INDUSTRY STUDIES ASSOCIATION
WORKING PAPER SERIES

Analysis of the Potential for Delay Propagation in Passenger Airline Networks

By

Amy Cohn
Global Airline Industry Program
Massachusetts Institute of Technology
Cambridge, MA 02139
and
University of Michigan
Ann Arbor, MI 48109

Shervin AhmadBeygi
Yihan Guan
University of Michigan
Ann Arbor, MI 48109

Peter Belobaba
Global Airline Industry Program
Massachusetts Institute of Technology
Cambridge, MA 02139

2007
Industry Studies Association
Working Papers

WP-2007-11
http://isapapers.pitt.edu/

Analysis of the Potential for Delay Propagation in Passenger Airline Networks

Shervin AhmadBeygi, Amy Cohn, and Yihan Guan
University of Michigan
Peter Belobaba
Massachusetts Institute of Technology

July 7, 2007

Abstract

In this paper, we analyze the potential for delays to propagate in passenger airline networks. The motivation for this research is the need to better understand the relationship between the scheduling of aircraft and crew members, and the operational performance of such schedules. In particular, when carriers decide how to schedule these costly resources, the focus is primarily on achieving high levels of utilization. The resulting plans, however, often have little slack, limiting the schedule's ability to absorb disruption; instead, initial flight delays may propagate to delay subsequent flights as well. Understanding the relationship between planned schedules and delay propagation is a requisite precursor to developing tools for building more robust airline plans. In this paper, we investigate this relationship using flight data provided by two major U.S. carriers, one traditional hub-and-spoke and one "low-fare" carrier operating a predominantly point-to-point network.

1 Introduction

The operation of a passenger airline requires the allocation of resources and development of schedule plans over complex networks. A large airline can operate over a thousand flight departures per day with several hundred aircraft and thousands of cockpit and cabin crew employees. These resources are costly – the direct operating costs of a 183-seat Boeing 757 aircraft were over $4300 per hour in 2006. The efficient utilization of costly resources is thus one of the key challenges faced by airlines hoping to control operating expenses in order to generate profits in an increasingly competitive fare environmen ([6]).

The operations research (OR) community has played a significant role in developing airline planning tools with the aim of optimizing the utilization of these costly resources ([1], [2]). In these optimization tools, which typically assume deterministic flight times and other parameters, it is desirable to generate schedules that have little if any slack between flights – by *turning* crews and aircraft quickly, (i.e. having them connect from one flight to another with minimal time in between) greater utilization of these resources can be achieved and unit operating costs can be minimized.

In practice, however, flight times are *not* deterministic. Departure delays arise due to mechanical problems, weather delays, ground holds, and other sources. Flights that depart on time can still be delayed in arrival due to causes such as air traffic control issues or re-routings to avoid inclement weather. In isolation, these delays are themselves costly. In a network structure, they can have an even greater impact – without adequate slack to absorb an initial *root delay*, subsequent flights may also be delayed as they await aircraft and crews from the initially delayed flight. We refer to this as *delay propagation*.

In fact, one would expect to see an inverse relationship between the planned level of resource utilization in a schedule and that schedule's operational robustness. In a planned schedule with high resource utilization, there is limited slack. This in turn limits the opportunity to absorb flight delays, which must instead propagate to subsequent flights. But what is the nature of this relationship? How can it be incorporated in the planning process to produce "better" schedules? And what constitutes "better" – how should the deterministic costs of an airline plan be traded off against the potential and much more uncertain costs of delay? These are all important and challenging questions that are beginning to receive significant attention from both the airline industry and the academic community.

To assist in these efforts, we have undertaken an empirical study of passenger airline flight networks and their potential for delay propagation as a function of their planned schedule. This study is based on flight data from two U.S. carriers, one traditional "legacy" hub-and-spoke carrier and one "low-fare" carrier operating a predominantly point-to-point network. Using the idea of *propagation trees* as our foundation, we examine how any single given flight delay, in the absence of other flight delays, can propagate through the network (the structure is a tree because one flight uses multiple resources, such as cockpit crews and aircraft, and therefore each flight delay can directly cause multiple subsequent delays, which in turn can continue to branch further). We analyze these trees for all flights in a given time period in order to gain insight into the distribution of slack throughout the system and the implications of this on delay propagation. We then use this analysis to address commonly-held assumptions about how delays propagate, provide insights into the relationship between planned resource utilization and operational robustness, and raise questions for further study.

This research makes several important contributions towards understanding the relationship between planned airline schedules and the operational performance of these schedules, a requisite precursor to developing more robust plans. First, we introduce the use of propagation trees as a visual and quantitative tool for assessing the ramifications of individual flight delays throughout the network. Second, we propose several new metrics for quantifying the impact of these initial delays as they propagate through the network. Third, we use these metrics to conduct a quantitative analysis of planned airline schedules, gaining insights into the relationship between scheduled slack and delay propagation. Fourth, we extend this analysis to substantiate (in some cases) and disprove (in other cases) commonly-held assumptions about delay propagation. Finally, we lay the groundwork for future research on incorporating measures of potential delay propagation in the airline schedule planning process.

The paper is organized as follows. In section 2, we outline the details of the study. We present our analysis in section 3. Section 4 contains our conclusions and suggested areas for future research.

2 Analytical Framework

2.1 Motivation

Consider a (hypothetical) airline plan that maximizes resource utilization in the sense that, for every crew and for every aircraft, the time between two consecutive flight assignments is as short as possible. [Depending on the context, this time is referred to as *connection time*, *turn time*, *sit time*, and *ground time*.] There is of course some minimum time between any pair of flight assignments that must observed – for example, an aircraft cannot be assigned to two subsequent flights unless there is adequate time between them for passengers from the first flight to de-plane, catering and cleaning tasks to be completed, and new passengers to board. We begin by assuming, in this hypothetical plan, that all crew and aircraft assignments exactly satisfy this minimum time between assignments. Such a planned schedule is ideal from the perspective that aircraft are being fully utilized and crews are not being paid for any excess time between flights

Suppose further that this schedule is implemented, and that an arbitrary flight is delayed in departure by thirty minutes (for example, due to a mechanical problem that must be fixed before take-off). Assuming that this delay is not compensated for by increasing the travel speed, the flight will have a thirty minute arrival delay as well. Because the cockpit crew does not have any extra slack time before their next flight, this thirty minute delay will propagate to that flight as well, causing it to also be delayed in take-off by thirty minutes. If the crew and the aircraft do not stay together, then the aircraft's next flight will also experience a thirty-minute flight delay. The cabin crew could cause a third thirty-minute flight delay if they separate from the aircraft and cockpit crew. Likewise, if flights are held for connecting passengers from this flight, then these flights will be delayed as well.

Now consider this set of flights that have been delayed as a result of the initial flight delay. [We will refer to the initial flight delay as the the *root delay*; it is also sometimes referred to as the *independent delay* in the literature (Lan et al [8]).] These flights will also arrive late at their destinations, resulting in a second layer of subsequent flight delays. In fact, a delayed aircraft will continue to propagate delay to all of its subsequent flights until the aircraft *goes off-rotation* (i.e. is removed from the flight network to undergo maintenance) or enters an overnight phase where there are no longer flights to be covered. Similarly, a crew (cockpit or cabin) will propagate delay to all of its subsequent flights until they *go off-duty* (i.e. their day's schedule is completed). If the original crews and aircraft do not stay together for subsequent flights, then each of these resources will ultimately cause other resources (i.e. other crews and aircraft) to enter this stream of delays as well.

The situation we present here considers two improbable extremes – on one hand, a schedule without any slack at all and on the other hand a delay cycle that propagates indefinitely. In reality, other factors prevent schedules from fully utilizing all resources to their maximum levels (for example, the market demands that influence flight times and frequencies). Furthermore, recovery alternatives (canceling flights, calling in reserve crews, etc.) frequently prevent delays from propagating fully. Nonetheless, taking into account both objectives – maximizing the profitability of a schedule under ideal conditions; minimizing the propagation of delays in operation – presents an important challenge for airline planners.

It is also a difficult challenge, not only because the planning problems are themselves so complex and because the real-world environment is highly stochastic, but also because measures for quantifying robustness (and the value of this robustness) are not well-defined. In this study, we take an important first step in developing metrics and tools for understanding passenger airline networks and the role of their structures in propagating flight delays.

2.2 Literature Survey

Within the literature, our research is most closely related to the work of Beatty et al [3], who also consider the impact of individual flight delays as they propagate across the network. Their research, which focuses on the relationship between time of day and delay propagation, simulates the movement of delayed cockpit crews, flight attendants, and aircraft through the network. In particular, they use the metric of "delay multiplier"to track the ratio of propagated delay minutes to the length of the initial delay.

More generally, several people have conducted empirical studies to better understand the causes and effects of flight delays. For example, Wang et al [15] use queuing models to examine how the response to propagated delays varies by airport. Queueing models are also used by Janic [7] to quantify the economic consequences of flight delays. Hsiao and Hansen [5] use a statistical model to investigate the impact of different factors such as time of day, congestion, and weather conditions on delay propagation. Tu et al [14] focus specifically on the long-term seasonal behavior of flight delays, while also taking into account additional short-term factors such as time of day. Rupp [10] considers factors such as weather condition, station type (hub vs. spoke), and seasonality to identify the most significant causes of delay.

Efforts to improve the robustness of airline schedules, so as to reduce the potential impacts of disruption, are also beginning to appear in the literature. Ehrgott and Ryan [4] focus on crew scheduling, considering the bi-criteria optimization problem of minimizing cost and maximizing robustness. Their measure of robustness considers the two options of keeping crew and aircraft together or increasing slack between flights when a change of aircraft is scheduled, so as to minimize propagation. Shebalov and Klabjan [12] also consider a bi-criteria crew scheduling problem. In their case, the objectives are minimizing cost and maximizing the number of *move-up crews* (the number of flights for which there exists an alternate crew scheduled for a later departure that could be "moved up" to cover this flight if its scheduled crew is delayed). Schaefer et al [11] instead use an approximation of the *expected operational cost* of a crew pairing in solving the crew scheduling problem as a single-objective optimization problem. Yen and Birge [16] formulate the crew scheduling problem as a stochastic integer program in order to incorporate the impact of delays.

In Rosenberger et al [9], the focus is on incorporating robustness in the fleet assignment problem. By constructing fleet assignments and aircraft rotations with many *short cycles* that could be canceled without violating flow balance, they seek to provide greater opportunity for canceling a limited number of flights without significant network impact. Smith and Johnson [13] also seek to improve robustness in the fleet assignment by imposing *station purity*, a limitation on the number of fleet types that land at/depart from each airport, which in turn provides greater flexibility in recovery operations.

Finally, the work of Lan et al [8] presents an integrated model that simultaneously assigns fleet types and flight departure times, with an emphasis on minimizing the impact of delays on passengers.

We extend this literature by conducting a detailed investigation of the relationship between scheduled plans and the potential for delays to propagate in operations, a requisite precursor to developing tools for

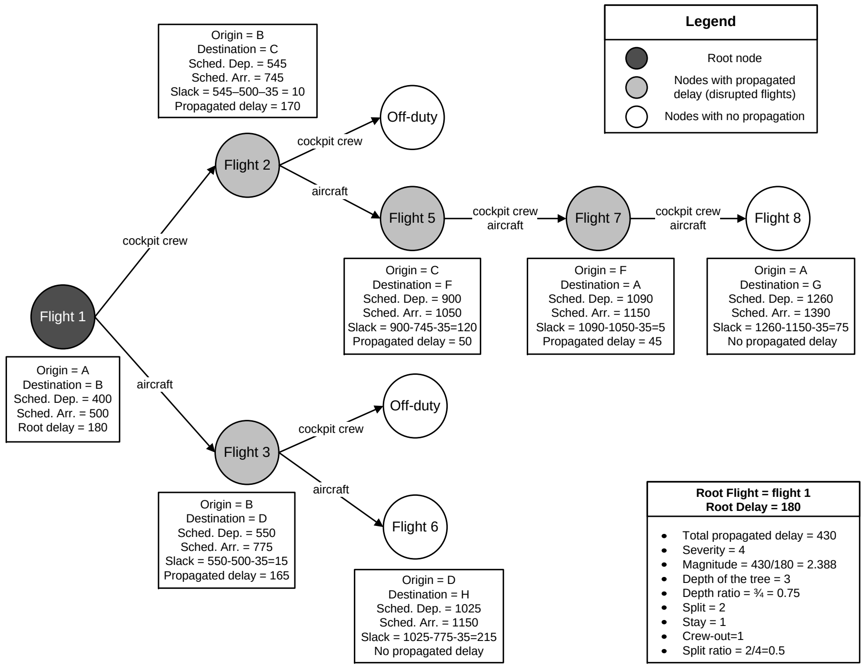


Figure 1: An example of a propagation tree.

building more robust plans.

2.3 Propagation Trees

A *propagation tree* is a simple but powerful visual and quantitative device that enables us to better understand how a root delay can, in the absence of other disruptions or schedule modifications, propagate through the network. Figure 1 provides an example of such a tree.

In this example Flight 1 is the root flight, which we suppose is delayed by 180 minutes. Note that this is an "independent" delay, caused by a mechanical problem, weather conditions, etc rather than because of an upstream flight delay. After landing, the cockpit crew of Flight 1 is scheduled to fly Flight 2 and the aircraft is scheduled to fly Flight 3. This is shown in Figure 1 by two arcs coming out of Flight 1. Assuming a minimum of 35 minutes between any two consecutive flight assignments, the cockpit crew has ten minutes of slack between Flights 1 and 2 in the original schedule. Thus, if Flight 1 is 180 minutes late in arriving, then 170 minutes of delay will propagate via the cockpit crew to Flight 2. The cockpit crew from Flights 1 and 2 goes off duty after Flight 2 and thus does not propagate further. However, the aircraft used on Flight 2 continues on to Flight 5. Because this connection initially had 50 minutes of slack and now the aircraft is 170 minutes late in arriving, 120 minutes of delay will propagate to Flight 5. Given that both the crew and the aircraft of Flight 5 connect to Flight 7, only one downstream flight can be affected. In this case, because the connection had 5 minutes of slack in the original schedule, 45 minutes of delay will propagate. Finally, the crew and the aircraft again stay together to connect from Flight 7 to Flight 8. Because there is 75 minutes of slack in the schedule for this connection, the remaining 45 minutes of delay will be absorbed and there will be no further propagation of delay from this flight.

Similarly, the aircraft assigned to Flight 1 next connects to Flight 3. Because there is 15 minutes of slack scheduled in this connection, 165 minutes of delay will propagate to Flight 3. The crew from Flight 3 next goes off-duty and therefore does not propagate any further delay. The aircraft continues on, connecting to Flight 6. This connection has sufficient slack (215 minutes), however, and thus the remaining delay from this branch of the tree is also absorbed and there is no further propagation.

By examining this propagation tree in its entirety, we observe that the original delay of 180 minutes to Flight 1 also leads to an additional 430 minutes of downstream delay, collectively impacting four other flights.

Note that we have only considered the impact of delay on two resources – the aircraft and the cockpit crews. We do this for a number of reasons. The first is data availability – these are the two resources reported in the data provided by the supporting carriers in our preliminary study. The second is importance – these are the two most costly resources, with cabin crews and passenger delays having less significant impact. The third is complexity – we felt that initially limiting our study to only two resources would enable us to develop an understanding of some of the critical drivers of delay propagation without being overwhelmed by the network interactions. We hope that the results of this first study will facilitate a second study that incorporates additional network resources (particularly, cabin crews and connecting passengers) as well.

We construct a distinct propagation tree for each scheduled flight and for a range of initial lengths of delay, to help develop an understanding of the relationship between the schedule and the potential for delays to propagate. But how can we process this collection of trees and their corresponding data? What information is of relevance?

We have developed several metrics that we suggest are of value in understanding how delays propagate:

- **Total propagated delay**:The sum of the delays (in minutes) imposed on downstream flights by an initial root delay in a propagation tree; note that the root delay is not included in this total. In Figure 1, the total propagated delay is 430 minutes.

- **Magnitude**: The ratio of total propagated delay to root delay. In Figure 1, the magnitude is 2.388. [This measure has also been referred to as *delay multiplier* (Beatty et al [3]).]

- **Severity**: The total number of disrupted flights, excluding the root flight itself. In Figure 1, the severity is 4.

- **Depth**: The number of nodes in the longest path in a propagation tree (not counting the root delay). In Figure 1, the depth is 3.

- **Depth ratio**: The ratio of depth to severity in a propagation tree. In Figure 1, the depth ratio is 0.75.

- **Stay**: The total number of nodes (disrupted flights) in which both the crew and the aircraft are the same as in the preceding node. In Figure 1, the stay is 1 (Flight 7).

- **Crew-out**: The total number of nodes (disrupted flights) in which the crew is not the same as the preceding node, because the crew in the preceding node has ended their pairing. In Figure 1, the crew-out is 1 (Flight 5).

- **Split**: The total number of nodes (disrupted flights) in which either the crew or the aircraft is not the same as the preceding flight, because these resources split to serve two different subsequent flights. In Figure 1, the split is 2 (Flights 2 and 3).

- **Split ratio**: The ratio of split to severity in a propagation tree. In Figure 1, the split ratio is 0.5.

In the remainder of this paper, we limit our focus primarily to the first five metrics for the sake of brevity. We introduce the remaining four metrics as well, however, because they are valuable in further understanding the impact of keeping crews and aircraft together in the schedule.

Through the use of these metrics, we are able to quantitatively evaluate the relationship between an individual root delay and the rest of the network. This provides insights into the relationship between airline schedules and their robustness. In particular, we use these metrics to address "conventional wisdom" from the airline and academic communities about these relationships. For example, we consider the following commonly-held beliefs:

- Propagated delays create significantly more impact than the original root delays themselves.

- A single delay can "snowball" through the entire network, affecting a large number of subsequent flights.

- Scheduling aircraft and crews to stay together can help to mitigate the impact of disruption.
- Delays that occur early in the day cause greater propagation than delays later in the day.
- It is most important to prevent delay propagation early in the day (in other words, slack should be more pronounced in the early parts of the schedule).

We will revisit these claims in section 3.2.

2.4 Study Parameters

In this study, we conduct analysis on data sets provided by two very different U.S. passenger airlines. The first is a traditional hub-and-spoke carrier that provides virtually full coverage of the U.S.; their daily flight schedule for the time period that we consider contains approximately 1700 flights. The second carrier is a point-to-point, "low-fare" carrier that focuses primarily on a specific and limited set of markets, many of them targeting leisure passengers. This carrier offers approximately 400 daily flights in the schedule that we consider.

For both carriers, we look at a single one-day "snapshot" of the schedule. Specifically, the carriers provided us with their complete schedule for a given day. This snapshot contains the complete set of domestic flights (international flights were not considered) scheduled to be operated on that day. For each flight, we were given the origin and destination, scheduled departure and scheduled arrival times, scheduled *tail number* (i.e. unique identifier for a specific airplane), and scheduled cockpit crew. This information allows us to in turn identify all pairs of sequential flights sharing a common aircraft, a common crew, or both.

For *each* flight in these networks, and for several different potential root delay lengths, we construct a distinct propagation tree. Recall that a propagation tree looks at a single root delay in isolation, assuming no other concurrent delays in the network. It also assumes that all delays propagate until they are absorbed – we do not consider recovery options such as canceling flights or calling in backup (*reserve*) crews. In constructing these trees, we assume a minimum of 35 minutes between all pairs of sequential flight assignments. We also assume a constant *minimum rest period* of 9.5 hours for all crews between one day's duty and the next. In addition to computing the individual metrics (as defined in section 2.1) for each of these propagation trees, we also compile aggregate statistics, looking at flights grouped by origin and by time of day. The sub-routine we use in order to generate the propagation trees is shown in Table 1.

```
for each value of root delay
for each flight
{
   initialize the propagation tree
   initialize the list of nodes
   create the root node and add it to the list
   while there is a node in the list
   {
     calculate the slack between the flight in the current node and the succeeding flights
     if there is not enough slack
     {
       create a new node and add it to the list
       update the propagation tree statistics
     }
     delete the current node
   }
}
```

Table 1: the sub-routine for constructing propagation trees

We analyze these results and discuss their implications in section 3. Before doing so, we conclude this section by noting limitations of our study. First, we only take into account aircraft and cockpit crews. Other resources (in particular, cabin crews and connecting passengers) can cause additional delay propagation. Second, we do not take into account interactions between delays. In reality, there is often correlation between delays (in particular, due to weather conditions) and thus propagation trees will impact one another. Third, we do not consider recovery options (canceling flights, calling in reserve crews, etc.). One of the challenges of doing so is that these decisions are rarely codified by the airlines, but instead are typically made in an ad hoc manner by SOC personnel, based on their experience and intuition. Fourth, we do not weight the probability of root delays. In our aggregate data, all root delays are treated equally. Finally, our data sets are restricted to only two carriers, and one specific day in each carrier's schedule. As a result, the analysis is by no means intended to make generalizable conclusions but rather to gain preliminary insights and to develop metrics and tools for further analysis.

3 Empirical Analysis

3.1 Case Study

In our case study, we consider flight data from two carriers, one hub-and-spoke legacy carrier and one point-to-point low-fare carrier. We look at all scheduled flights for a single day. The first network has approximately 1700 flights, the other approximately 400 flights.

For each of these flights, and for each potential delay, ranging from 15 minutes to 180 minutes in increments of 15, we construct the propagation tree. [We do not consider delays longer than 180 minutes because such delays would typically lead to cancelation or other recovery methods, rather than delaying subsequent flights.] In our analysis, we examine both individual trees and the aggregation of their metrics.

3.1.1 Highest-Impact Root Delays

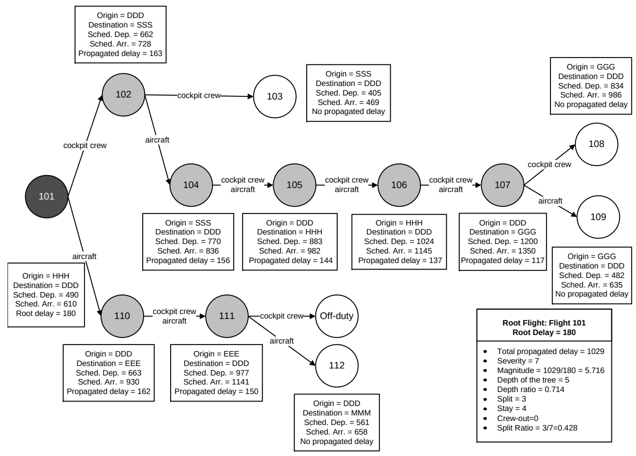
We begin our analysis by first seeking to identify the maximum impact that any one individual root delay can have on the planned schedule. Specifically, we identify the maximum severity, depth, and magnitude that can be achieved as a result of a single root delay. [Of course, these levels will be achieved under the maximum root delay length, which is 180 minutes in our study.]

Maximum Severity Recall that severity refers to the *number* of downstream flights that are impacted as a result of a root flight delay. In the hub-and-spoke network that we consider, in the worst case a single flight delay can impact seven other flights (maximum severity). This occurs only in four cases, i.e. there are only four flights for which a 180 minute root delay can lead to seven other subsequent flight delays. The corresponding propagation tree for one of these cases is illustrated in Figure 2(a). Observe that this root delay leads to 1029 total propagated delay minutes, and thus a magnitude of 5.716. The depth of this tree is five.

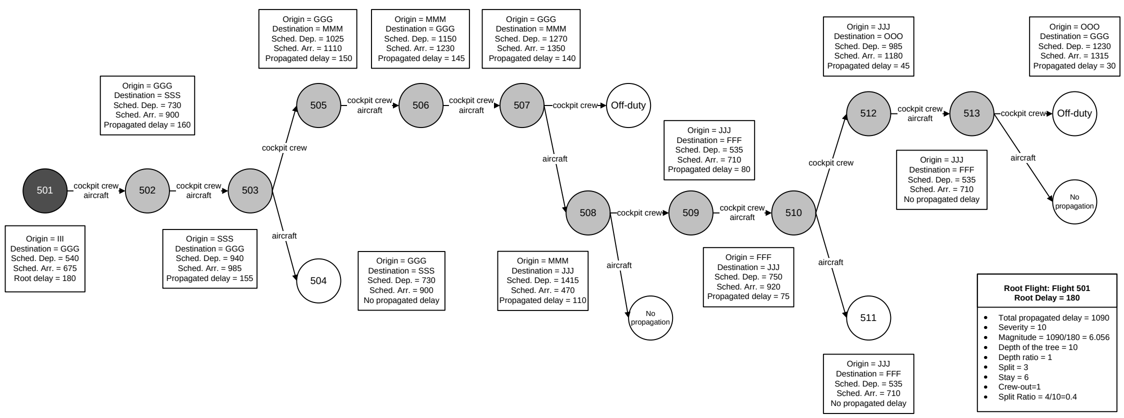
We observe similar results in the point-to-point network (Figure 2(b)). In this case, there is a single flight that, in response to a 180 minute delay, impacts 10 other flights (maximum severity); the depth is also ten. This tree has a magnitude of 6.056, associated with 1090 total propagated delay minutes.

There are two interesting observations that stem from these results. First, it seems logical that those root delays leading to the highest severity would come as a function of the most frequent splitting of resources – for example, each delayed flight in the propagation tree leading to two subsequent flight delays (one due to the aircraft and a second due to the crew), which in turn each lead to two subsequent delays, etc. However, the propagation trees that exhibit the greatest severity do not in fact show this sort of exponential growth. Instead, these trees have only one or two branches.

The second observation is that these extreme severity cases are quite rare. Although the maximum severity levels are 7 and 10, respectively, for the two networks, the vast majority of flights have severities that are significantly smaller. Table 2 provides the breakdown of severity values for the complete set of flights. Specifically, for each severity value from zero (no delay propagation) to ten (the maximum severity), this table provides the number and percentage of flights achieving that severity given a 180 minute root delay. It is interesting to note that, even with a root delay as large as 180 minutes, more than 38% (40%) of



(a) Hub-and-Spoke Carrier



(b) point-to-point Carrier

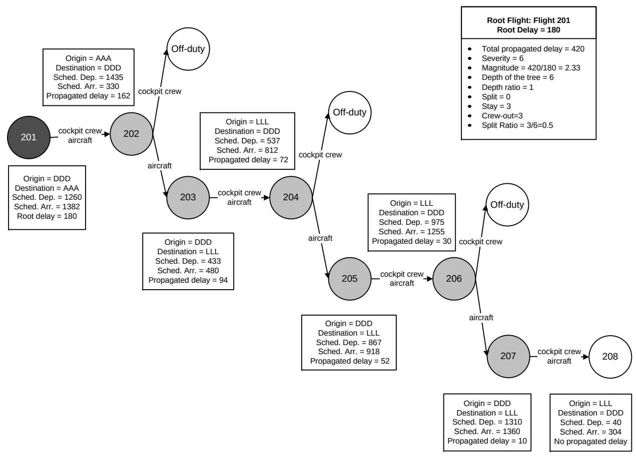
Figure 2: Propagation trees corresponding to the maximum severity.

	hub-and-spoke		point-to-point	
severity	number of flights	percent of flights	number of flights	percent of flights
10	0	0.00%	1	0.24%
9	0	0.00%	1	0.24%
8	0	0.00%	3	0.73%
7	4	0.23%	4	0.98%
6	6	0.35%	5	1.22%
5	20	1.16%	14	3.41%
4	68	3.96%	18	4.39%
3	201	11.69%	36	8.78%
2	303	17.63%	65	15.85%
1	460	26.76%	99	24.15%
0	657	38.22%	164	40.00%
sum	1719	100.00%	410	100.00%

Table 2: Severity breakdown for root delays of 180 minutes.

the root flights do not propagate at all, and almost 98% (93%) of the root flights propagate to four or fewer downstream flights.

Maximum Depth "Depth" refers to the longest path in a propagation tree. In the hub-and-spoke network, the maximum depth of any propagation tree is six; this is achieved by two flights, one of which is depicted in Figure 3. In the point-to-point network, the maximum depth is ten. This is associated with the same flight (Figure 2(b)) that leads to the propagation tree of maximum severity for that carrier. Table 3 demonstrates that these extreme cases are again quite rare, and that the typical propagation tree has much lower depth.



Hub-and-Spoke Carrier

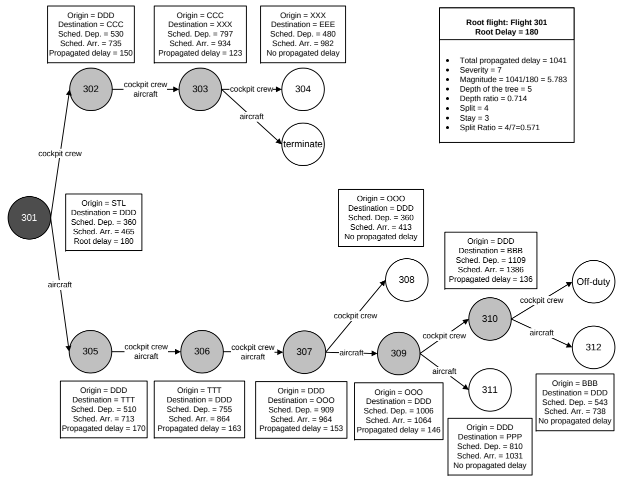
Figure 3: Propagation trees corresponding to the maximum depth.

It is interesting to note that there is not a dramatic difference between depth and severity – that is, a large number of the propagation trees are really *propagation chains*. In some cases, this is because the crew and the aircraft are scheduled to remain together. In those cases where the resources do separate, one resource may have enough to slack to absorb the disruption while the other propagates the delay. Alternatively, the crew may go off duty or the aircraft may go out of rotation and thus not propagate the delay. Finally, we note that because our data sets do not include international flights, some resources terminate prematurely in our analysis.

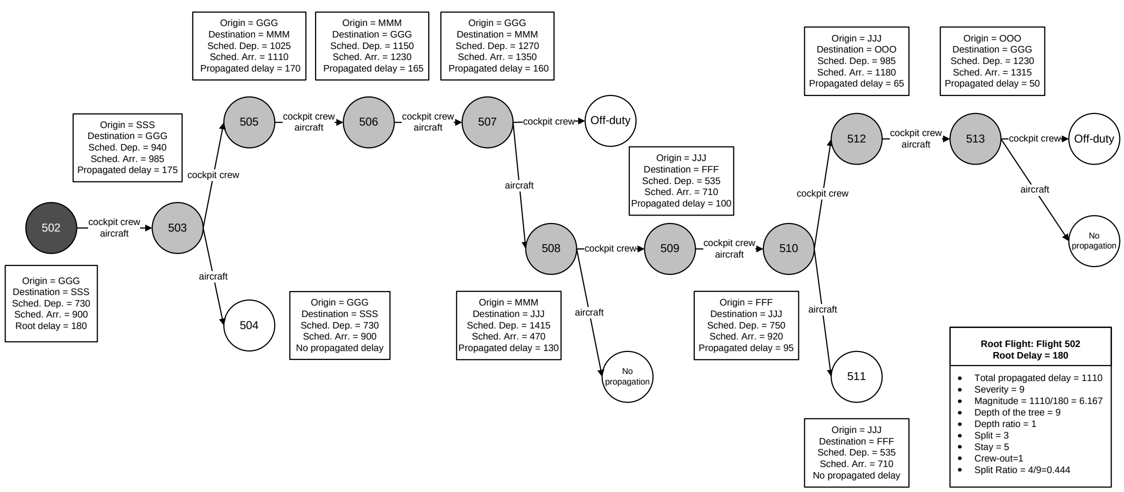
We further investigate the relationship between depth and severity in Table 4. Although it is commonly assumed that a major source of delay propagation in airline networks is the splitting of resources, we observe in this table that for many trees, such splitting does not occur. In fact, in fewer than 3% of the root delays is there a depth ratio strictly between zero (i.e. no propagation at all) and one (severity equals depth and hence no splitting).

Maximum Magnitude We next consider the maximum magnitude, i.e. the ratio of propagated delay minutes to root delay minutes. In the point-to-point network, the maximum magnitude is 6.16, which corresponds to 1110 additional minutes of propagated delay as a result of the original root delay. In this tree (Figure 4(b)), both the severity and depth are 9 (close to the maximum value for both of these metrics as well).

In the hub-and-spoke network, however, the propagation tree with highest magnitude – 5.78, corresponding to 1041 additional minutes of propagated delay – looks quite different (see Figure 4(a)). In this case, the tree does in fact split, and has a depth ratio of about 0.70. In general, one would expect higher magnitude to occur in trees with more splitting. At each new level of the tree, the propagated delay is dampened by any slack between the connecting flights, and thus propagation trees with lower depth-to-severity ratios (i.e. more splitting) will tend to have higher magnitude.



(a) Hub-and-Spoke Carrier



(b) point-to-point Carrier

Figure 4: Propagation trees corresponding to the maximum magnitude.

depth	hub-and-spoke		point-to-point	
	number of flights	percent of flights	number of flights	percent of flights
10	0	0.00%	1	0.24%
9	0	0.00%	1	0.24%
8	0	0.00%	2	0.49%
7	4	0.23%	3	0.73%
6	2	0.12%	4	0.98%
5	20	1.16%	13	3.17%
4	68	3.96%	19	4.63%
3	202	11.75%	37	9.02%
2	302	17.57%	64	15.61%
1	468	27.23%	102	24.88%
0	657	38.22%	164	40.00%
sum	1719	100.00%	410	100.00%

Table 3: Depth breakdown for root delays of 180 minutes.

depth ratio	hub-and-spoke		point-to-point	
	number of flights	percent of flights	number of flights	percent of flights
1	1033	60.09%	235	57.32%
(0,1)	29	1.69%	11	2.68%
0	657	38.22%	164	40.00%
sum	1719	100.00%	410	100.00%

Table 4: Depth ratio breakdown for root delay of 180 minute.

This idea is demonstrated more clearly in Figure 5. In this figure, we have two different flights with root delays of fifteen minutes each. Both root delays lead to a severity of two, and all flights have a slack of five minutes. However, Flight A splits (depth equals one-half severity) whereas the crew and aircraft of flight E remain together for the next two flights (depth equals severity). As a result, Flight A yields 33% more propagated delay. More generally, assuming equal severity, we would expect to see greater magnitude values for propagation trees with lower depth ratios.

We conclude this section by providing a summary of magnitude values across the complete set of flights in Table 5.

magnitude	hub-and-spoke		point-to-point	
	number of flights	percent of flights	number of flights	percent of flights
(6,7]	0	0.00%	2	0.49%
(5,6]	3	0.17%	3	0.73%
(4,5]	12	0.70%	9	2.20%
(3,4]	62	3.61%	14	3.41%
(2,3]	198	11.52%	42	10.24%
(1,2]	316	18.38%	73	17.80%
(0,1]	471	27.40%	103	25.12%
0	657	38.22%	164	40.00%
sum	1719	100.00%	410	100.00%

Table 5: Magnitude breakdown for root delays of 180 minutes.

3.1.2 Range of Propagation Tree Characteristics

In the previous section, we identified the most extreme impacts of root delays, in terms of magnitude, depth, and severity. We also observed that these extreme cases were quite rare. Furthermore, they all occurred

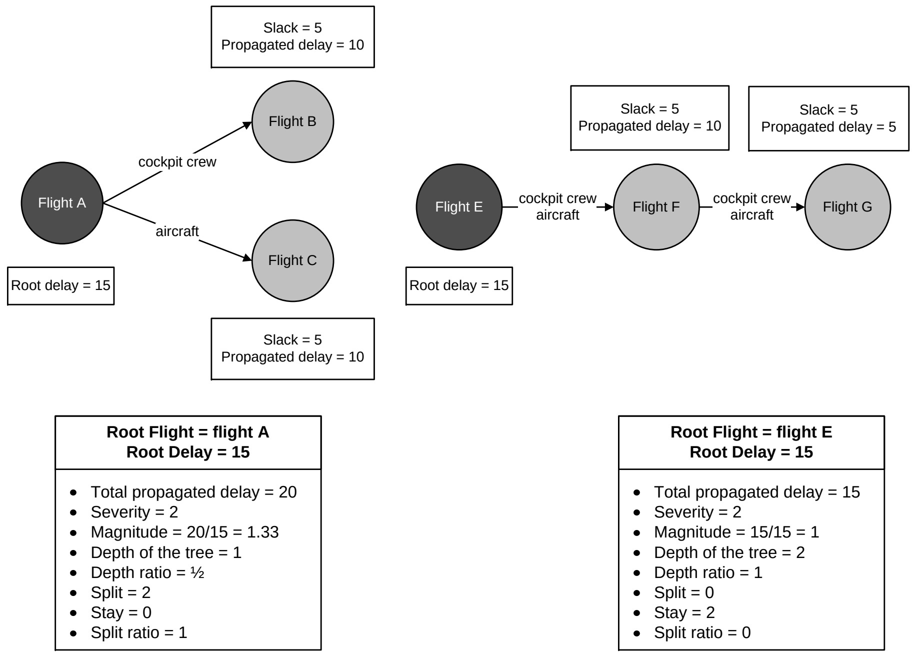


Figure 5: Assuming equal values for severity, propagation trees with depth ratio closer to one tend to have a lower magnitude measure.

as a result of the most lengthy root delay (180 minutes). In this section, we present the complete set of root delays, including not only the full set of flights but also the full range (15 minutes to 180 minutes) of potential lengths of delay.

In Table 6 and Figures 6(a) and 6(b), we present system-wide statistics on propagation severity. Table 6 lists the range of root delay lengths (from 15 minutes to 180 minutes), the maximum severity achieved by any flight for that length of root delay, and the average severity across all flights given that length of root delay.

Observe that the maximum severity level quickly jumps in the initial increases in flight delay, but then remains constant (for the hub-and-spoke carrier) or grows much more slowly (for the point-to-point carrier) as the amount of delay increases. The average severity grows a little more steadily but nonetheless also quickly reaches near-maximum levels, then shows very little growth beyond this point. This fact is even more evident in Figures 6(a) and 6(b), which show the percentage of flights achieving each given severity level for each value of the root delay. Again, we see a sharp change in the graph at a relatively low length of root delay, and then the graph remains nearly constant beyond this point. [Tables 7 and 8, and the corresponding Figure 7(a), 7(b), 8(a), and 8(b) demonstrate similar behavior for depth and magnitude as well.]

In other words, the network appears to reach a *saturation point*, and increasing the length of the root delay beyond this point does not dramatically change the corresponding severity. Why is this? Consider, for a given root flight, constructing the *exhaustive connection tree* – this is like a propagation tree, but at each flight in the tree, one or two new arcs are automatically created to represent the next flight (or flights) of the crew and aircraft (there is no notion of delay in this tree). Thus, the tree continues to grow until the resources leave the network (crews going off duty, aircraft going out of rotation), or the connection time exceeds 180 minutes and thus no delay would ever propagate.

This exhaustive connection tree is a superset of all possible propagation trees for that root flight (i.e. the tree for each root length of delay). Different propagation trees from the same exhaustive connection tree can vary in two ways. First, they can have different nodes (i.e. flights). Depending on the length of the root delay, one propagation tree may contain fewer nodes than another because the shorter delay is absorbed in

root delay	hub-and-spoke max	hub-and-spoke average (all)	hub-and-spoke average (only nonzero)	point-to-point max	point-to-point average (all)	point-to-point average (only nonzero)
15	1	0.12	1.00	1	0.19	1.00
30	4	0.73	1.41	4	0.54	1.37
45	6	0.98	1.70	5	0.78	1.51
60	6	1.08	1.82	5	0.95	1.74
75	7	1.13	1.88	5	1.06	1.91
90	7	1.15	1.91	6	1.16	2.03
105	7	1.17	1.92	7	1.22	2.11
120	7	1.18	1.93	8	1.29	2.21
135	7	1.19	1.94	8	1.32	2.26
150	7	1.21	1.96	9	1.36	2.31
165	7	1.21	1.97	10	1.40	2.34
180	7	1.22	1.98	10	1.41	2.36

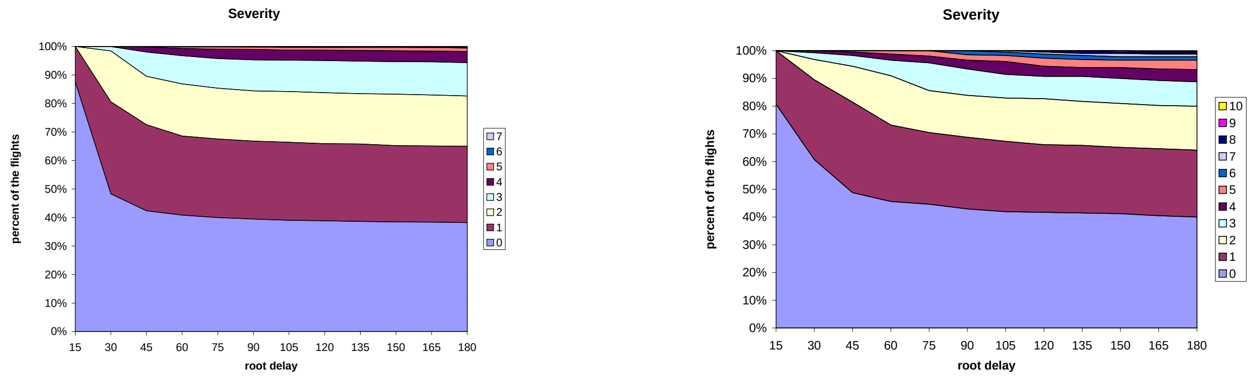
Table 6: System-wide statistics on delay propagation: severity

root delay	hub-and-spoke max	hub-and-spoke average (all)	hub-and-spoke average (only nonzero)	point-to-point max	point-to-point average (all)	point-to-point average (only nonzero)
15	1	0.12	1.00	1	0.19	1.00
30	3	0.72	1.40	4	0.54	1.37
45	4	0.97	1.68	5	0.78	1.51
60	5	1.07	1.81	5	0.95	1.74
75	5	1.11	1.86	5	1.05	1.89
90	5	1.14	1.88	6	1.15	2.01
105	5	1.15	1.89	7	1.20	2.08
120	5	1.16	1.90	8	1.26	2.16
135	5	1.17	1.91	8	1.29	2.21
150	5	1.18	1.92	9	1.32	2.25
165	6	1.19	1.93	10	1.36	2.28
180	6	1.20	1.94	10	1.37	2.28

Table 7: System-wide statistics on delay propagation: depth

root delay	hub-and-spoke max	hub-and-spoke average (all)	hub-and-spoke average (only nonzero)	point-to-point max	point-to-point average (all)	point-to-point average (only nonzero)
15	1.00	0.15	0.53	1.00	0.17	0.77
30	2.10	0.41	0.78	2.50	0.34	0.85
45	3.29	0.59	1.02	3.33	0.49	0.95
60	3.97	0.71	1.19	3.75	0.60	1.11
75	4.37	0.79	1.31	4.00	0.69	1.25
90	4.64	0.85	1.40	4.17	0.77	1.35
105	4.91	0.89	1.47	4.48	0.83	1.44
120	5.18	0.93	1.52	4.83	0.89	1.53
135	5.38	0.96	1.56	5.22	0.94	1.60
150	5.54	0.98	1.60	5.60	0.98	1.67
165	5.67	1.00	1.63	5.91	1.02	1.71
180	5.78	1.02	1.66	6.17	1.05	1.75

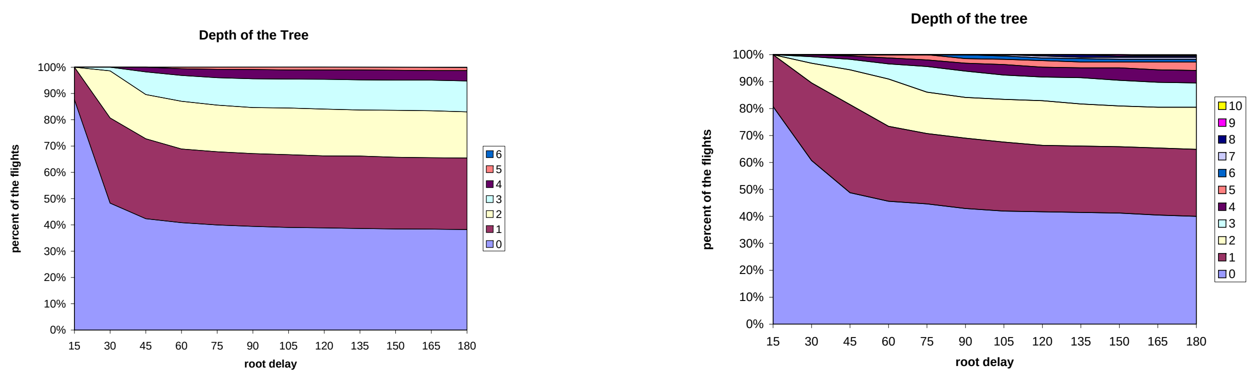
Table 8: System-wide statistics on delay propagation: magnitude



(a) Hub-and-Spoke Carrier

(b) point-to-point Carrier

Figure 6: The overall trend corresponding to the severity of the propagation trees.



(a) Hub-and-Spoke Carrier

(b) point-to-point Carrier

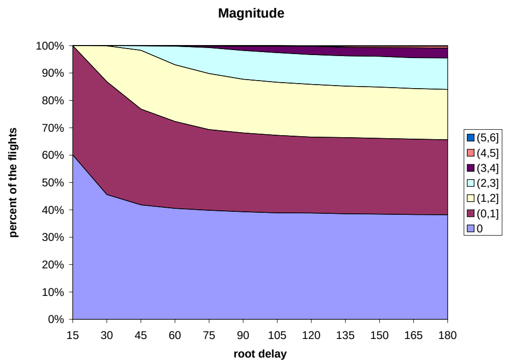
Figure 7: The overall trend corresponding to the depth of the propagation trees.

one tree but the longer delay continues to propagate in another. A second way that propagation trees for the same root flight can vary is in the lengths of the arcs – that is the actual departure times of the delayed flights. If there is 30 minutes of slack between two connecting flights and the first flight experiences a 60 minute root delay, then 30 minutes will propagate across this connection, whereas in a second propagation tree corresponding to a 90 minute root delay, 60 minutes will instead propagate.

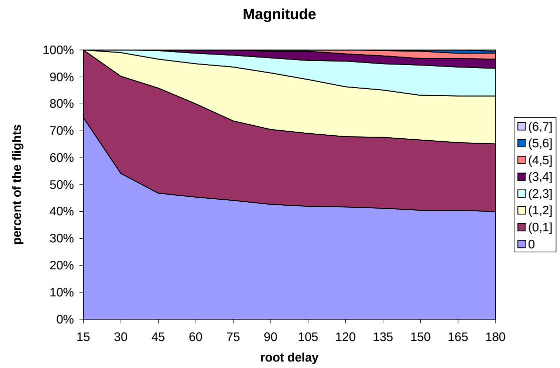
With this in mind, it is clear that at some point, as the length of root delay increases, the resulting propagation tree will eventually contain all of the nodes of the exhaustive connection tree. This in turn defines the severity and depth. As the length of root delay grows even longer, neither the severity nor the depth can increase – all flights that could be delayed are in fact being delayed. The total accumulated delay, however, does in fact continue to grow – this can be thought of as "stretching" the tree, with every flight (i.e. node) moving further forward in time. This is observed in Table 9 and Figures 9(a) and 9(b), which demonstrates that the total number of propagated delay minutes grows much more sharply across the shorter lengths of root delay, then becomes roughly linear (but certainly continues to increase) at this saturation point.

3.1.3 Range of Characteristics by Flight Category

In the previous section, we observed that many root delays do *not* propagate; those that do range significantly in the impact of their propagation. In this section, we question whether there are groups of flights that exhibit



(a) Hub-and-Spoke Carrier

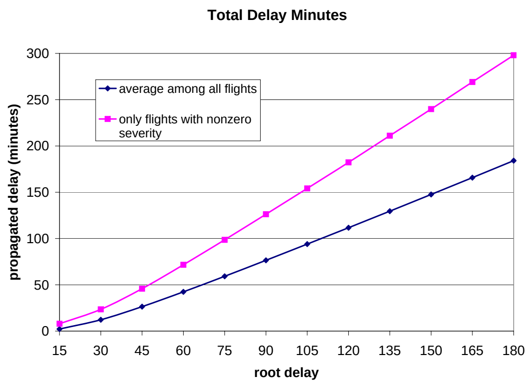


(b) point-to-point Carrier

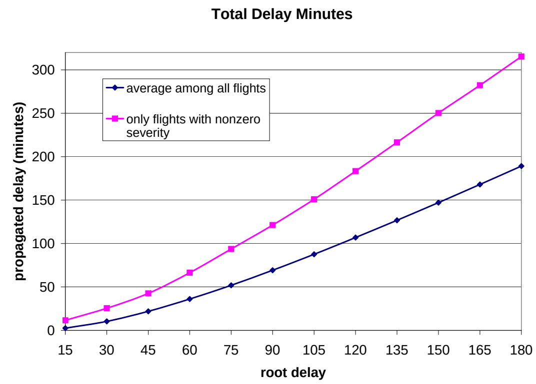
Figure 8: The overall trend corresponding to the magnitude of the propagation trees.

	hub-and-spoke			point-to-point		
root delay	max	average (all)	average (only nonzero)	max	average (all)	average (only nonzero)
15	15	2.24	7.93	15	2.52	11.58
30	63	12.25	23.51	75	10.33	25.47
45	148	26.44	45.83	150	21.91	42.60
60	238	42.43	71.65	225	36.11	66.37
75	328	59.20	98.60	300	51.91	93.71
90	418	76.45	126.24	375	69.23	121.28
105	516	93.97	154.14	470	87.63	150.96
120	621	111.68	182.26	580	106.91	183.40
135	726	129.56	211.09	705	126.74	216.50
150	831	147.61	239.83	840	147.14	250.26
165	936	165.79	269.12	975	168.02	282.32
180	1041	184.13	298.04	1110	189.21	315.35

Table 9: System-wide statistics on delay propagation: total propagated delay minutes



(a) Hub-and-Spoke Carrier



(b) point-to-point Carrier

Figure 9: Average total propagated minutes of the propagation trees.

common behaviors in their propagation patterns. More specifically, we consider two factors: departure time of the flights (time of day) and origin station.

Time of Day In this section we examine whether time of day impacts the characteristics of the propagation trees. We start by dividing the flights into three categories based on their departure time:

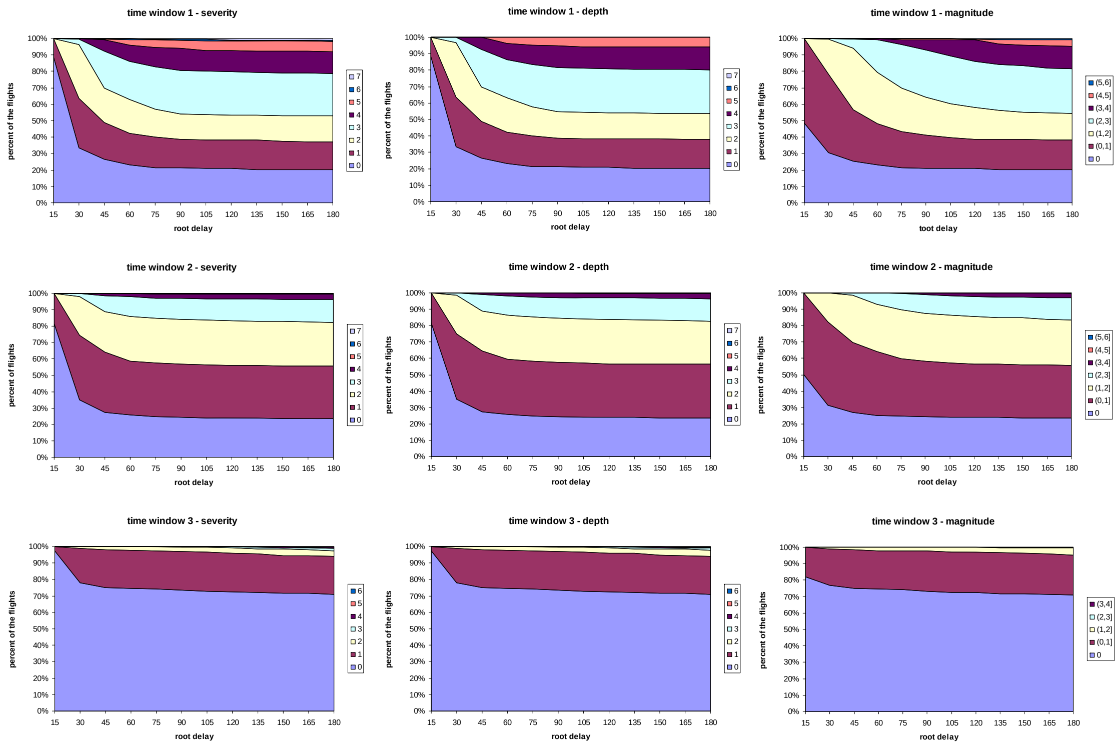
1. departures between midnight and 8 AM.
2. departures between 8 AM and 4 PM.
3. departures between 4 PM and midnight.

We continue, as in our earlier analysis, to focus on individual root delays (i.e. we are not looking at interactions between flights, such as those occurring during peak periods of congestion or times of severe weather disruption at a station). In this analysis, we are instead interested in whether the departure time of day, in and of itself, influences the structure of the propagation tree.

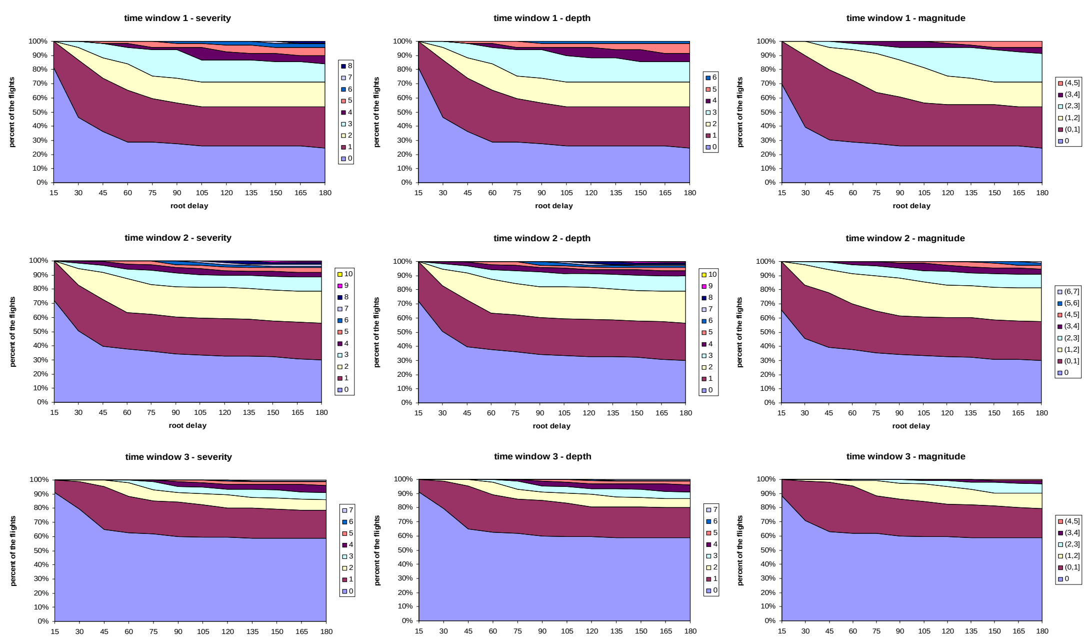
Although many carriers operate a limited number of "red-eye" flights overnight, the majority of flights typically occur between early morning (e.g. 6 AM) and late evening (e.g. 10 PM). In addition, crew members are frequently assigned to duties that start in the morning and end in the evening, with their rest periods occurring overnight, and scheduled aircraft maintenance is typically planned during the overnight period as well. Therefore, this introduces a natural "down time" in the system, in which any remaining delay propagations would typically be absorbed. As a result, it is reasonable to suspect that flights originating early in the day would have greater opportunity to propagate than flights originating later in the day. We consider this hypothesis in Figures 10(a) and 10(b).

In this figure, the top graphs show the trends in severity, depth, and magnitude of propagation trees for flights in the first time window (midnight to 8 AM); there are 266 (69) flights in this category. Similarly, the graphs in the middle row show the propagation tree characteristics for the flights in the second time window (8 AM to 4 PM); there are 908 (186) of these. Finally, the bottom row shows the characteristics of flights which depart during the third time window (8 PM to midnight); there are 545 (155) of these. Clearly, these figures demonstrate a strong relationship between the time of the original (root) delay and the degree to which that delay propagates.

Of course, there are exceptions. The most prevalent of these are propagation trees that include red-eye flights spanning the overnight lull which typically absorbs residual delay. Nonetheless, the opportunities for delay propagation decrease significantly as departure time moves later in the day. For example, in the hub-and-spoke network, if you look at the impact of a 180 minute root delay on the flights with the 100 latest departures, the average severity is 0.48 and the average number of propagated delay minutes is 35.08,



(a) Hub-and-Spoke Carrier



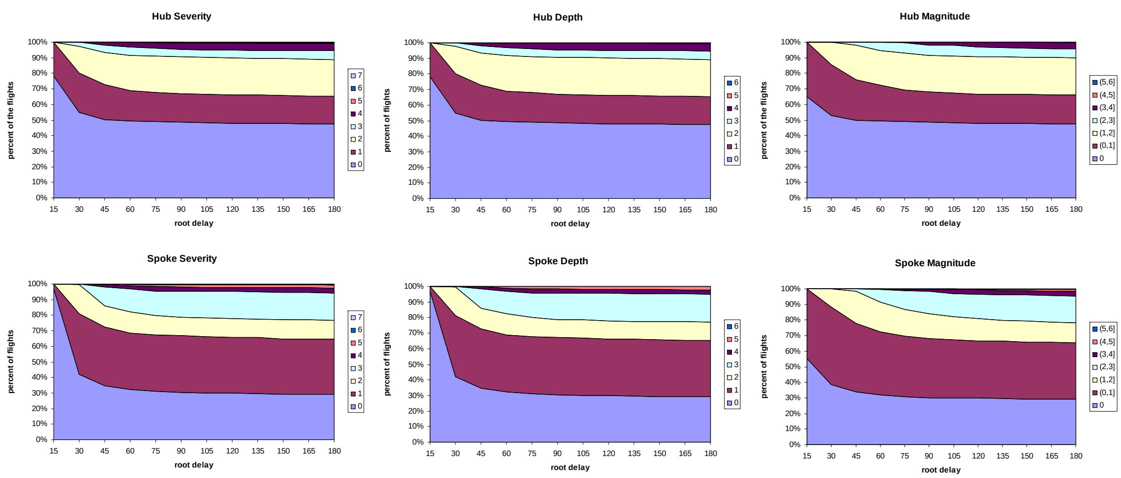
(b) point-to-point Carrier

Figure 10: Trends in the propagation measures categorized based on the departure time of the flights.

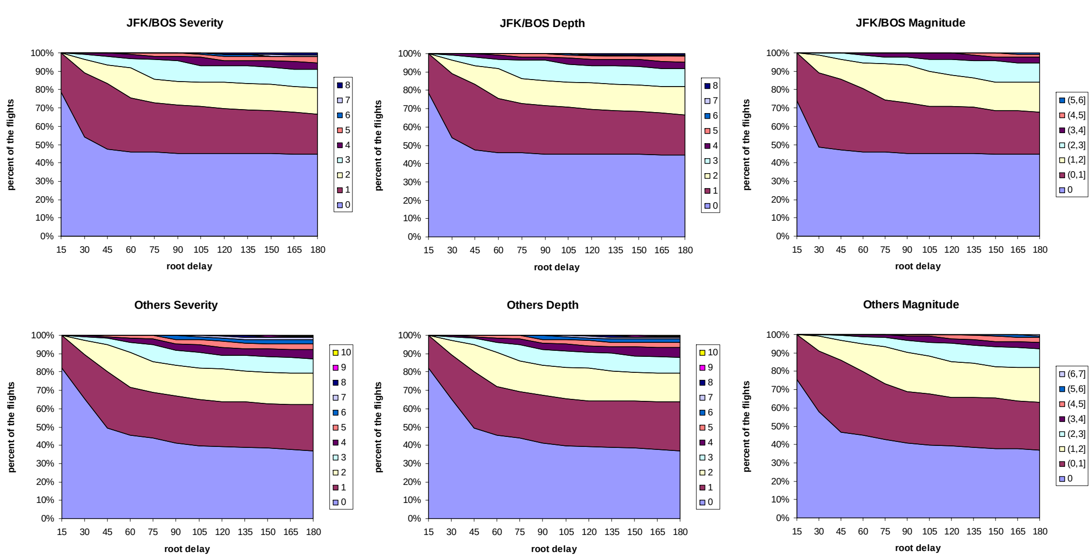
whereas for the flights with the 100 earliest departures, the average severity is 2.48 and the average number of propagated delay minutes is 368.35.

Similarly, in the point-to-point network, if you look at the impact of a 180 minute root delay on the flights with the 25 latest departures, the average severity is one and the average number of propagated delay minutes is 114, whereas for the flights with the 25 earliest departures, the average severity is 2.36 and the average number of propagated delay minutes is 304.88.

Hub vs. Spoke Another key characteristic of a flight, in addition to its departure time, is its origin airport. In particular, we question whether flights originating from hub vs spoke stations (or, in the case of the point-to-point carrier, the two highest-volume stations, which serve the majority of flights, vs the remaining stations) demonstrate different behaviors in their propagation trees. As demonstrated in Figures 11(a) and 11(b), this does in fact seem to be the case. Specifically, the *layers* of severity, depth, and magnitude are on average lower for the hub airports than for the other stations.



(a) Hub-and-Spoke Carrier



(b) point-to-point Carrier

Figure 11: Trends in the propagation measures categorized based on the departure station size.

We theorize that this occurs as a combination of four factors. First, the majority of flights originate at a hub station, terminate at a hub station, or both. Second, as observed earlier, most propagation "trees" are in fact actually chains, with very limited splitting. As a result, the propagation tree associated with one flight might be virtually identical to the tree associated with its connecting flight, except for the addition of the original flight itself. Third, crews often start and end their pairings at hubs; similarly, aircraft often go out of rotation at hubs as well. Finally, there is often more slack between flights at the hubs (both to allow for connecting passengers and also because of the larger number of connecting opportunities), whereas sparse spokes often turn aircraft (and crews) as soon as possible.

As a result, if a flight from a hub to a spoke experiences a root delay, it will often experience a propagated delay at the spoke as well; in the more general sense, we commonly see even numbers of arcs in the tree. Conversely, delays to a flight originating at a spoke typically are either absorbed at the hub or else propagate for at least two more flights (hub to spoke and then spoke back to hub).

3.2 Observations

In the previous sections, we presented empirical results concerning the potential for root flight delays to propagate downstream in an airline network. In this section, we re-visit five commonly held viewpoints about delay propagation, and discuss these viewpoints in the context of our observations.

1. **Propagated delays create significantly more impact than the original root delays themselves.**

   Because of the interconnected use of multiple constrained resources, it is commonly assumed that the propagation of a delay in a flight network has greater impact than the root delay itself. On the one hand, our observations suggest that many flights do not, in fact, propagate root delays. Even with root delays of up to 180 minutes (and taking into account the caveats of section 3.1.2), nearly 40% of the flights have no propagating effect. Furthermore, many flights that do propagate do so to a limited degree, impacting only one or two additional flights. On the other hand, this does not mean that flight delays do not propagate, nor that the impact is not of significance. In particular, we note that for flights which *do* show delay propagation, for all but the briefest of root delay lengths (e.g. of less than 30 minutes), the magnitude for roughly half the flights is more than one, i.e. the propagated delay more than doubles the initial root delay, substantiating the importance of addressing delay propagation in network planning efforts.

2. **A single delay can "snowball" through the entire network.**

   The notion of a single flight delay propagating rampantly across the network, progressively expanding in its impact, is not in fact seen in our empirical observations. This may be partially attributed to the fact that we are not considering cabin crews and connecting passengers. [On the other hand, we are also not considering recovery decisions, instead assuming all delays propagate until absorbed.] More likely, the key "buffers" that have the greatest impact on limiting the propagation of delays are:

   - crews going off-duty,
   - crews and aircraft remaining together (and thus preventing one delay from causing two subsequent downstream delays),

   and

   - periods of decreased activity in the network (particularly in late evening, but also during lulls throughout the day).

Again, this does not suggest that delay propagation is not an important concern, but that these propagations are fairly localized and should be addressed correspondingly.

3. **Keeping aircraft and crews together can help to mitigate the impact of disruption.**

   As suggested in the previous section, keeping aircraft and crews together appears to have tremendous benefit in reducing delay propagation. In particular, it implies that any one flight delay can lead to at most one subsequent delay, whereas splitting the crew and aircraft can potentially result in two downstream delays. It is interesting to note, however, that keeping crews and aircraft together is not sufficient to avoid delay propagation, and in fact, some of the flight delays with the maximum severity actually correspond to a propagation tree in which most flight connections do keep the crew and aircraft together. What is not clear from the data, and merits further study, is whether or not efforts to keep crews and aircraft together have any negative impact on delay propagation (i.e. lead to reduced amounts of slack between connecting flight pairs).

4. **Delays that occur early in the day can cause greater propagation than delays later in the day.**

   It seems logical that root flight delays early in the day will lead to greater propagation than root delays later in the day, because there are more opportunities for delay. This logic is premised on the notion, however, that there is a natural "break" at the end of the day which serves as a guaranteed buffer for any delays still propagating, and furthermore that such breaks do not occur at other times of the day. In the data that we considered this is predominantly, but not completely, true. For example, some flight delays do in fact propagate overnight, because the crews have short overnight rest periods - beginning their rest period late may force them to delay their first flight the subsequent day. In addition, there are times during the day where flight volumes decrease and thus early morning flight disruptions are absorbed into the system, rather than propagating through the day and into the evening slack period. Nonetheless, we see a significant difference in delay propagation from one flight to another when factoring by departure time of day, with all three metrics (severity, depth, and magnitude) decreasing as the origin time of the root flight increases later into the day. Similar results have been reported in the literature by Beatty et al [3], Hsiao and Hansen [5] and Tu et al [14].

5. **It is most important to prevent delay propagation early in the day.**

   The previous paragraph seems to support the conventional wisdom that it is most important to prevent delay propagations early in the day, i.e. given a fixed amount of slack in the system, that slack yields greater benefit when added to early morning flight connections. This logic, however, is premised on the notion that the disruption will actually occur. All other things being equal, a disrupted flight early in the day will in fact benefit more substantially from increased slack to absorb disruption than will a flight disrupted later in the day. On the other hand, slack early in the day also has less probability of being used.

   Consider, for example, the simple case where a crew and aircraft stay together for n consecutive flights, with zero slack between each flight pair. Suppose that we delay the first flight in the chain by five minutes. Then this delay will propagate until the end of the chain, resulting in an additional $5 * (n-1)$ extra minutes of delay. This propagated delay could be eliminated by adding five minutes of slack after the first flight.

   At the other extreme, consider a five minute delay to the second-to-last flight in the chain. This root delay would only lead to five minutes of additional propagated delay (associated with the final flight in the chain). Adding five minutes of slack to this connection would only save five minutes of propagated delay, rather than $5 * (n - 1)$, again suggesting that adding slack earlier in the day is more beneficial than later in the day.

Now consider, however, the case where we add five minutes of slack to the first connection, but the first flight is *not* delayed. Then this slack would provide no benefit, and any subsequent root delay in the chain would propagate fully. On the other hand, putting the slack between the last pair of flights would save five minutes of propagated delay for *any* root delay in the chain. In fact, all other characteristics (and, in particular, the probability of a root delay) being equal, it can be shown that the optimal location for the slack is actually in the middle of the chain. This is the trade-off point where the expected delay is minimized, trading off the length of the propagation and the probability of the root delay. Of course, actual networks are much more complex. Thus, the question of where slack can most greatly benefit the network becomes a more challenging issue, and worthy of future research.

4 Conclusions and Future Research

In this paper, we have investigated the relationship between planned aircraft and crew schedules and the potential for delay propagation as these schedules are implemented. This research makes several important contributions towards understanding this relationship, a requisite precursor to developing more robust plans. First, we introduce the use of propagation trees as a visual and quantitative tool for assessing the ramifications of individual flight delays throughout the network. Second, we propose several metrics for quantifying the impact of these initial delays as they propagate through the network. Third, we use these metrics to conduct a quantitative analysis of planned airline schedules, gaining insights into the relationship between scheduled slack and delay propagation. Fourth, we extend this analysis to substantiate (in some cases) and disprove (in other cases) commonly-held assumptions about delay propagation. Finally, we lay the groundwork for future research on incorporating measures of potential delay propagation in the airline schedule planning process. This future research includes the following four topics:

First, we see value in extending the complexity of this analysis – taking into account the probabilities of the occurrence of different root delays, recognizing correlations (for example, weather-based) between groups of root delays, adding in the propagation due to cabin crews and connecting passengers, and incorporating recovery decisions, such as crew swaps and flight cancelations.

Second, our research can be useful in helping to identify mechanisms for strategically using slack in the system to mitigate the impact of disruption, by recognizing where this slack can provide the greatest benefit.

Third, there is important work left to be done in quantifying the *value* of increased robustness – what is the trade-off between improved robustness (i.e. decreasing the likelihood of delay propagation) and scheduled costs (i.e. the cost of a plan if it operates without disruption)?

Finally, once the value of robustness can be quantified, it is possible to begin incorporating metrics of robustness within the planning process itself.

Acknowledgments:

We gratefully acknowledge the financial support of the Alfred P. Sloan Foundation Industry Studies Program and the MIT Global Airline Industry Program.

References

[1] C. Barnhart, A. Cohn, E. Johnson, D. Klabjan, G. Nemhauser, and P. Vance. *Handbook of Transportation Science*, chapter Airline Crew Scheduling. Kluwer's International Series, 2003.

[2] C. Barnhart and K. Talluri. *Design and Operation of Civil and Environmental Engineering Systems*, chapter Airline Operations Research, pages 435–469. John Wiley and Sons, New York, 1997.

[3] R. Beatty, R. Hsu, L. Berry, and J. Rome. Preliminary evaluation of flight delay propagation through an airline schedule. 1998. Proceedings of the 2^{nd} USA/Europe Air Traffice Management R&D Seminar.

[4] M. Ehrgott and D. M. Ryan. Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-Criteria Decision Analysis*, 11(3):139 – 150, 2002.

[5] C. Hsiao and M. Hansen. An econometric analysis of us airline flight delays with time-of-day effects. 2006. Proceedings of TRB 2006 Annual Meeting.

[6] http://www.transtats.bts.gov. Dot form 41 traffic and financial reports, 2006.

[7] M. Janic. Modeling the large scale disruption of an airline network. *Journal of Transportation Engineering*, 131(4):249 – 260, 2005.

[8] S. Lan, J.P. Clarke, and C. Barnhart. Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. *Transportation Science*, 40(1):15 – 28, 2006.

[9] J. Rosenberger, E. Johnson, and G. Nemhauser. A robust fleet-assignment model with hub isolation and short cycles. *Transportation Science*, 38(3):357 – 369, 2004.

[10] N. Rupp. Investigating the causes of flight delays. 2007. Working paper, Department of Economics, East Carolina University.

[11] A. Schaefer, E. Johnson, J. Kleywegt, and G. Nemhauser. Airline crew scheduling under uncertainty. *Transportation Science*, 39(3):340 – 348, 2005.

[12] S. Shebalov and D. Klabjan. Robust airline crew pairing: Move-up crews. *Transportation Science*, 40(3):300 – 312, 2006.

[13] B. Smith and E. Johnson. Robust airline fleet assignment: Imposing station purity using station decomposition. *Transportation Science*, 40(4):497 – 516, 2006.

[14] Y. Tu, M. Ball, and W. Jank. Estimating flight departure delay distributions - a statistical approach with long-term trend and short-term pattern. 2006. Working paper, Department of Decision and Information Technologies, University of Maryland.

[15] P. Wang, L. Schaefer, and L. Wojcik. Flight connections and their impacts on delay propagation. Technical report, 2003. MITRE.

[16] J. Yen and J. Birge. A stochastic programming approach to the airline crew scheduling problem. *Transportation Science*, 40(1):3 – 14, 2006.