inkn} = \sum_{k \in A} \sum_{n \in P} S_{inkn}^{\text{arr}} + u_i, \quad \forall i \in F$$

Moreover, we ensure that every flight is assigned a new departure time and a new arrival time, even if these coincide with the original scheduled times. So we have that:

$$\sum_{k \in A} \sum_{n \in P} x_{ink} = 1, \quad \forall i \in F$$

$$\sum_{k \in A} \sum_{n \in P} y_{ink} = 1, \quad \forall i \in F$$

Slot constraints

The second set of constraints relates to the slot limits that we wish to apply to the $K$ airports. These constraints ensure that at no period $n \in P$ the scheduled departures and the scheduled arrivals will exceed the departure and arrival slot limits, respectively, at each of the $K$ slot-constrained airports. Furthermore, we introduce a
constrain that limits the sum of arrivals and departures in any period $n$:

\[
\sum_{i \in F} x_{nk} \leq C_{nk}^{\text{dep}}, \quad \forall n \in P, \, k \in A
\]
\[
\sum_{i \in F} y_{nk} \leq C_{nk}^{\text{arr}}, \quad \forall n \in P, \, k \in A
\]
\[
\sum_{i \in F} x_{nk} + \sum_{i \in F} y_{nk} \leq C_{nk}^T, \quad \forall n \in P, \, k \in A
\]

**Aircraft connectivity constraints**

The next set of constraints ensures that aircraft will fly the same itineraries after the new schedule is created as they did under the original schedule. This is achieved by connecting pairs of sequential flights that are operated by the same aircraft. Hence, if a morning flight of an aircraft into a slot-controlled airport is moved to an earlier period then the afternoon flight of that aircraft will also have to be moved. These constraints increase the complexity of the model since there are many aircraft that visit busy airports multiple times in a day. We also wish to test the effect of reducing the turnaround time between two flights operated by the same aircraft. To this effect, we introduce two different sets of constraints for two cases: a) keep the ground turnaround time between the two flights the same; and b) allow the turnaround time to be shorter than originally scheduled but greater than a minimum value that depends on the type of aircraft, the airline and the type of airport (hub or local airport). Hence we write the following two sets of constraints that we shall utilize in
separate optimization models:

\[ a) \quad (\sum_{k \in A} \sum_{n \in P} x_{ink} n - \sum_{k \in A} \sum_{n \in P} y_{jnk} n) z_{ij} = (\sum_{k \in A} \sum_{n \in P} S^{dep}_{ink} n - \sum_{k \in A} \sum_{n \in P} S^{arr}_{jnk} n) z_{ij}, \]

\[ \forall i, j \in F, i \neq j \]

\[ b) \quad (\sum_{k \in A} \sum_{n \in P} x_{ink} n - \sum_{k \in A} \sum_{n \in P} y_{jnk} n) z_{ij} \leq (\sum_{k \in A} \sum_{n \in P} S^{dep}_{ink} n - \sum_{k \in A} \sum_{n \in P} S^{arr}_{jnk} n) z_{ij}, \]

\[ \forall i, j \in F, i \neq j \]

\[ (\sum_{k \in A} \sum_{n \in P} x_{ink} n - \sum_{k \in A} \sum_{n \in P} y_{jnk} n) z_{ij} \geq \beta_{ij}, \quad \forall i, j \in F, i \neq j \]

**Passenger connectivity constraints**

The final set of constraints ensures that all passenger connections can be achieved under the new schedule at each of the \( K \) airports where we apply slot constraints. In other words, any passenger who is originating from any airport \( A \) and destined to any airport \( C \), but connecting through \( B \) where \( B \in \{1, \ldots, K\} \), will be able to fly the exact same itinerary \{A-B-C\} operated by the same two flights. This ensures that, if we apply slot constraints, especially at a hub airport, the banks of the “hubbing” airline will not be affected: the time when the banks occur might be shifted, but all connections within the bank will be respected. We allow however, for the connection time to change, but never below a threshold \( \alpha_k \) in order to allow for adequate transfer time. Hence we write:

\[ (\sum_{k \in A} \sum_{n \in P} n x_{ink} n - \sum_{k \in A} \sum_{n \in P} y_{jnk} n) p a x_{ij} \geq p a x_{ij} \alpha_k \quad \forall i, j \in F, i \neq j, k \in \{1, \ldots, K\} \]
Formulation

In summary the optimization model is formulated as follows:

\[
\begin{align*}
\text{minimize} & \quad \lambda \max_{i \in F} |u_i| + \sum_{i \in F} |u_i| \\
\text{subject to:} & \quad \sum_{k \in A \land n \in P} x_{ink} n = \sum_{k \in A \land n \in P} s_{ink}^{\text{dep}} + u_i, \quad \forall i \in F \\
& \quad \sum_{k \in A \land n \in P} y_{ink} n = \sum_{k \in A \land n \in P} s_{ink}^{\text{arr}} + u_i, \quad \forall i \in F \\
& \quad \sum_{k \in A \land n \in P} x_{ink} = 1, \quad \forall i \in F \\
& \quad \sum_{k \in A \land n \in P} y_{ink} = 1, \quad \forall i \in F \\
& \quad \sum_{i \in F} x_{ink} \leq c_{nk}^{\text{dep}}, \quad \forall n \in P, k \in A \\
& \quad \sum_{i \in F} y_{ink} \leq c_{nk}^{\text{arr}}, \quad \forall n \in P, k \in A \\
& \quad \sum_{i \in F} x_{ink} + \sum_{i \in F} y_{ink} \leq c_{nk}^T, \quad \forall n \in P, k \in A \\
& \quad \left(\sum_{k \in A \land n \in P} x_{ink} n - \sum_{k \in A \land n \in P} y_{jnk} n\right) z_{ij} \leq \left(\sum_{k \in A \land n \in P} s_{ink}^{\text{dep}} - \sum_{k \in A \land n \in P} s_{jnk}^{\text{arr}}\right) z_{ij}, \\
& \qquad \forall i, j \in F, i \neq j \\
& \quad \left(\sum_{k \in A \land n \in P} x_{ink} n - \sum_{k \in A \land n \in P} y_{jnk} n\right) z_{ij} \geq \beta_{ij}, \quad \forall i, j \in F, i \neq j \\
& \quad \left(\sum_{k \in A \land n \in P} n x_{ink} - \sum_{k \in A \land n \in P} y_{jnk} n\right) p a x_{ij} \geq p a x_{ij} \alpha_k, \quad \forall i, j \in F, i \neq j, k \in \{1, \ldots, K\} \\
\end{align*}
\]

\[x_{ink} \in [0, 1] \]

\[y_{ink} \in [0, 1]\]

5.3.6 Model Development and Computation of Model Solution

The objective function introduced in the previous section is clearly non-linear but can be made linear through a simple transformation. We introduce variables \( w \) and
\(d_i\) and two additional sets of constraints:

\[
\begin{align*}
    u_i &\leq w \quad \forall i \in F \\
    -u_i &\leq w \quad \forall i \in F
\end{align*}
\]

and

\[
\begin{align*}
    u_i &\leq d_i \quad \forall i \in F \\
    -u_i &\leq d_i \quad \forall i \in F
\end{align*}
\]

Then the objective function becomes:

\[
\text{minimize} \quad \lambda w + \sum_{i \in F} d_i
\]

Even though the model developed is an integer program (IP), which is NP-hard in general, the problem size has been kept small to eliminate computational issues. As defined the IP has \(O(f^2)\) constraints and \(O(f)\) variables. This means that with one slot constrained airport \((K = 1)\), \(F\) contains approximately 2000 flights (depending on the size of the airport), hence 2000 variables and 4 million constraints. However we can reformulate the problem to reduce the size of the aircraft connectivity constraints.

Sorting the flights in the set \(F\) according to their aircraft tail number and then according to their scheduled departure time we guarantee that a random flight \(i \in F\) can only be connected to flight \(i - 1\). Hence we can change the parameters \(z_{ij}\) and \(\beta_{ij}\) to \(z_i\) and \(\beta_i\) respectively, to reflect whether flight \(i\) is connected to flight \(i - 1\). So the aircraft connectivity constraints become:

\[
\begin{align*}
    \left(\sum_{k \in A} \sum_{n \in P} x_{ink} - \sum_{k \in A} \sum_{n \in P} y_{(i-1)nk}\right)z_i &\leq \left(\sum_{k \in A} \sum_{n \in P} S_{ink}^{\text{dep}} - \sum_{k \in A} \sum_{n \in P} S_{(i-1)nk}^{\text{arr}}\right)z_i, \\
    \forall i &\in \{2, \ldots, f\}
\end{align*}
\]

\[
\begin{align*}
    \left(\sum_{k \in A} \sum_{n \in P} x_{ink} - \sum_{k \in A} \sum_{n \in P} y_{(i-1)nk}\right)z_i &\geq \beta_i, \quad \forall i \in \{2, \ldots, f\}
\end{align*}
\]

As described below, we have tested the model with and without the passenger
connectivity constraints. By ignoring passenger constraints the optimization problem has only \(O(f)\) constraints—instead of \(O(f^2)\)—when reformulating the aircraft connectivity constraints as described above. Then, in the case of \(K = 1\) and \(f = 2000\) there are 2000 constraints. With this transformation, we were able to reduce the computational time in CPLEX from a few hours down to approximately 30 minutes. For \(K = 2\) the number of variables and the number of constraints will nearly double, but as \(K\) increases the the rate at which the number of the variables and aircraft connectivity constraints will decrease. The reasoning behind this is that when increasing the slot controlled airports from \(k\) to \(k + 1\) many of the aircraft included in the optimization for the \(k\) airports may also fly at some point in the day through the \((k + 1)^{th}\) airport. That means that the set of aircraft corresponding to the airports \(\{1, \ldots, k\}\) and the set of aircraft corresponding to airport \((k + 1)\) intersect. The slot constraints will increase linearly with \(K\).

Overall, it was decided that the primary objective of the optimization model should be to minimize the maximum displacement. The reasoning behind this choice is that, if we optimize for the total displacement, the solution of the optimization model will have the tendency to displace aircraft that fly few flights a day more than it displaces aircraft that operate many daily flights. For example, at a busy hub airport, there are many aircraft that operate long-haul flights and thus might only perform two flights a day. Such aircraft will visit the hub only once per day—one arrival at the hub and one departure from it. If these flights are scheduled during peak times and we only optimize for the total displacement across all flights, they would tend to be displaced more than flights operated by aircraft that fly multiple short-haul flights during the day. A displacement of 2 hours for an aircraft flying two long haul flights daily will lead to 4 hours of total displacement displacement. In contrast, for an aircraft flying 5 short haul flights daily, a displacement of one of its flights by 1 hour will lead to 5 hours of total displacement in the aircraft’s schedule. Hence, when optimizing for \(\sum_{i \in F} |u_i|\) and a long haul and a short haul flight are scheduled to operate at roughly the same time period, the choice, in the presence of slot controls, will generally be to displace the long haul flight rather than the short
haul one.

It turns out that in order to reduce the solution time in CPLEX it is necessary to break the original optimization problem into two optimization problems solved sequentially: in the first problem we minimize only the maximum displacement obtaining a solution $w^*$; in the second problem, we minimize the total displacement, while constraining the maximum displacement to equal the result of the first model, i.e. $d_i \leq w^*$.

Furthermore, by first solving without including the passenger constraints, we obtain a lower bound on the maximum displacement that would be necessary if passenger connection constraints were included. If, for example, the result of the model without passenger constraints is $w^* = 2$, then it is guaranteed that the optimization result with passenger constraints can never be less than 2. In addition, to speed up the solution, we may also try to infer an upper bound for the maximum displacement that would be necessary in the presence of passenger constraints, e.g., $w^* = 4$ in the above example. In this way we can reduce significantly the search space of the optimization algorithm in the most general case (i.e., when passenger constraints are included) and consequently the computational effort.

5.3.7 Experimental Setup

As EWR is one of the busiest US airports and at the same time the hub of one of the largest US carriers, Continental Airlines, it was decided to test there the application of slot constraints. We chose to execute the optimization model with the schedule of one of the busy days at EWR in 2007. On 07/13/2007, 1252 flights were scheduled to depart/arrive at/from EWR, making that day one of the busiest 5% of days in 2007, as shown in Figure 5-8.

In the absence of a publicly available OAG database, we obtained the aircraft schedules from the Individual Flights ASPM database [19]. On that day there were very few cancelations according to the On-time Performance database of the Bureau of Transportation and Statistics (BTS) [53], as good weather prevailed throughout the country. Hence the ASPM database includes most of the flights scheduled to operate
that day (ASPM only includes non-cancelled flights). We corrected for any missing tail numbers from the Individual Flights database, by making inferences based on the airline operating the flight, the aircraft type, the time of the day and the origin and destination of the flight. For example, if a flight without a tail number was operated by a B737 and was scheduled to arrive at EWR from Raleigh-Durham at 12:00, and another flight without a tail number entry was also operated by a B737 and was scheduled to depart from EWR to Raleigh-Durham at 13:00, then it was assumed that the flights were operated by the same aircraft. We also used the online database of flightstats.com [9] to account for international flights in and out of EWR that were missing from the ASPM database. In total 1910 flights were included in the optimization (out of which 1252 flights operated directly into or out of EWR and the remaining at airports visited by the corresponding aircraft in-between visits to EWR). The original schedule of arrivals and departures for that day at EWR is shown in Figure 5-9.

The parameters $\beta_{ij} \forall i, j \in \{2 \ldots f\}$ are obtained from the analysis we performed in Chapter 3 on the subject of the estimation of the minimum turnaround times as a function of the aircraft type, the operating airline and the type of airport where the connection takes place (a hub or any other airport).

To obtain the passenger connections at EWR, we used a database developed by Barnhart, Vaze and Fearing [4] through a discrete choice model that estimates the exact itineraries flown by all passengers in 2007 in the United States from publicly available aggregate data from BTS [53]. They trained their model using a real set of passenger itinerary data of a major US airline for one quarter in 2007. According to this passenger itinerary database, there were 1800 distinct passenger connections on 07/13/2007 at EWR; where by distinct we mean that multiple passengers connecting between the exact same flights are considered as one.

We tested the model and the effect of demand smoothing for EWR under various sets of inputs. Table 5.2 provides a summary of the different tests. We chose three levels of slot constraints. In the first level, we set the limit for arrivals and departures very near the expected VFR capacity of EWR as estimated in Section 5.2.2 and, in the
second and third, near the IFR capacity for the most common runway configuration under these conditions—4R|4L. The numbers for slot limits given as $X/Y$ in Table 5.2 signify that slot limits alternate from one quarter to the next from X to Y; for example 20/19 “total slots” translates to 78 slots per hour. In order to put these slot limits in perspective and relate them to the analysis presented in Section 5.2.2, 37% of all departures in 2007 were scheduled in periods when there were more than 11 scheduled departures and 46% were scheduled in periods with more than 10 scheduled departures. In terms of scheduled arrivals the respective numbers are 27% and 37%.
Table 5.2: Test scenarios for demand smoothing at EWR.

<table>
<thead>
<tr>
<th>Total slots per 15mins</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total slots per 15mins = 20</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arr slots per 15mins = 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dep slots per 15 mins = 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total slots per 15mins = 20/19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arr slots per 15mins = 10/11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dep slots per 15 mins = 11/10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Aircraft Connectivity Constraints: | |
| (a) First set | √ | | | |
| Aircraft Connectivity Constraints: | | | | | |
| (b) Second set | | | | | |
| Passenger Connectivity Constraints | | | | | |
| with α = 30mins | | | | | |

5.4 Optimization results

In this section, we concentrate on the optimization results obtained from the 5 different tests. From the first row of Table 5.3 it can be seen that, for all five tests, the maximum schedule displacement across all flights is only two 15-minute periods. The second row of Table 5.3 shows the result of the second optimization step, where we minimize the total displacement. Clearly, by changing the aircraft connectivity constraints from Test 1 to Test 2 and hence allowing the optimizer to reduce the connection time between flights, the total schedule displacement is reduced by 109 periods (from 382 to 273), which is equivalent to 27.25 hours. The total number of displaced flights also drops by 41, as shown in the third row of Table 5.3. This is achieved by reducing the turn time between 104 of the flights as shown in the last row of Table 5.3. Furthermore, after introducing passenger constraints, there is only a small increase in the total displacement, from 273 in Test 2 to 275 in Test 3 and, surprisingly, a reduction in the number of displaced flights.

In Table 5.3 we also clearly notice that the introduction of lower slot limits in Tests 4 and 5 lead to a large increase both in the total schedule displacement and in the total number of displaced flights. Even though the maximum displacement
remains the same, the total displacement increases by 172 periods from Test 3 to Test 4, and by an additional 203 periods from Test 4 to Test 5.

Table 5.4 shows the number of displaced arrivals and departures and the resulting total schedule displacement for all the flights operating directly in/out of EWR (a total of 1252 flights out of the 1910 flights included in the optimization, in contrast to Table 5.3 which included statistics about all 1910 flights in the optimization). We observe that in all 5 Tests a greater number of departures than arrivals is displaced. This is an expected result, since as presented in Sections 5.2.2 and 5.3.7 departures at EWR are more “concentrated” than arrivals and exceed the slot limits more often during the day.

Table 5.3: Optimization results for all flights in $F$ in EWR tests (1 period = 15 minutes).

<table>
<thead>
<tr>
<th></th>
<th>Test1</th>
<th>Test2</th>
<th>Test3</th>
<th>Test4</th>
<th>Test5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w^*$ (in periods)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$\sum_{i \in F}</td>
<td>u_i</td>
<td>$ (in periods)</td>
<td>382</td>
<td>273</td>
<td>275</td>
</tr>
<tr>
<td>Displaced flights</td>
<td>281</td>
<td>240</td>
<td>237</td>
<td>363</td>
<td>491</td>
</tr>
<tr>
<td>Average displacement</td>
<td>1.36</td>
<td>1.14</td>
<td>1.16</td>
<td>1.23</td>
<td>1.32</td>
</tr>
<tr>
<td>per displaced flight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flights with reduced turn</td>
<td>-</td>
<td>106</td>
<td>104</td>
<td>125</td>
<td>163</td>
</tr>
<tr>
<td>time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4: Optimization results for flights in $F$ that operate directly in/out of EWR (1 period = 15 minutes).

<table>
<thead>
<tr>
<th></th>
<th>Test1</th>
<th>Test2</th>
<th>Test3</th>
<th>Test4</th>
<th>Test5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displaced EWR arrivals</td>
<td>104</td>
<td>78</td>
<td>79</td>
<td>135</td>
<td>207</td>
</tr>
<tr>
<td>Displaced EWR departures</td>
<td>124</td>
<td>135</td>
<td>130</td>
<td>192</td>
<td>229</td>
</tr>
<tr>
<td>Total displaced flights at EWR</td>
<td>228</td>
<td>213</td>
<td>209</td>
<td>327</td>
<td>436</td>
</tr>
<tr>
<td>$\sum_{i \in F}</td>
<td>u_i</td>
<td>$ of EWR arrivals (in periods)</td>
<td>146</td>
<td>84</td>
<td>90</td>
</tr>
<tr>
<td>$\sum_{i \in F}</td>
<td>u_i</td>
<td>$ of EWR departures (in periods)</td>
<td>184</td>
<td>160</td>
<td>156</td>
</tr>
<tr>
<td>$\sum_{i \in F}</td>
<td>u_i</td>
<td>$ of all EWR flights (in periods)</td>
<td>330</td>
<td>244</td>
<td>246</td>
</tr>
</tbody>
</table>

Table 5.5 shows the percentage share of the flights in $F$ operated by each airline.

In Table 5.6 we present for each airline the number of its flights that are displaced for
each optimization test, as well as each airline's share of the total number of displaced flights. In Table 5.7 we also show, for every airline, the average displacement for every displaced flight. As expected Continental Airlines (COA) and its regional partner Continental Express (BTA) that use EWR as a hub have, by far, the most displaced flights in all 5 tests. Moreover, even though COA operates around 30% of all flights, as shown in Table 5.5, its share of the number of flights displaced is more than 50% in Test 1. By allowing shorter connection times in Test 2, we observe that COA's share of displaced flights is reduced from 55% to less than 40%, even when introducing passenger connections constraints in Tests 3, 4 and 5. In contrast, BTA's share of displaced flights increases, when we allow smaller connection times from 33 to 55 displaced flights. This can be attributed to the fact that BTA flies on a tighter schedule than COA, i.e., with less slack embedded into the ground turnaround times, and hence with little room for significant reduction in BTA's connection times. For all other airlines, their share of schedule displacement is approximately the same as their share of flights. We may also observe that although the average amount of displacement changes between tests (Table 5.3), it does not vary significantly amongst airlines (Table 5.6) within each test.

Table 5.5: Percentage of flights operated by each airline at EWR tests.

<table>
<thead>
<tr>
<th></th>
<th>% of total flights</th>
</tr>
</thead>
<tbody>
<tr>
<td>COA</td>
<td>30%</td>
</tr>
<tr>
<td>BTA</td>
<td>32%</td>
</tr>
<tr>
<td>AAL</td>
<td>4%</td>
</tr>
<tr>
<td>NWA</td>
<td>5%</td>
</tr>
<tr>
<td>UAL</td>
<td>3%</td>
</tr>
<tr>
<td>DAL</td>
<td>4%</td>
</tr>
<tr>
<td>JBU</td>
<td>2%</td>
</tr>
</tbody>
</table>

Finally, Figure 5-10 depicts the arrival and departure schedules per 15 minutes that resulted from the demand smoothing model for Tests 3, 4 and 5. We also show the original schedule for the sake of comparison. We exclude, however, the schedules from Tests 1 and 2, as these tests were run with exactly the same slot limits as Test 3 and thus their resulting schedules were almost identical with those from Test 3.
Table 5.6: Optimization results: Number of flights displaced by airline in EWR tests.

<table>
<thead>
<tr>
<th>Airline</th>
<th>Test1</th>
<th>Test2</th>
<th>Test3</th>
<th>Test4</th>
<th>Test5</th>
</tr>
</thead>
<tbody>
<tr>
<td>COA</td>
<td>144 (55%)</td>
<td>88 (37%)</td>
<td>87 (37%)</td>
<td>140 (39%)</td>
<td>191 (39%)</td>
</tr>
<tr>
<td>BTA</td>
<td>33 (13%)</td>
<td>55 (23%)</td>
<td>52 (22%)</td>
<td>82 (23%)</td>
<td>113 (23%)</td>
</tr>
<tr>
<td>AAL</td>
<td>9 (3%)</td>
<td>14 (6%)</td>
<td>20 (8%)</td>
<td>18 (5%)</td>
<td>19 (4%)</td>
</tr>
<tr>
<td>NWA</td>
<td>4 (2%)</td>
<td>10 (4%)</td>
<td>10 (4%)</td>
<td>13 (4%)</td>
<td>16 (3%)</td>
</tr>
<tr>
<td>UAL</td>
<td>10 (4%)</td>
<td>11 (5%)</td>
<td>9 (4%)</td>
<td>15 (4%)</td>
<td>18 (4%)</td>
</tr>
<tr>
<td>DAL</td>
<td>4 (2%)</td>
<td>7 (3%)</td>
<td>8 (3%)</td>
<td>6 (2%)</td>
<td>10 (2%)</td>
</tr>
<tr>
<td>JBU</td>
<td>4 (2%)</td>
<td>4 (2%)</td>
<td>4 (2%)</td>
<td>6 (2%)</td>
<td>8 (2%)</td>
</tr>
</tbody>
</table>

Table 5.7: Optimization Results: Average displacement by airline in EWR tests.

<table>
<thead>
<tr>
<th>Airline</th>
<th>Test1</th>
<th>Test2</th>
<th>Test3</th>
<th>Test4</th>
<th>Test5</th>
</tr>
</thead>
<tbody>
<tr>
<td>COA</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>BTA</td>
<td>1.5</td>
<td>1.0</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>AAL</td>
<td>1.8</td>
<td>1.14</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>NWA</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>UAL</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>DAL</td>
<td>1.8</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td>JBU</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

It should be pointed out that, compared to the original schedule, the new schedules show the following differences: the morning departure peak starts approximately 30 minutes earlier at 5:45 and ends 15 minutes later at 9:30, while the afternoon arrival peak starts 30 minutes earlier at 12:30 and ends 30 minutes later at 21:00.

Overall, the most important observation from the tests we performed at EWR is that even when the slot limits are set to, as low as, 10 departures and 10 arrivals per 15 minutes (in Test 5) there are enough slots to accommodate all the traffic of a busy day in EWR without having to shift the schedule arrival or departure time of any flight by more than 30 minutes.

5.5 Testing in AND

In order to assess the benefits from the modified schedules produced by the demand smoothing model and the different levels of slots, we tested the new schedules in
AND. We shall compare various statistics of the delays associated with the schedules of Tests 3, 4 and 5 from Table 5.2 with the original schedule before smoothing the demand (we exclude schedules from Tests 1 and 2, as these tests were run with exactly the same slot limits as Test 3 and thus their resulting profiles at EWR were almost identical with those from Test 3). In the schedules run by AND we do, of course, also include all the other flights that remained unaffected by the changes in EWR, as originally scheduled (approximately 32,000 flights). We compare the delays under two scenarios:

**Scenario 1:** EWR operates under optimum conditions (VFR). We assume that configuration 22L|22R is used before 13:00 with more emphasis on departures, configuration 11, 22L|22R from 13:00 to 18:00 with more emphasis on arrivals and configuration 22L|22R after 18:00 with a balance between arrival and de-
departure throughput.

**Scenario 2**: EWR operates under IFR conditions. We assume that configuration $4R|4L$ is used for the entire day but with more emphasis on departures before 13:00 and more emphasis on arrivals after 13:00.

In both Scenarios “by more emphasis” on arrivals we mean that departure throughput is sacrificed for a higher arrival throughput as EWR is expecting more arrivals—referring to Figure 5-4 this is achieved as we move right-wise on the expected capacity line. The opposite holds when “emphasizing more” on departures. For example, in Scenario 2, when the same configuration is used throughout the day ($4R|4L$), the morning capacity is assumed to be 10 departures and 8 arrivals per 15 minutes, while in the afternoon it is assumed to be 9 departures and 10 arrivals per 15 minutes.

We break down our analysis into two parts. In the first we look at the local delay changes at EWR, while in the second we discuss the benefits of demand smoothing at the network level.

### 5.5.1 Local Delays at EWR

In this section we present and discuss the results from the AND test at the local level at EWR. Figure 5-11 depicts the average delay per 15 minutes due to *local* congestion at EWR for both arrivals and departures (local arrival delay and local departure delay respectively), under Scenario 1. The first observation is that the local departure delay (Figure 5-11b) has decreased throughout the day as a result of the smoothing of the schedule. Especially in the morning hours when there is a large demand for departures during a 3 hour span, the local departure delay is reduced by as much as 80%. Clearly the smoothed schedule of Test 5, which has the lowest slot limit, also produces the lowest local departure delay.

The benefits of demand smoothing are not as obvious when it comes to the local arrival delay (shown in figure 5-11a) as they were for the local departure delay. There is only a small decrease during the morning hours until 11:00am since the arrival demand was already low during that time. Interestingly, in the early afternoon,
we clearly see that as the slots are reduced (from 11 in test 3 to 10/11 in test 4 and 10 in test 5) the local arrival delay increases compared to the original schedule. The reason is that, in the three tests, the peak afternoon arrival profile has been moved by the optimization model to a slightly earlier time due to demand smoothing, starting at around 12:45 for Test 3, 12:15 for Test 4 and 11:45 for Test 5, rather than at 13:00 in the original schedule (Figures 5-9 and 5-10). At the same time, the runway configuration to allow for a higher arrival throughput only changes at 13:00. Hence, under the schedules of Tests 3, 4 and 5 there is insufficient arrival capacity to accommodate the arrival demand between 12:00 and 13:00 and a longer arrival queue is formed. This issue will be dealt with in section 5.5.3.

Overall, we notice that with the application of slot limits the average local arrival and local departure delay at EWR can be reduced, as shown in Table 5.8. Especially in the case of departures, which under the original schedule were much more peaked than the arrivals, the reduction in average delay is estimated to be 51% (from 12.5 minutes to 6.1) if we introduce slot limits at the level of 10 per quarter (Test 5). Moreover, the maximum local delay is greatly reduced with the application of slot control, signifying that delays are more evenly spread among flights. For the schedule of Test 4, for example, the maximum local arrival delay drops from 21.9 minutes to 16.6 while the maximum local departure delay drops from 37.3 to 13.4 minutes.

Table 5.8: Local delays at EWR in VFR (in parenthesis is shown the percentage change from the original schedule)

<table>
<thead>
<tr>
<th>Delays (in minutes)</th>
<th>Original</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg arrival</td>
<td>7.7</td>
<td>6.8 (-12%)</td>
<td>6.9 (-10%)</td>
<td>6.9 (-10%)</td>
</tr>
<tr>
<td>Max arrival</td>
<td>21.9</td>
<td>12.8 (-42%)</td>
<td>16.6 (-24%)</td>
<td>18.8 (-14%)</td>
</tr>
<tr>
<td>Avg departure</td>
<td>12.5</td>
<td>7.8 (-38%)</td>
<td>6.6 (-47%)</td>
<td>6.1 (-51%)</td>
</tr>
<tr>
<td>Max departure</td>
<td>37.3</td>
<td>17.9 (-52%)</td>
<td>13.4 (-64%)</td>
<td>10.8 (-71%)</td>
</tr>
</tbody>
</table>

Turning to Scenario 2, as shown in Figure 5-12, we notice again that the arrival delay is higher from 11:00 to 16:00 when we introduce slot limits, for the same reasons as noted above in connection with Scenario 1. In the late afternoon and evening, however, we clearly see the benefit of the arrival slots contraints, as the delay for the schedule of Test 5 is up to 20 minutes lower than that of the original schedule. On
average, for the entire the day, the arrival delay remains roughly the same as for the original schedule in the cases of Tests 3 and 4 and is reduced by 7% for the slot limits of Test 5 (Table 5.9).

Under Scenario 2, the reduction of local delays is again more evident for departures, especially before 12:30. The lowest slot limits (Test 5) clearly produce, once
again, the lowest departure delays. The most interesting observation in this case is that the departure delay profiles from the four schedules are very close to each other after 13:00. For the schedule of Test 3—which is also the closest to the original schedule—this happens because the slot limit of 11 departures per quarter is very close to the original departure schedule in the afternoon. For Tests 4 and 5, however, this behavior is the result of delay propagation.

In Figure 5-13 we plot the scheduled demand profile and the adjusted demand profile at the end of the day at EWR (as defined in Chapter 2) for the schedule of Test 5. In Figure 5-13b, we notice that, due to delay propagation, the departure schedule does not remain smooth after 14:00. This happens because arriving aircraft at EWR experience a high local arrival delay and are therefore not ready to depart on time for their next flight. We do not observe the same effect from delay propagation for the arrival schedule except only after 21:00, the reason being that the departure delay in the morning is small and flights that departed delayed in the morning from EWR and returned later in the day at EWR were able to compensate for most of that delay through the slack in their schedules. However, we would expect to see a much less smooth arrival demand profile when a few other airports generate high delays and flights start arriving late at EWR due to congestion elsewhere in the system.

Table 5.9 shows the average and maximum local delays at EWR under Scenario 2. With 10 slots for departures (Test 5) throughout the day, we expect the reduction of average local departure delays to be of the order of 30%.

<table>
<thead>
<tr>
<th>Local Delays</th>
<th>Original</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg arrival</td>
<td>21.8</td>
<td>21.9 (0%)</td>
<td>21.8 (0%)</td>
<td>20.4 (-7%)</td>
</tr>
<tr>
<td>Max arrival</td>
<td>51.2</td>
<td>39.8 (-22%)</td>
<td>37.6 (-27%)</td>
<td>33.2 (-35%)</td>
</tr>
<tr>
<td>Avg departure</td>
<td>25.9</td>
<td>20.7 (-20%)</td>
<td>17.9 (-31%)</td>
<td>17.3 (-33%)</td>
</tr>
<tr>
<td>Max departure</td>
<td>57.8</td>
<td>56.1 (-3%)</td>
<td>50.4(-12%)</td>
<td>48.5((-16%)</td>
</tr>
</tbody>
</table>

Finally, we plot in Figure 5-14 the average total flight arrival delay per hour. The total flight arrival delay, as explained in Chapter 2, is defined as the sum of the upstream delay of a flight and the local arrival delay that the flight experiences.
Figure 5-12: Delay profiles at EWR in IFR conditions under the original schedule and the new schedules from Tests 3, 4 and 5.

at EWR at the time of arrival. With the presence of many aircraft that visit EWR multiple times in a day, we notice that the differences in the average total flight arrival delay between the original and the smoothed schedules are greater than when just looking at the local arrival delay (shown in Figures 5-11a and 5-12a). The reason is the large reduction of the local departure delay. Flights that take off in the morning
from EWR experience lower delays under the smoothed schedules and thus carry less upstream delay when they return to EWR in the afternoon. In particular, the average total flight arrival delay with the smoothed schedules drops by more than 20% in VFR conditions and by around 19% in IFR conditions, as shown in Table 5.10. Thus, the reduction of the delay is much higher when considering the total flight arrival delay rather than just the local arrival delay at EWR (comparing with Tables 5.8 and 5.9, which show approximately a 10% delay reduction in VFR and 7% in IFR conditions).
Figure 5-14: Flight arrival delay profile at EWR in VFR and IFR conditions for the four different schedules.

Table 5.10: Average flight arrival delay at EWR (shown in parenthesis is the percentage reduction from the original schedule)

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFR</td>
<td>8.9</td>
<td>6.8 (-23%)</td>
<td>6.9 (-22%)</td>
<td>6.9 (-22%)</td>
</tr>
<tr>
<td>IFR</td>
<td>26.9</td>
<td>23.8 (-11%)</td>
<td>23.2 (-14%)</td>
<td>21.7 (-19%)</td>
</tr>
</tbody>
</table>

5.5.2 Network delays

In this section we present and discuss the network effects of introducing slot limits at EWR. Table 5.11 shows the total upstream delay at 25 out of the 34 airports of the
AND network. Under scenario 1 the delay reduction ranges from 0% to 45% for Test 3, 0% to 51% for Test 4 and 0% to 54% for Test 5. As expected the upstream delay at JFK and LGA is not affected by the slot limits at EWR since there are no aircraft visiting more than one of the three New York airports in one day. Furthermore, we notice that the upstream delay at EWR is reduced by 98%, since the new low delays at EWR can be compensated by the slack in the schedules of airlines. In Scenario 1, the total upstream delay at all the 34 airports of the AND network is reduced from 11,700 minutes to 9,300 with the schedule of Test 3 (-21%), to 9,100 with that of Test 4 (-22%) and 9,000 with the schedule of Test 5 (-23%).

Similarly, under Scenario 2, the maximum upstream delay reduction at any single airport other than EWR is 39% for Test 3, 47% for Test 4 and 49% for Test 5 (all at Houston International airport which is COA’s other principal hub). The upstream delay at EWR shows a smaller reduction compared to Scenario 1 (ranging from 63% to 74%) due to the congestion caused by the IFR conditions at the airport that leads to high delays. Overall, the total upstream delay at the 34 airports of the AND network is reduced from 21,400 minutes to 16,300 with the schedule of Test 3 (-24%), 14,800 with the schedule of Test 4 (-31%) and 14,500 with that of Test 5 (-32%).

The enormous impact on network wide delays in the United States just by smoothing the demand at one busy airport (EWR), should be highlighted here. Without sacrificing any of the demand, but only by rationalizing the schedule at EWR, a reduction of at least 21% in downstream delays when EWR is operating in VFR conditions and 24% when operating in IFR conditions may be expected (depending on the level of slots the reduction can be much higher).

5.5.3 Modified runway utilization

As described in Section 5.4 the demand smoothing model reallocates many early-afternoon arrivals to earlier times. In the case of Test 5, which has the lowest slot limits, this leads to 6 more scheduled arrivals from 12:00pm to 1:30pm than in the original schedule. As indicated in the previous section, we assumed in our original computational tests that the time when the runway configuration changes from
Table 5.11: Total daily upstream delays (in minutes) at selected airports of the AND network for Scenarios 1 and 2 (shown in parenthesis is the percentage reduction from the original schedule)

<table>
<thead>
<tr>
<th>Airport</th>
<th>Test 3 (Scenario 1)</th>
<th>Test 4 (Scenario 1)</th>
<th>Test 5 (Scenario 1)</th>
<th>Test 3 (Scenario 2)</th>
<th>Test 4 (Scenario 2)</th>
<th>Test 5 (Scenario 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>772 (-14%)</td>
<td>767 (-14%)</td>
<td>746 (-17%)</td>
<td>1283 (-11%)</td>
<td>1208 (-16%)</td>
<td>1142 (-21%)</td>
</tr>
<tr>
<td>BOS</td>
<td>1102 (-6%)</td>
<td>1089 (-7%)</td>
<td>1081 (-8%)</td>
<td>1369 (-9%)</td>
<td>1315 (-13%)</td>
<td>1283 (-15%)</td>
</tr>
<tr>
<td>CLE</td>
<td>366 (-25%)</td>
<td>355 (-28%)</td>
<td>354 (-28%)</td>
<td>656 (-34%)</td>
<td>573 (-43%)</td>
<td>540 (-46%)</td>
</tr>
<tr>
<td>CLT</td>
<td>349 (-13%)</td>
<td>342 (-15%)</td>
<td>343 (-15%)</td>
<td>690 (-11%)</td>
<td>619 (-21%)</td>
<td>636 (-18%)</td>
</tr>
<tr>
<td>CVG</td>
<td>340 (-14%)</td>
<td>326 (-18%)</td>
<td>326 (-18%)</td>
<td>475 (-10%)</td>
<td>442 (-16%)</td>
<td>451 (-14%)</td>
</tr>
<tr>
<td>DCA</td>
<td>794 (-8%)</td>
<td>790 (-8%)</td>
<td>804 (-7%)</td>
<td>987 (-12%)</td>
<td>946 (-15%)</td>
<td>957 (-14%)</td>
</tr>
<tr>
<td>DEN</td>
<td>162 (-19%)</td>
<td>157 (-22%)</td>
<td>156 (-22%)</td>
<td>227 (-25%)</td>
<td>210 (-30%)</td>
<td>209 (-31%)</td>
</tr>
<tr>
<td>DFW</td>
<td>337 (-20%)</td>
<td>325 (-23%)</td>
<td>324 (-23%)</td>
<td>683 (-14%)</td>
<td>621 (-21%)</td>
<td>597 (-25%)</td>
</tr>
<tr>
<td>DTW</td>
<td>267 (-27%)</td>
<td>257 (-30%)</td>
<td>250 (-32%)</td>
<td>476 (-23%)</td>
<td>445 (-28%)</td>
<td>404 (-35%)</td>
</tr>
<tr>
<td>EWR</td>
<td>10 (-98%)</td>
<td>12 (-98%)</td>
<td>15 (-98%)</td>
<td>1050 (-63%)</td>
<td>727 (-74%)</td>
<td>731 (-74%)</td>
</tr>
<tr>
<td>FLL</td>
<td>282 (-18%)</td>
<td>275 (-20%)</td>
<td>276 (-20%)</td>
<td>430 (-17%)</td>
<td>406 (-21%)</td>
<td>402 (-22%)</td>
</tr>
<tr>
<td>IAH</td>
<td>264 (-30%)</td>
<td>257 (-32%)</td>
<td>254 (-33%)</td>
<td>559 (-39%)</td>
<td>489 (-47%)</td>
<td>466 (-49%)</td>
</tr>
<tr>
<td>JFK</td>
<td>127 (0%)</td>
<td>127 (1%)</td>
<td>126 (0%)</td>
<td>137 (-28%)</td>
<td>126 (-33%)</td>
<td>143 (-25%)</td>
</tr>
<tr>
<td>LAX</td>
<td>157 (-29%)</td>
<td>150 (-32%)</td>
<td>149 (-32%)</td>
<td>272 (-28%)</td>
<td>240 (-37%)</td>
<td>259 (-32%)</td>
</tr>
<tr>
<td>LGA</td>
<td>755 (0%)</td>
<td>755 (0%)</td>
<td>754 (0%)</td>
<td>760 (-3%)</td>
<td>762 (-3%)</td>
<td>760 (-3%)</td>
</tr>
<tr>
<td>MDW</td>
<td>139 (-8%)</td>
<td>137 (-10%)</td>
<td>137 (-10%)</td>
<td>193 (-20%)</td>
<td>171 (-29%)</td>
<td>182 (-24%)</td>
</tr>
<tr>
<td>MIA</td>
<td>255 (-10%)</td>
<td>256 (-10%)</td>
<td>249 (-12%)</td>
<td>409 (-11%)</td>
<td>401 (-13%)</td>
<td>393 (-15%)</td>
</tr>
<tr>
<td>MSP</td>
<td>154 (-22%)</td>
<td>141 (-29%)</td>
<td>142 (-29%)</td>
<td>289 (-22%)</td>
<td>258 (-31%)</td>
<td>238 (-36%)</td>
</tr>
<tr>
<td>ORD</td>
<td>767 (-14%)</td>
<td>772 (-13%)</td>
<td>764 (-14%)</td>
<td>1190 (-15%)</td>
<td>1138 (-19%)</td>
<td>1110 (-21%)</td>
</tr>
<tr>
<td>PDX</td>
<td>65 (-45%)</td>
<td>62 (-47%)</td>
<td>66 (-44%)</td>
<td>137 (-37%)</td>
<td>117 (-46%)</td>
<td>115 (-47%)</td>
</tr>
<tr>
<td>PHL</td>
<td>193 (-9%)</td>
<td>198 (-7%)</td>
<td>198 (-7%)</td>
<td>197 (-11%)</td>
<td>200 (-10%)</td>
<td>197 (-11%)</td>
</tr>
<tr>
<td>PIT</td>
<td>291 (-18%)</td>
<td>268 (-25%)</td>
<td>270 (-24%)</td>
<td>568 (-16%)</td>
<td>510 (-25%)</td>
<td>503 (-26%)</td>
</tr>
<tr>
<td>SFO</td>
<td>92 (-38%)</td>
<td>73 (-51%)</td>
<td>69 (-54%)</td>
<td>282 (-16%)</td>
<td>258 (-23%)</td>
<td>222 (-34%)</td>
</tr>
<tr>
<td>STL</td>
<td>160 (-31%)</td>
<td>147 (-36%)</td>
<td>145 (-37%)</td>
<td>373 (-19%)</td>
<td>327 (-29%)</td>
<td>330 (-28%)</td>
</tr>
<tr>
<td>TPA</td>
<td>174 (-17%)</td>
<td>177 (-15%)</td>
<td>171 (-18%)</td>
<td>263 (-15%)</td>
<td>263 (-15%)</td>
<td>243 (-21%)</td>
</tr>
</tbody>
</table>

22L|22R to 11, 22L|22R at 13:00 in order to accommodate the increasing arrival traffic of the original schedule remains unchanged. We modify this assumption in this section to explore the benefits that may be obtained from a more timely change of the runway configuration, designed to improve the utilization of the runway capacity. Note that the new schedules resulting from the smoothing of the demand at EWR clearly requires more arrival capacity earlier in the day. Hence we perform an additional set of tests, in which the runway configuration is switched from 22L|22R to 11, 22L|22R at 12:00 instead of 13:00.

Figure 5-15 shows the local delay profiles at EWR for the original schedule and the schedule of Test 5 under Scenario 1 after the change in the runway configuration has been switched to 12:00. The increase in the arrival throughput due to the earlier
change of the runway configuration, results in a large reduction in the arrival delay from 12:00 until 16:15. Moreover, even though the departure capacity decreases by one operation per quarter from 12:00 to 13:00, there is no significant departure delay increase, due to the fact that during that period the demand for departures is small anyway. Hence, the benefits of demand smoothing can be much higher than the results presented earlier, when the runway configuration utilized is also modified taking into account the changes in the traffic profiles.

The reduction of local arrival delay in the afternoon leads to an average local arrival delay for the day at EWR of 5.4 minutes, −30% lower than the delay created by the original schedule. Furthermore, the daily average total flight arrival delay is also 5.4 minutes, which is approximately 3 minutes lower than the average total flight delay of the original schedule at EWR (Table 5.13). The aggregate upstream delay at all the airports of the AND network drops to 8,900 minutes, which is a 24% reduction from the original schedule.

Table 5.12: Local delays at EWR in VFR (shown in parenthesis is the percent change from the original schedule).

<table>
<thead>
<tr>
<th>Delays (in minutes)</th>
<th>Original</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 5 modified config.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg arrival</td>
<td>7.7</td>
<td>6.8 (-12%)</td>
<td>6.9 (-10%)</td>
<td>6.9 (-10%)</td>
<td>5.4 (-30%)</td>
</tr>
<tr>
<td>Max arrival</td>
<td>21.9</td>
<td>12.8 (-42%)</td>
<td>16.6 (-24%)</td>
<td>18.8 (-14%)</td>
<td>10.8 (-51%)</td>
</tr>
<tr>
<td>Avg departure</td>
<td>12.5</td>
<td>7.8 (-38%)</td>
<td>6.6 (-47%)</td>
<td>6.1 (-51%)</td>
<td>6.0 (-52%)</td>
</tr>
<tr>
<td>Max departure</td>
<td>37.3</td>
<td>17.9 (-52%)</td>
<td>13.4 (-64%)</td>
<td>10.8 (-71%)</td>
<td>10.7 (-71%)</td>
</tr>
</tbody>
</table>

Table 5.13: Average flight arrival delay at EWR (in minutes), (shown in parenthesis is the percent reduction from the original schedule).

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Test 5</th>
<th>Test 5 modified config.</th>
</tr>
</thead>
<tbody>
<tr>
<td>VFR</td>
<td>8.9</td>
<td>6.8 (-23%)</td>
<td>6.9 (-22%)</td>
<td>6.9 (-22%)</td>
<td>6.0 (-32%)</td>
</tr>
</tbody>
</table>

Concluding, a smooth schedule at EWR combined with a minor modification of a runway configuration utilization is estimated to provide a local delay reduction that may be as high as 30% for arrivals and 50% departures. The average total flight arrival delay at EWR, which takes into account delay propagation, may be reduced
by 32%. Of course, these delay reductions are achieved without altering the total number of flights operating at EWR and by respecting all passenger connections taking place there. The implementation of such “demand-smoothing” policy would, of course, require the close coordination of airlines with authorities.
5.6 Summary

In this Chapter we presented a demand-smoothing model for busy airports that modifies the original operations schedule in response to the imposition of slot constraints. The model is formulated as an Integer Program. Starting from a given original schedule, the IP re-schedules some flights in order to comply with the slot limits that have been imposed, while at the same time constraining all aircraft to fly their original itineraries and all connecting passengers to fly their booked itineraries. The objective of the model is to minimize both the maximum schedule displacement experienced by any single flight and the aggregate schedule displacement across all flights.

Using this optimization model in tandem with AND, we carried out an extensive study of the situation at EWR, considering the 2007 schedule of arrivals and departures there along with the capacity of the airport. We concluded that the slot limits that are currently being enforced at EWR are insufficient to tackle the problem of congestion there. In the process, we also looked at different capacity metrics and proposed the expected capacity metric as the one most appropriate for planning purposes, such as determining the level of slot limits.

The demand-smoothing model was applied to a busy day in 2007 at EWR in a series of tests involving different realistic values for the slot limits. It was shown that for all these slot limits there exists a feasible schedule that may accommodate all the daily traffic, even when the slots are set to meet the IFR capacity of the airport, while respecting all constraints related to aircraft itineraries and passenger connections. Furthermore, it was observed that with slot limits near the IFR capacity of the airport, the maximum flight displacement was only 30 minutes while a total of 491 flights were displaced with an average displacement of 20 minutes.

We also estimated the delays associated with three different levels of slot controls using the Airport Network Delays model. For the first time in the literature an airport network model is used to estimate the local and system-wide effects of introducing slot controls at a busy airport. It is shown that, when setting the slot limits near the IFR capacity of EWR, the delay savings due to local congestion at EWR are
for arrivals of the order of 10% and for departures of the order of 50%, when EWR operates under VFR conditions. Similarly, under IFR conditions the reduction of delay is 7% for arrivals and 30% for departures. When also considering the upstream delay carried by flights while executing their daily schedules, the average flight arrival delay at EWR is reduced by 22% when EWR operates under VFR conditions. The system-wide delay savings from the application of slot limits at EWR was estimated to be of the order of 20% (a reduction of roughly 45 aircraft-hours daily). Finally, it was shown that by switching runway configurations at EWR earlier in the afternoon together with the application of slots, the arrival delay can drop even further, by as much as 32%.

Overall, it was shown that by smoothing the demand at a busy US airport, which is achieved with minor modifications of the current schedule, while respecting all of the existing demand, enormous delay savings can be achieved both in a local and system-wide scale.
Chapter 6

Conclusions

In this Thesis we developed an analytical, dynamic and stochastic model of large networks of airports that aims to facilitate a) the study of system-wide queuing phenomena, as they develop and propagate, and b) the efficient exploration and the macroscopic analyses at the network level of a broad range of demand management policies and infrastructure improvements. In this chapter we summarize the work presented in this Thesis and the main conclusions of our research, and describe possible future directions.

6.1 Summary of Thesis

We started in Chapter 1 by presenting a literature review of existing modeling tools of the NAS as a whole, as well as of research concerned with the application of demand management policies at airports. We drew two important conclusions in relation to the existing macroscopic research in the air transportation sector. First, we confirmed that there exists a gap in the NAS models between detailed micro-simulations that require extensive effort both in terms of input preparation and computation, and macroscopic models (simulations and analytical) that are simple to use, but typically lack aircraft itinerary tracking capabilities and credible queuing models of airport congestion. Second, the review pointed out that, despite the voluminous existing work on airport demand management, very limited research has been done to date
on the impact of applying slot controls at an airport on system-wide delays and delay propagation. This Thesis presents our work on the first topic in Chapters 2 through 4, while in Chapter 5 we describe our research on the second topic.

In Chapter 2 we presented the Airport Network Delays (AND) model, a stochastic and dynamic queuing network representation of air transportation networks. AND applies a decomposition approach, by iterating between a queuing engine—delay estimator—and a delay propagation algorithm. The queuing engine utilizes an approximation to the $M(t)/E_k(t)/1$ queuing system to estimate how delays occur at individual airports due to changes in the demand and/or the capacity at each airport during the course of a day. The delay propagation algorithm tracks how the aircraft fly through the network in small time steps—typically of 15 minutes length. At every time step the algorithm assigns improved delay estimates to every flight that has completed an operation (take-off and landing) by that time, and updates the schedules of flights for the rest of the day. The updated schedules are then used by the queuing engine to re-estimate the delays for the remainder of the day. The principal assumptions of the AND model are that: a) arrivals at the airport queues follow a time-varying Poisson distribution, b) airport service times follow a time-varying Erlang distribution, c) each runway system is modeled as a single server or as two servers, depending on local conditions, d) the airspace system is un-capacitated and e) airlines do not take any actions to recover from large delays.

In Chapter 3 we presented our work on a) the development of the various components of the AND model, b) the validation of the model and, c) some tests with several network-wide traffic and capacity scenarios that yield new insights on interactions among airports and on the downstream impact of delays incurred at any given airport. On the first topic we performed a statistical analysis of aircraft ground turnaround times, one of the fundamental variables that determine delay propagation. More specifically, our analysis provided estimates of the minimum turn times of aircraft, and of the slack embedded in airline schedules, as a function of the airline operating the aircraft, the type of airport where the connection takes place and the aircraft type. The data used in the analysis have been filtered carefully in order to
focus on connections that have potentially utilized the slack in their scheduled turn time. We concluded that, on average, airlines tend to turn their aircraft faster outside their hubs, and that the differences between Low Cost and Network Carriers in terms of how efficiently they perform their ground operations are statistically insignificant.

In relation to the validation of the AND model we presented, first, our work on modeling the airport departure process as a $M(t)/E_k(t)/1$ queuing system. The correct calibration of the model in terms of the server characteristics (service rate and Erlang order) is essential and, for this reason, we described in detail the overall methodology for estimating these two parameters. We showed that our departure process model is able to predict departure delays at Boston and Newark Airports with greater than 90% accuracy. The above analysis provided strong evidence that the DELAYS approximation to the $M(t)/E_k(t)/1$ queuing system is adequate for describing queues within an airport network model, such as AND. We then expanded our validation efforts to the entire AND model. We simulated two different busy days in 2007 with AND and compared the results with observed data from ASPM [19]. We concluded that AND is able to predict with reasonable accuracy aggregate delays in the network of US airports, as AND's results closely followed the delay profiles that were observed in reality. Hence, we are confident that AND is an appropriate tool for exploring at a "macroscopic", approximate level the types of issues that it was designed for.

We also presented a detailed example based on one of the families of tests conducted with the AND model with the intent of emphasizing the capabilities of the model and the insights that the model provides in relation to inter-airport interactions. We showed that delay propagation tends to "smoothen" daily airport demand profiles and push more demands into late evening hours. Such phenomena are especially evident at hub airports, where some flights may benefit considerably (by experiencing reduced delays) from the changes that occur in the scheduled demand profile as a result of delays and delay propagation. In addition, we showed that the upstream delay—i.e. the delay incurred elsewhere in the system—may often account for more than half of the total flight arrival delay at most of the airports included in
In Chapter 4 we presented two of the most important extensions implemented in the AND model. First, we described an algorithm that models Ground Delay Programs (GDP); a central coordination tool of the Air Traffic Management System. The GDP algorithm developed for AND consists of three main parts: a) GDP initiation, b) slot assignment and c) the decision on whether to terminate or continue the GDP. The algorithm is based on the maintenance of a priority queue, in the form of a binary tree, to store GDP exempt and non-exempt flights. The heap structure makes the very efficient the operation of adding/removing flights to/from the queue. We validated the GDP algorithm against observed data obtained from ASPM [19] over a day when ORD was operating under low IFR conditions and was issuing arrival slots for most of the day. The GDP algorithm offered a clear improvement on the delay estimation of AND.

The second extension presented in Chapter 4 was an alternative deterministic queuing engine to estimate delays. We implemented a queuing system with non-stationary and deterministic demand and service rates \((D(t)/D(t)/1)\). We compared the results of the deterministic to the stochastic, \(M(t)/E_k(t)/1\), queuing models. We showed that the higher the delays in the system, the closer are the delay estimates from the two queuing models. We also provided insights on how delay propagation affects the comparison between the deterministic and stochastic AND models. We then provided an example of how the "stochastic" and the "deterministic" versions of AND may be used to provide approximate estimates of the benefits obtainable from increased predictability of aircraft trajectories and processing times—both of which count among the principal objectives of the NextGen and Sesar systems.

In Chapter 5 we began with a detailed study of the relationship between demand and capacity at Newark International Airport (EWR), considering the 2007 schedule of arrivals and departures there along with the capacity of the airport. We concluded that the slot limits that are currently being enforced at EWR are insufficient to tackle the problem of congestion there. We also showed that 46% of all departures in 2007 were scheduled in periods when there were more scheduled departures than the IFR
departure capacity of the airport. Similarly, 37% of scheduled arrivals were in periods when there were more scheduled arrivals than the IFR arrival capacity of the airport.

Motivated by this analysis, we then presented a demand-smoothing model for busy airports that modifies the original operations schedule in response to the imposition of slot constraints. The model is formulated as a Mixed Integer Program. Starting from a given original schedule, the MIP re-schedules some flights in order to comply with the slot limits that have been imposed, while at the same time constraining all aircraft to fly their original itineraries and maintaining the feasibility of the booked itineraries of all connecting passengers. The objective of the model is to minimize the maximum schedule displacement experienced by any single flight, as well as the aggregate schedule displacement across all flights.

The demand-smoothing model was applied to a busy day at EWR in 2007 in a series of tests involving different realistic values of the slot limits. It was shown, that for all these slot limits there exists a feasible schedule that may accommodate all the daily traffic (even when the slots are set as low as the IFR capacity of the airport), while respecting all constraints related to aircraft itineraries and passenger connections. Furthermore, it was observed that with slot limits near the IFR capacity of the airport, the maximum flight displacement was only 30 minutes while a total of 491 flights were displaced with an average displacement of 20 minutes.

We also estimated the delays associated with three different levels of slot controls at EWR using the AND model—the first time that an airport network model is used to estimate the local and system-wide effects of introducing slot controls at a busy airport. It was shown that, when setting the slot limits near the IFR capacity of EWR, the delay savings due to local congestion at EWR of the order of 10% are for arrivals and of the order of 50% for departures, when EWR operates under VFR conditions. Similarly, under IFR conditions the reduction of delay is 7% for arrivals and 30% for departures. When also considering the upstream delay carried by flights while executing their daily schedules, the average flight arrival delay at EWR is reduced by 22% when EWR operates under VFR conditions. The system-wide delay savings from the application of slot limits at EWR was estimated to be of the order
of 20% (a reduction of roughly 45 aircraft-hours daily). Finally, it was shown that the combination of switching runway configurations at EWR earlier in the afternoon (compared to what is done today) and the application of slot controls, results in a further drop of the arrival delay at EWR, by as much as 32%. Overall, it was shown that by smoothing the demand at one busy US airport—which can be achieved with relatively minor modifications of the current schedule without eliminating any flights—great delay savings can be achieved both at a local and a system-wide level.

6.2 Future Research

This section identifies some of the most important possible directions of research that would extend the products and findings of this thesis.

Airline reaction models

One of the main limitations of AND is that it does not capture airline reactions to congestion. Typically, when airlines experience heavy congestion, they take a number of actions to recover from "irregular operations". These include:

- Flight cancellations.
- Utilization of spare aircraft at hubs.
- Swapping of aircraft assignments to flights and re-routing some aircraft as a result.
- Crew substitutions.

These actions are meant to reduce the delay experienced by an airline during a period of extreme congestion. By not modeling these real-time interventions, AND tends to overestimate delays on days of extreme congestion. The addition of an airline recovery model, will improve the capabilities of AND, as long as it does not impact greatly the computational performance and ease of execution of the original model. With the addition of this feature, AND could also be used to test different airline recovery
optimization models, or the robustness of airline schedules to poor weather and other disturbances.

Passenger delays estimation model

The delays experienced by passengers have been shown by recent research ([4] and [54]) to be greater than flight delays; the main causes of this are flight cancellations, that leave passengers waiting until the next available flight, and missed connections of passengers on multi-leg itineraries. Vaze and Fearing [4] have developed a passenger delay estimator that consists of a discreet choice model, which is used to estimate historical passenger itineraries, and a heuristic algorithm used to re-accommodate disrupted passengers.

An integrated model that incorporates this passenger delay estimator within AND would provide estimates of passenger misconnections and re-bookings based on the flight delay estimates of AND and subsequently generate estimates of passenger delays. In addition, such a model would provide the capability of testing the network-wide impacts of approaches for reducing passenger delays, e.g. by delaying flight departures in order to await passengers from delayed arrivals. It would also provide a metric for a flight cancellation model within AND. In relation to the latter, Vaze and Fearing [4] have shown that both the number of passengers and the number of connecting passengers on a flight have a negative correlation with the probability of that flight being cancelled.

By running the integrated model we would be able to identify the airports that currently account for most of the passenger delays in the United States. It would also be possible to perform various case studies, such as estimating the impact of airport capacity changes at busy airports, not only on flight delays and delay propagation, but also, on passenger delays.

We have already performed some preliminary research on this topic. As described in Chapter 2, AND estimates the probability density function (PDF) of delays per airport and time of day. In our preliminary work we have compared two methods for establishing whether a passenger is misconnected or not. In the first we use
the expected value of flight delays to determine whether a passenger has missed a connection between two flights. In the second method we estimate the probability of a misconnection by convolving the PDFs of the delays of the two flights between which a passenger is expected to connect. Further research is necessary, however, in order to incorporate correctly the passenger delay estimator within AND.

Models of en-route congestion

As the demand for air travel increases, several parts of the United States airspace may also become congested and a queuing model that will predict delays due to en route congestion may become necessary, even at the macroscopic analysis level. In Europe, a significant portion of the arrival delay at airports is already attributed to en route congestion due to the complicated structure of the European airspace and the fact that major air routes are more concentrated than in the US. Research in this area should include the development of queuing models that predict en route delays and their validation, in a similar way as described in this thesis for the airport departure model in Section 3.2. Moreover, this topic will require extensive data mining in order to segregate delays attributed to airport congestion from delays attributed to airspace congestion.

Slot control at US airports

In the United States, airports operate mostly without slot constraints on a first-come, first-served basis with only occasional prior coordination (mainly through Ground Delay Programs). On the other hand, European airports have declared capacities that are set close to their poor-weather capacities and assign slots to airlines accordingly. The US practice leads to high congestion and high delays even on good weather days, while the European system may be underutilizing capacity under visual meteorological conditions.

As we showed in Chapter 5, there is the potential for great delay savings from the implementation of slot controls at some busy airports in the US. But there is still the major question of how to implement efficiently slot controls. In that respect, a possible
future research direction is the development of stochastic optimization models that
generate vectors of daily arrival/departure slots that minimize the expected delays
while maximizing the expected efficiency (throughput) across all airport conditions
(weather, runway configurations, noise restrictions etc.). This is a vast topic that must
include consideration of difficult trade-offs, as well as of the, often divergent, interests
of the several types of airport stakeholders (airlines, passengers, airport operators, air
navigation service providers, central and regional governments).
Appendix A

Estimated Minimum Turnaround Times
Table A.1: Average minimum turn times by airline, by aircraft type and by airport type

<table>
<thead>
<tr>
<th>Airline</th>
<th>Aircraft</th>
<th>Airport Type</th>
<th>Estimated Min. Turnaround Time</th>
<th># of flights in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A320</td>
<td>H</td>
<td>57</td>
<td>2842</td>
</tr>
<tr>
<td>B6</td>
<td>A320</td>
<td>NH</td>
<td>43</td>
<td>6518</td>
</tr>
<tr>
<td>B6</td>
<td>E190</td>
<td>H</td>
<td>39</td>
<td>1411</td>
</tr>
<tr>
<td>B6</td>
<td>E190</td>
<td>NH</td>
<td>31</td>
<td>1885</td>
</tr>
<tr>
<td>WN</td>
<td>B737-300</td>
<td>H</td>
<td>22</td>
<td>901</td>
</tr>
<tr>
<td>WN</td>
<td>B737-300</td>
<td>NH</td>
<td>23</td>
<td>14991</td>
</tr>
<tr>
<td>WN</td>
<td>B737-500</td>
<td>H</td>
<td>22</td>
<td>165</td>
</tr>
<tr>
<td>WN</td>
<td>B737-500</td>
<td>NH</td>
<td>22</td>
<td>2643</td>
</tr>
<tr>
<td>WN</td>
<td>B737-700</td>
<td>H</td>
<td>24</td>
<td>1697</td>
</tr>
<tr>
<td>WN</td>
<td>B737-700</td>
<td>NH</td>
<td>24</td>
<td>13417</td>
</tr>
<tr>
<td>F9</td>
<td>B717-200</td>
<td>H</td>
<td>46</td>
<td>2736</td>
</tr>
<tr>
<td>F9</td>
<td>B717-200</td>
<td>NH</td>
<td>29</td>
<td>5848</td>
</tr>
<tr>
<td>F9</td>
<td>B737-700</td>
<td>H</td>
<td>51</td>
<td>1498</td>
</tr>
<tr>
<td>F9</td>
<td>B737-700</td>
<td>NH</td>
<td>34</td>
<td>2862</td>
</tr>
<tr>
<td>DL</td>
<td>B737-800</td>
<td>H</td>
<td>52</td>
<td>1106</td>
</tr>
<tr>
<td>DL</td>
<td>B737-800</td>
<td>NH</td>
<td>43</td>
<td>1621</td>
</tr>
<tr>
<td>DL</td>
<td>B757-200</td>
<td>H</td>
<td>62</td>
<td>2981</td>
</tr>
<tr>
<td>DL</td>
<td>B757-200</td>
<td>NH</td>
<td>47</td>
<td>3046</td>
</tr>
<tr>
<td>DL</td>
<td>B767-300</td>
<td>H</td>
<td>71</td>
<td>475</td>
</tr>
<tr>
<td>DL</td>
<td>B767-300</td>
<td>NH</td>
<td>55</td>
<td>665</td>
</tr>
<tr>
<td>DL</td>
<td>B767-400</td>
<td>H</td>
<td>74</td>
<td>121</td>
</tr>
<tr>
<td>DL</td>
<td>B767-400</td>
<td>NH</td>
<td>63</td>
<td>363</td>
</tr>
<tr>
<td>DL</td>
<td>MD88</td>
<td>H</td>
<td>51</td>
<td>3941</td>
</tr>
<tr>
<td>DL</td>
<td>MD88</td>
<td>NH</td>
<td>36</td>
<td>5666</td>
</tr>
<tr>
<td>DL</td>
<td>MD90</td>
<td>H</td>
<td>44</td>
<td>440</td>
</tr>
<tr>
<td>DL</td>
<td>MD90</td>
<td>NH</td>
<td>35</td>
<td>408</td>
</tr>
</tbody>
</table>
Table A.2: Average minimum turn times by airline, by aircraft type and by airport type

<table>
<thead>
<tr>
<th>Airline</th>
<th>Aircraft</th>
<th>Airport Type</th>
<th>Estimated Min. Turnaround Time (mins)</th>
<th># of flights in sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>B737-300</td>
<td>H</td>
<td>51</td>
<td>1948</td>
</tr>
<tr>
<td>CO</td>
<td>B737-300</td>
<td>NH</td>
<td>37</td>
<td>1958</td>
</tr>
<tr>
<td>CO</td>
<td>B737-500</td>
<td>H</td>
<td>45</td>
<td>4579</td>
</tr>
<tr>
<td>CO</td>
<td>B737-500</td>
<td>NH</td>
<td>35</td>
<td>3009</td>
</tr>
<tr>
<td>CO</td>
<td>B737-700</td>
<td>H</td>
<td>50</td>
<td>961</td>
</tr>
<tr>
<td>CO</td>
<td>B737-700</td>
<td>NH</td>
<td>39</td>
<td>1248</td>
</tr>
<tr>
<td>CO</td>
<td>B737-800</td>
<td>H</td>
<td>50</td>
<td>961</td>
</tr>
<tr>
<td>CO</td>
<td>B737-800</td>
<td>NH</td>
<td>44</td>
<td>3394</td>
</tr>
<tr>
<td>CO</td>
<td>B737-900</td>
<td>H</td>
<td>58</td>
<td>474</td>
</tr>
<tr>
<td>CO</td>
<td>B737-900</td>
<td>NH</td>
<td>45</td>
<td>580</td>
</tr>
<tr>
<td>CO</td>
<td>B757-200</td>
<td>H</td>
<td>65</td>
<td>288</td>
</tr>
<tr>
<td>CO</td>
<td>B757-200</td>
<td>NH</td>
<td>50</td>
<td>870</td>
</tr>
<tr>
<td>CO</td>
<td>B757-300</td>
<td>H</td>
<td>71</td>
<td>800</td>
</tr>
<tr>
<td>CO</td>
<td>B757-300</td>
<td>NH</td>
<td>57</td>
<td>1379</td>
</tr>
<tr>
<td>CO</td>
<td>B767-400</td>
<td>H</td>
<td>92</td>
<td>23</td>
</tr>
<tr>
<td>CO</td>
<td>B767-400</td>
<td>NH</td>
<td>74</td>
<td>77</td>
</tr>
<tr>
<td>UA</td>
<td>A319</td>
<td>H</td>
<td>43</td>
<td>1773</td>
</tr>
<tr>
<td>UA</td>
<td>A319</td>
<td>NH</td>
<td>37</td>
<td>3257</td>
</tr>
<tr>
<td>UA</td>
<td>A320</td>
<td>H</td>
<td>44</td>
<td>2834</td>
</tr>
<tr>
<td>UA</td>
<td>A320</td>
<td>NH</td>
<td>38</td>
<td>6087</td>
</tr>
<tr>
<td>UA</td>
<td>B737-300</td>
<td>H</td>
<td>42</td>
<td>3443</td>
</tr>
<tr>
<td>UA</td>
<td>B737-300</td>
<td>NH</td>
<td>34</td>
<td>4420</td>
</tr>
<tr>
<td>UA</td>
<td>B737-500</td>
<td>H</td>
<td>41</td>
<td>1363</td>
</tr>
<tr>
<td>UA</td>
<td>B737-500</td>
<td>NH</td>
<td>33</td>
<td>2106</td>
</tr>
<tr>
<td>UA</td>
<td>B747-400</td>
<td>H/NH</td>
<td>66</td>
<td>10</td>
</tr>
<tr>
<td>UA</td>
<td>B757-200</td>
<td>H</td>
<td>52</td>
<td>2970</td>
</tr>
<tr>
<td>UA</td>
<td>B757-200</td>
<td>NH</td>
<td>49</td>
<td>4476</td>
</tr>
<tr>
<td>UA</td>
<td>B767-300</td>
<td>H</td>
<td>64</td>
<td>269</td>
</tr>
<tr>
<td>UA</td>
<td>B767-300</td>
<td>NH</td>
<td>63</td>
<td>523</td>
</tr>
<tr>
<td>UA</td>
<td>B777-200</td>
<td>H</td>
<td>77</td>
<td>85</td>
</tr>
<tr>
<td>UA</td>
<td>B777-200</td>
<td>NH</td>
<td>64</td>
<td>148</td>
</tr>
</tbody>
</table>
Appendix B

Estimated Nominal Block Time
Table B.1: Nominal block time components by route (in minutes), as estimated by Skaltsas [50]

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>Nominal Taxi-out</th>
<th>Nominal Flight</th>
<th>Nominal Taxi-in</th>
<th>Nominal Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>LGA</td>
<td>12</td>
<td>95</td>
<td>6</td>
<td>113</td>
</tr>
<tr>
<td>BOS</td>
<td>LGA</td>
<td>13</td>
<td>40</td>
<td>6</td>
<td>58</td>
</tr>
<tr>
<td>BWI</td>
<td>LGA</td>
<td>9</td>
<td>45</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>CLE</td>
<td>LGA</td>
<td>12</td>
<td>64</td>
<td>6</td>
<td>81</td>
</tr>
<tr>
<td>CLT</td>
<td>LGA</td>
<td>11</td>
<td>72</td>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>CVG</td>
<td>LGA</td>
<td>13</td>
<td>83</td>
<td>7</td>
<td>104</td>
</tr>
<tr>
<td>DCA</td>
<td>LGA</td>
<td>11</td>
<td>39</td>
<td>5</td>
<td>55</td>
</tr>
<tr>
<td>DEN</td>
<td>LGA</td>
<td>11</td>
<td>180</td>
<td>5</td>
<td>196</td>
</tr>
<tr>
<td>DFW</td>
<td>LGA</td>
<td>11</td>
<td>164</td>
<td>5</td>
<td>180</td>
</tr>
<tr>
<td>DTW</td>
<td>LGA</td>
<td>15</td>
<td>70</td>
<td>7</td>
<td>91</td>
</tr>
<tr>
<td>FLL</td>
<td>LGA</td>
<td>13</td>
<td>134</td>
<td>5</td>
<td>153</td>
</tr>
<tr>
<td>IAD</td>
<td>LGA</td>
<td>14</td>
<td>41</td>
<td>5</td>
<td>61</td>
</tr>
<tr>
<td>IAH</td>
<td>LGA</td>
<td>14</td>
<td>169</td>
<td>6</td>
<td>188</td>
</tr>
<tr>
<td>MCI</td>
<td>LGA</td>
<td>10</td>
<td>133</td>
<td>6</td>
<td>148</td>
</tr>
<tr>
<td>MCO</td>
<td>LGA</td>
<td>12</td>
<td>120</td>
<td>6</td>
<td>138</td>
</tr>
<tr>
<td>MDW</td>
<td>LGA</td>
<td>8</td>
<td>93</td>
<td>6</td>
<td>107</td>
</tr>
<tr>
<td>MEM</td>
<td>LGA</td>
<td>14</td>
<td>120</td>
<td>8</td>
<td>142</td>
</tr>
<tr>
<td>MIA</td>
<td>LGA</td>
<td>13</td>
<td>137</td>
<td>5</td>
<td>155</td>
</tr>
<tr>
<td>MSP</td>
<td>LGA</td>
<td>13</td>
<td>123</td>
<td>8</td>
<td>144</td>
</tr>
<tr>
<td>ORD</td>
<td>LGA</td>
<td>12</td>
<td>91</td>
<td>4</td>
<td>107</td>
</tr>
<tr>
<td>PIT</td>
<td>LGA</td>
<td>11</td>
<td>54</td>
<td>6</td>
<td>71</td>
</tr>
<tr>
<td>STL</td>
<td>LGA</td>
<td>11</td>
<td>111</td>
<td>5</td>
<td>127</td>
</tr>
<tr>
<td>ATL</td>
<td>DCA</td>
<td>12</td>
<td>71</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>BOS</td>
<td>DCA</td>
<td>13</td>
<td>66</td>
<td>4</td>
<td>82</td>
</tr>
<tr>
<td>CLE</td>
<td>DCA</td>
<td>11</td>
<td>49</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>CLT</td>
<td>DCA</td>
<td>11</td>
<td>49</td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>CVG</td>
<td>DCA</td>
<td>13</td>
<td>56</td>
<td>5</td>
<td>74</td>
</tr>
<tr>
<td>DEN</td>
<td>DCA</td>
<td>11</td>
<td>162</td>
<td>5</td>
<td>178</td>
</tr>
<tr>
<td>DFW</td>
<td>DCA</td>
<td>12</td>
<td>136</td>
<td>4</td>
<td>151</td>
</tr>
<tr>
<td>DFW</td>
<td>DCA</td>
<td>15</td>
<td>58</td>
<td>6</td>
<td>79</td>
</tr>
<tr>
<td>EWR</td>
<td>DCA</td>
<td>12</td>
<td>40</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>FLL</td>
<td>DCA</td>
<td>12</td>
<td>113</td>
<td>4</td>
<td>129</td>
</tr>
<tr>
<td>IAH</td>
<td>DCA</td>
<td>14</td>
<td>140</td>
<td>4</td>
<td>159</td>
</tr>
<tr>
<td>JFK</td>
<td>DCA</td>
<td>19</td>
<td>46</td>
<td>4</td>
<td>68</td>
</tr>
<tr>
<td>LAS</td>
<td>DCA</td>
<td>14</td>
<td>225</td>
<td>4</td>
<td>242</td>
</tr>
<tr>
<td>LAX</td>
<td>DCA</td>
<td>12</td>
<td>255</td>
<td>4</td>
<td>271</td>
</tr>
<tr>
<td>LGA</td>
<td>DCA</td>
<td>10</td>
<td>39</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>MCI</td>
<td>DCA</td>
<td>10</td>
<td>109</td>
<td>4</td>
<td>123</td>
</tr>
<tr>
<td>MCO</td>
<td>DCA</td>
<td>11</td>
<td>96</td>
<td>4</td>
<td>110</td>
</tr>
<tr>
<td>MEM</td>
<td>DCA</td>
<td>14</td>
<td>90</td>
<td>6</td>
<td>110</td>
</tr>
<tr>
<td>MIA</td>
<td>DCA</td>
<td>13</td>
<td>114</td>
<td>4</td>
<td>131</td>
</tr>
<tr>
<td>MSP</td>
<td>DCA</td>
<td>13</td>
<td>109</td>
<td>6</td>
<td>128</td>
</tr>
<tr>
<td>ORD</td>
<td>DCA</td>
<td>12</td>
<td>75</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>PHL</td>
<td>DCA</td>
<td>12</td>
<td>26</td>
<td>4</td>
<td>42</td>
</tr>
<tr>
<td>PHX</td>
<td>DCA</td>
<td>12</td>
<td>215</td>
<td>4</td>
<td>230</td>
</tr>
<tr>
<td>SEA</td>
<td>DCA</td>
<td>14</td>
<td>256</td>
<td>4</td>
<td>273</td>
</tr>
<tr>
<td>SLC</td>
<td>DCA</td>
<td>15</td>
<td>203</td>
<td>4</td>
<td>223</td>
</tr>
<tr>
<td>STL</td>
<td>DCA</td>
<td>11</td>
<td>87</td>
<td>4</td>
<td>102</td>
</tr>
<tr>
<td>TPA</td>
<td>DCA</td>
<td>9</td>
<td>101</td>
<td>4</td>
<td>114</td>
</tr>
</tbody>
</table>
Table B.2: Nominal block time components by route (in minutes), as estimated by Skaltsas [50]

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>Nominal Taxi-out</th>
<th>Nominal Flight</th>
<th>Nominal Taxi-in</th>
<th>Nominal Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATL</td>
<td>EWR</td>
<td>14</td>
<td>95</td>
<td>7</td>
<td>115</td>
</tr>
<tr>
<td>BOS</td>
<td>EWR</td>
<td>15</td>
<td>46</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>BWI</td>
<td>EWR</td>
<td>10</td>
<td>33</td>
<td>7</td>
<td>50</td>
</tr>
<tr>
<td>CLE</td>
<td>EWR</td>
<td>12</td>
<td>59</td>
<td>8</td>
<td>79</td>
</tr>
<tr>
<td>CLT</td>
<td>EWR</td>
<td>12</td>
<td>73</td>
<td>6</td>
<td>91</td>
</tr>
<tr>
<td>CVG</td>
<td>EWR</td>
<td>12</td>
<td>84</td>
<td>7</td>
<td>103</td>
</tr>
<tr>
<td>DCA</td>
<td>EWR</td>
<td>11</td>
<td>39</td>
<td>7</td>
<td>57</td>
</tr>
<tr>
<td>DEN</td>
<td>EWR</td>
<td>12</td>
<td>178</td>
<td>7</td>
<td>197</td>
</tr>
<tr>
<td>DFW</td>
<td>EWR</td>
<td>13</td>
<td>165</td>
<td>7</td>
<td>184</td>
</tr>
<tr>
<td>DTW</td>
<td>EWR</td>
<td>15</td>
<td>68</td>
<td>7</td>
<td>90</td>
</tr>
<tr>
<td>FLL</td>
<td>EWR</td>
<td>14</td>
<td>137</td>
<td>7</td>
<td>158</td>
</tr>
<tr>
<td>IAD</td>
<td>EWR</td>
<td>14</td>
<td>45</td>
<td>7</td>
<td>66</td>
</tr>
<tr>
<td>IAH</td>
<td>EWR</td>
<td>14</td>
<td>164</td>
<td>8</td>
<td>187</td>
</tr>
<tr>
<td>LAS</td>
<td>EWR</td>
<td>11</td>
<td>241</td>
<td>8</td>
<td>260</td>
</tr>
<tr>
<td>LAX</td>
<td>EWR</td>
<td>12</td>
<td>268</td>
<td>8</td>
<td>288</td>
</tr>
<tr>
<td>MCI</td>
<td>EWR</td>
<td>10</td>
<td>131</td>
<td>7</td>
<td>147</td>
</tr>
<tr>
<td>MCO</td>
<td>EWR</td>
<td>12</td>
<td>119</td>
<td>7</td>
<td>138</td>
</tr>
<tr>
<td>MEM</td>
<td>EWR</td>
<td>12</td>
<td>124</td>
<td>7</td>
<td>143</td>
</tr>
<tr>
<td>MIA</td>
<td>EWR</td>
<td>14</td>
<td>139</td>
<td>7</td>
<td>161</td>
</tr>
<tr>
<td>MSP</td>
<td>EWR</td>
<td>13</td>
<td>122</td>
<td>7</td>
<td>142</td>
</tr>
<tr>
<td>ORD</td>
<td>EWR</td>
<td>12</td>
<td>89</td>
<td>7</td>
<td>108</td>
</tr>
<tr>
<td>PHX</td>
<td>EWR</td>
<td>11</td>
<td>236</td>
<td>7</td>
<td>253</td>
</tr>
<tr>
<td>PIT</td>
<td>EWR</td>
<td>11</td>
<td>60</td>
<td>7</td>
<td>79</td>
</tr>
<tr>
<td>SAN</td>
<td>EWR</td>
<td>13</td>
<td>264</td>
<td>8</td>
<td>285</td>
</tr>
<tr>
<td>SEA</td>
<td>EWR</td>
<td>12</td>
<td>265</td>
<td>8</td>
<td>285</td>
</tr>
<tr>
<td>SFO</td>
<td>EWR</td>
<td>12</td>
<td>276</td>
<td>7</td>
<td>295</td>
</tr>
<tr>
<td>SLC</td>
<td>EWR</td>
<td>15</td>
<td>218</td>
<td>6</td>
<td>240</td>
</tr>
<tr>
<td>STL</td>
<td>EWR</td>
<td>10</td>
<td>112</td>
<td>7</td>
<td>129</td>
</tr>
<tr>
<td>BOS</td>
<td>JFK</td>
<td>14</td>
<td>41</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>ATL</td>
<td>JFK</td>
<td>13</td>
<td>97</td>
<td>9</td>
<td>119</td>
</tr>
<tr>
<td>BWI</td>
<td>JFK</td>
<td>12</td>
<td>37</td>
<td>8</td>
<td>57</td>
</tr>
<tr>
<td>CLE</td>
<td>JFK</td>
<td>14</td>
<td>66</td>
<td>8</td>
<td>88</td>
</tr>
<tr>
<td>CLT</td>
<td>JFK</td>
<td>14</td>
<td>75</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>CVG</td>
<td>JFK</td>
<td>13</td>
<td>88</td>
<td>9</td>
<td>109</td>
</tr>
<tr>
<td>DCA</td>
<td>JFK</td>
<td>11</td>
<td>40</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td>DEN</td>
<td>JFK</td>
<td>13</td>
<td>183</td>
<td>7</td>
<td>203</td>
</tr>
<tr>
<td>DFW</td>
<td>JFK</td>
<td>13</td>
<td>169</td>
<td>8</td>
<td>191</td>
</tr>
<tr>
<td>DTW</td>
<td>JFK</td>
<td>15</td>
<td>72</td>
<td>9</td>
<td>96</td>
</tr>
<tr>
<td>FLL</td>
<td>JFK</td>
<td>14</td>
<td>131</td>
<td>7</td>
<td>152</td>
</tr>
<tr>
<td>HOU</td>
<td>JFK</td>
<td>12</td>
<td>171</td>
<td>6</td>
<td>189</td>
</tr>
<tr>
<td>IAD</td>
<td>JFK</td>
<td>15</td>
<td>44</td>
<td>6</td>
<td>65</td>
</tr>
<tr>
<td>IAH</td>
<td>JFK</td>
<td>14</td>
<td>169</td>
<td>9</td>
<td>192</td>
</tr>
<tr>
<td>LAS</td>
<td>JFK</td>
<td>13</td>
<td>248</td>
<td>7</td>
<td>269</td>
</tr>
<tr>
<td>LAX</td>
<td>JFK</td>
<td>14</td>
<td>272</td>
<td>8</td>
<td>293</td>
</tr>
<tr>
<td>MCO</td>
<td>JFK</td>
<td>12</td>
<td>117</td>
<td>7</td>
<td>135</td>
</tr>
<tr>
<td>MIA</td>
<td>JFK</td>
<td>14</td>
<td>133</td>
<td>9</td>
<td>155</td>
</tr>
<tr>
<td>MSP</td>
<td>JFK</td>
<td>13</td>
<td>124</td>
<td>9</td>
<td>146</td>
</tr>
<tr>
<td>ORD</td>
<td>JFK</td>
<td>15</td>
<td>94</td>
<td>7</td>
<td>116</td>
</tr>
</tbody>
</table>
Table B.3: Nominal block time components by route (in minutes), as estimated by Skaltsas [50]

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>Nominal Taxi-out</th>
<th>Nominal Flight</th>
<th>Nominal Taxi-in</th>
<th>Nominal Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHL</td>
<td>JFK</td>
<td>14</td>
<td>26</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>PHX</td>
<td>JFK</td>
<td>12</td>
<td>239</td>
<td>7</td>
<td>259</td>
</tr>
<tr>
<td>PIT</td>
<td>JFK</td>
<td>13</td>
<td>60</td>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>SAN</td>
<td>JFK</td>
<td>14</td>
<td>270</td>
<td>7</td>
<td>291</td>
</tr>
<tr>
<td>SEA</td>
<td>JFK</td>
<td>15</td>
<td>269</td>
<td>8</td>
<td>292</td>
</tr>
<tr>
<td>SFO</td>
<td>JFK</td>
<td>14</td>
<td>282</td>
<td>8</td>
<td>303</td>
</tr>
<tr>
<td>SLC</td>
<td>JFK</td>
<td>15</td>
<td>219</td>
<td>8</td>
<td>243</td>
</tr>
<tr>
<td>STL</td>
<td>JFK</td>
<td>13</td>
<td>119</td>
<td>9</td>
<td>140</td>
</tr>
<tr>
<td>TPA</td>
<td>JFK</td>
<td>12</td>
<td>124</td>
<td>7</td>
<td>143</td>
</tr>
<tr>
<td>ATL</td>
<td>ORD</td>
<td>13</td>
<td>87</td>
<td>6</td>
<td>106</td>
</tr>
<tr>
<td>BOS</td>
<td>ORD</td>
<td>12</td>
<td>124</td>
<td>6</td>
<td>142</td>
</tr>
<tr>
<td>BWI</td>
<td>ORD</td>
<td>10</td>
<td>94</td>
<td>5</td>
<td>110</td>
</tr>
<tr>
<td>CLE</td>
<td>ORD</td>
<td>12</td>
<td>53</td>
<td>6</td>
<td>71</td>
</tr>
<tr>
<td>CLT</td>
<td>ORD</td>
<td>12</td>
<td>88</td>
<td>6</td>
<td>105</td>
</tr>
<tr>
<td>CVG</td>
<td>ORD</td>
<td>13</td>
<td>45</td>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>DCA</td>
<td>ORD</td>
<td>11</td>
<td>90</td>
<td>5</td>
<td>107</td>
</tr>
<tr>
<td>DEN</td>
<td>ORD</td>
<td>11</td>
<td>104</td>
<td>5</td>
<td>121</td>
</tr>
<tr>
<td>DEN</td>
<td>ORD</td>
<td>11</td>
<td>99</td>
<td>6</td>
<td>116</td>
</tr>
<tr>
<td>DFW</td>
<td>ORD</td>
<td>14</td>
<td>43</td>
<td>6</td>
<td>62</td>
</tr>
<tr>
<td>DTW</td>
<td>ORD</td>
<td>13</td>
<td>107</td>
<td>6</td>
<td>125</td>
</tr>
<tr>
<td>EWR</td>
<td>ORD</td>
<td>13</td>
<td>156</td>
<td>6</td>
<td>174</td>
</tr>
<tr>
<td>FLL</td>
<td>ORD</td>
<td>13</td>
<td>86</td>
<td>5</td>
<td>104</td>
</tr>
<tr>
<td>IAD</td>
<td>ORD</td>
<td>13</td>
<td>115</td>
<td>6</td>
<td>135</td>
</tr>
<tr>
<td>IAH</td>
<td>ORD</td>
<td>13</td>
<td>112</td>
<td>6</td>
<td>139</td>
</tr>
<tr>
<td>JFK</td>
<td>ORD</td>
<td>13</td>
<td>169</td>
<td>6</td>
<td>187</td>
</tr>
<tr>
<td>LAS</td>
<td>ORD</td>
<td>13</td>
<td>195</td>
<td>6</td>
<td>213</td>
</tr>
<tr>
<td>LAX</td>
<td>ORD</td>
<td>12</td>
<td>108</td>
<td>6</td>
<td>126</td>
</tr>
<tr>
<td>LGA</td>
<td>ORD</td>
<td>12</td>
<td>137</td>
<td>6</td>
<td>155</td>
</tr>
<tr>
<td>MCO</td>
<td>ORD</td>
<td>12</td>
<td>74</td>
<td>6</td>
<td>93</td>
</tr>
<tr>
<td>MEM</td>
<td>ORD</td>
<td>13</td>
<td>153</td>
<td>6</td>
<td>173</td>
</tr>
<tr>
<td>MIA</td>
<td>ORD</td>
<td>13</td>
<td>50</td>
<td>6</td>
<td>68</td>
</tr>
<tr>
<td>MSP</td>
<td>ORD</td>
<td>12</td>
<td>99</td>
<td>6</td>
<td>117</td>
</tr>
<tr>
<td>PHL</td>
<td>ORD</td>
<td>11</td>
<td>161</td>
<td>6</td>
<td>178</td>
</tr>
<tr>
<td>PHX</td>
<td>ORD</td>
<td>11</td>
<td>64</td>
<td>5</td>
<td>81</td>
</tr>
<tr>
<td>PIT</td>
<td>ORD</td>
<td>12</td>
<td>194</td>
<td>6</td>
<td>211</td>
</tr>
<tr>
<td>SAN</td>
<td>ORD</td>
<td>13</td>
<td>193</td>
<td>6</td>
<td>211</td>
</tr>
<tr>
<td>SEA</td>
<td>ORD</td>
<td>14</td>
<td>201</td>
<td>5</td>
<td>221</td>
</tr>
<tr>
<td>SFO</td>
<td>ORD</td>
<td>13</td>
<td>144</td>
<td>6</td>
<td>164</td>
</tr>
<tr>
<td>SLC</td>
<td>ORD</td>
<td>10</td>
<td>41</td>
<td>6</td>
<td>57</td>
</tr>
<tr>
<td>STL</td>
<td>ORD</td>
<td>11</td>
<td>135</td>
<td>5</td>
<td>151</td>
</tr>
</tbody>
</table>
Bibliography


223


[52] U.S. Congress Joint Economic Committee. Your flight has been delayed again: Flight delays cost passengers, airlines, and the u.s. economy billions, 2008. Washington, DC.


