to be submitted to *Operations Research* special issue
Operations Research for the Public Interest

Building Reliable Air-Travel Infrastructure Using Stochastic Models of Airline Networks

Mazhar Arıkan
Purdue University, West Lafayette, IN 47907, marikan@purdue.edu

Vinayak Deshpande
Purdue University, West Lafayette, IN 47907, vinayak@purdue.edu

Milind Sohoni
Indian School of Business, Hyderabad, India, milind_sohoni@isb.edu

Airlines' on-time performance was at its worst level in 2007 since 1995. A recent report by the Joint Economic Committee of the US Congress (Schumer and Maloney 2008) has estimated that the total cost to the US economy due to flight delays was as much as $41 billion in 2007. The goal of this paper is to build stochastic models of airline networks and utilize publicly available data to answer policy questions related to metrics for passenger on-time performance, bottleneck airports, robustness of airline schedules, and bottleneck flights.

The contribution of this paper is two-fold. First, we develop stochastic models, using empirical data, to analyze the propagation of delays through air-transportation networks. Our stochastic models allow us to develop three important robustness measures for airline networks. Second, our analysis enables us to make policy recommendations regarding managing bottleneck resources in the air-travel infrastructure, which if addressed, could lead to a significant improvement in air-travel reliability.

Key words: air-travel infrastructure, schedule robustness, empirical analysis, stochastic models
History: Last revised: May 31, 2010

1. Introduction

Flight delays have been a growing issue and they have reached an all-time high in recent years. According to the US Department of Transportation (DOT), a flight is considered as delayed if it arrives at the destination gate 15 minutes or more after its scheduled arrival time. The statistics show that there were 1,804,028 arrival delays out of a total of 7,455,458 commercial flight operations in the US in 2007[1]. Furthermore, in 2007, the airlines' on-time performance was at its worst level since 1995 when DOT first started to collect detailed on-time performance data[2]. Flight delays have a significant impact on the US economy. A recent report by the Joint Economic Committee of the US Congress, chaired by Senator Charles Schumer (Schumer and Maloney 2008), has estimated

[1] http://www.transtats.bts.gov/HomeDrillChart.asp.

[2] Airlines have reported on-time performance to DOT since 1987. Reporting was modified in 1995 to include reporting of mechanical delays, which had not been included in the original rule. Monthly reports are released in the Air Travel Consumer Report.

that the total cost to the US economy due to flight delays was as much as $41 billion in 2007. This includes an estimated $19 billion in operating costs to the airlines, as well as $12 billion in passenger delay costs. The report also estimated that flight delays resulted in consumption of 740 million additional gallons of jet fuel, costing an additional $1.6 billion in fuel costs, and releasing an additional 7.1 million metric tons of climate-disrupting carbon dioxide in the atmosphere (an equivalent of driving 1.57 million Toyota Prius hybrid cars for one year). The goal of this paper is to propose solutions for improving the reliability of the air-travel infrastructure through Operations Research based models.

Flight delays are typically attributed to two factors: (*i*) the randomness in the *intrinsic* travel time for a scheduled flight (which is the travel time excluding propagated delays), and (*ii*) the propagation of this randomness through the air-travel network and infrastructure. Our goal is to model both of these factors that cause travel delays. The randomness in travel time is affected by independent and collaborative decisions made by multiple agencies within the air-transportation network such as airlines, airports, and regulatory agencies like the Federal Aviation Administration (FAA). For example, airport capacity, which is driven by several factors such as number of runways, weather disruptions, etc., has a significant impact on the total travel time. Airline scheduling and resource utilization decisions impact congestion at airports, which in turn impacts the total travel time. Since most airline flight networks are tightly coupled, and resources such as airports are shared across networks, any delay caused within the system (either due to a single flight delay or a ground-stop at an airport) propagates across the entire transportation network and impacts the performance of the air-transportation infrastructure as a whole.

The goal of our research is to answer the following policy questions by building stochastic models and utilizing publicly available data. Which network based passenger-centric metrics could be used by the FAA to measure on-time performance and schedule robustness? Which are the bottleneck airports in the US air-travel infrastructure (i.e., airports that cause most delay propagation)? How would increasing airport capacity at these airports alleviate delay propagation? Which airlines have the least robust schedules? How could these schedules be made more robust? Which flight in an aircraft rotation is a bottleneck flight (and, hence, deserves managerial attention)?

To answer these questions, we collected publicly available data from several data sources including the FAA, the Bureau of Transportation Statistics (BTS), and aircraft manufacturer web sites. Our primary data source is the on-time performance data set published by the BTS. We collected data on approximately 21 million domestic flights flown by major airlines over a three year period (2005-2007). We first build a stochastic model of intrinsic (i.e., non-propagated) travel time by analyzing factors that affect this travel-time such as route, airport, airline, congestion, etc. This intrinsic travel-time distribution is then combined with a stochastic model of delay propagation to obtain the network impact of a delayed flight. These stochastic models are then utilized to define several metrics of airline network robustness.

Our analysis reveals that most airlines have negative block-time buffers built in their schedule, potentially driven by the 15 minute buffer used by DOT in reporting on-time performance. We recommend that the on-time performance reporting be modified so that on-time really means "on-time". We also construct a way of measuring "passenger" on-time arrival probability, as compared to flight on-time arrival performance reported by the BTS, and recommend that a metric that measures schedule reliability from a passenger's perspective be reported. We rank different airlines based on the magnitude of delay propagation in their networks, and recommend that airlines should judiciously allocate schedule block-time and ground-time buffers by analyzing the stochastic propagation of delays. We also construct a network metric to identify "bottleneck" flights in airline networks. Airlines should focus their managerial attention on such bottleneck flights to improve the robustness of their schedules. Our airport analysis reveals that the top five airports with the most network impact are Chicago O'Hare, Atlanta, Dallas Fort Worth, Phoenix, and Newark. Our stochastic model was then applied to study the impact of alleviating congestion at these airports. Our analysis reveals that a 15% increase in capacity at these airports will result in an average reduction in network delays by approximately 10%.

Our paper is organized as follows. In §2, we briefly review the literature on airline flight delays and schedule robustness. §3 explains the data used in the study and the stochastic models developed. Different measures for robustness and service reliabilities based on our stochastic models are introduced in §4. In §5, we develop policy recommendations to improve the reliability of the air-transportation infrastructure. §6 concludes the paper with a brief summary of findings and implications.

2. Literature Review

This paper draws on four streams of research on airline flight delays and schedule robustness: (*i*) research identifying factors that affect block-times / block-time estimation, (*ii*) flight delay propagation models, (*iii*) performance metrics which consider passenger delays, and (*iv*) performance metrics which consider flight schedule robustness.

Factors that have an effect on block-times have been a subject of much discussion recently in both academia and popular press. In a Wall Street Journal article, McCartney (2007, August 13) explains airlines' decision of using smaller airplanes as one of the reasons of increased congestion and delays at airports. The same author reports that scheduled block-times are higher compared to 10 years ago although planes can fly faster and navigate better with more advanced technology (McCartney 2007, May 29), (McCartney 2010, February 4). This block-time inflation is explained by increased variability in the air-travel infrastructure, airlines efforts to improve their on-time rankings, and slower flying speeds of smaller regional jets preferred by airlines. Academic research regarding block-time forecasting mostly focuses on individual segments of block-times. Shumsky (1995) uses a linear regression model to predict taxi-out times, while Mueller and Chatterji (2002) compare different distributions for departure, arrival, and en route delays. Hebert and Dietz (1997), Willemain et al. (2003), and Tu et al. (2008) are among others who study the estimation of different segments of block-times. In contrast, we construct a probabilistic model of total block-time instead of breaking it up into segments such as departure delay, taxi-times, air-time, etc., and propose several explanatory variables to predict the total block-time. Unlike most of the studies mentioned above, which focus on specific airports and shorter time periods, our model uses data on all the domestic flights by all major airlines flown in the US during 2005 through 2007 consisting of more than 21 million observations.

This study contributes to the robust airline scheduling literature (for a review, see Ball et al. (2007) and Barnhart and Cohn (2004)) by developing a stochastic model of delay propagation by using publicly available industry-wide data. Researchers have conducted several studies regarding modeling flight delay propagation. AhmadBeygi et al. (2008a) and Lapp et al. (2008) analyze how a flight delay propagates through the network using a tree structure (i.e., propagation tree). AhmadBeygi et al. (2008b) propose a method that reduces delay propagation by redistributing existing slack in the flight schedule. Lan et al. (2006) and Lan (2003) formulate a delay propagation problem as a mixed integer programming problem. The computational results based on their model reveal a considerable reduction in total propagated delays and in the number of disrupted passengers. The distinction between our study and these studies is in the modeling of block-times of downstream flight legs. We use a stochastic model of block-times of all flight legs, whereas these papers assume deterministic block-times given a random delay for the "root" flight. It is easy to show that the deterministic approach used in the prior literature overestimates the amount of delay propagated in the airline network, and, hence, our stochastic model is a better suited for capturing network robustness.

There exist a limited number of studies considering passenger delays (i.e., "passenger-centric" measures). On-time performance of "flights" has been the main focus of researchers in flight schedule reliability literature. Bratu and Barnhart (2005), by using their *Passenger Delay Calculator*, find that flight leg delays significantly underestimate passenger delays and that connecting passengers are more likely to be disrupted than local passengers. Their analysis is based on booking and schedule information of a major US airline during a one month (August 2000) period. Using an algorithm, they re-accommodate disrupted passengers (either due to cancellations or delays) to available flights by keeping track of total delay times of all disrupted passengers. Our analysis is different because we use data over a longer time horizon (3 years) and across the entire network of all airlines and all airports, and besides we use a stochastic model to compute delay propagation. By employing the concept of a *disrupted passenger*, which is introduced in Bratu and Barnhart (2005), Tien et al. (2009) model a passenger-based metric called *Average Passenger Delay*. This metric is calculated by using five inputs (i.e., fraction of passenger itineraries that are direct flights, fraction of canceled flights, average flight delay, average delay of disrupted passengers, and probability of missing a connecting flight). Using historical DOT data, Tien et al. (2009) find that average per passenger delay is equal to 35.95 minutes. Their findings are similar to those of Bratu and Barnhart (2005): the impact of flight delays on passengers are much severe than actual flight delays. In similar studies, Wang (2007) and Sherry et al. (2007) develop models that are designed to estimate flight-by-flight passenger trip delay using publicly available data from DOT and propose new passenger delay metrics. They suggest a route level passenger delay metric called *expected value of passenger trip delay* which is calculated by using airline on-time performance and T-100 databases for 1030 routes between OEP-35 airports. One of the drawbacks of Tien et al. (2009) and Sherry et al. (2007) is that their passenger delay calculation is on the aggregate level (i.e., industry-level and route-level respectively) and hence does not allow to compare different airlines or airports at an individual flight level. However, our study proposes a passenger delay metric at the individual flight level which enables policy makers to analyze different parts of the airline network. For instance, our model could potentially be used to assist FAA's Future Airport Capacity Task (FACT) team by pinpointing airports that create the highest network impact and hence most need the additional capacity (see, The MITRE Corporation (2007)). The contribution of our study in this area of research is to propose a novel method for finding average passenger on-time probability for each airline, by considering the probability that a passenger on a multi-leg itinerary would make a connection for each individual flight.

Studies that analyze flight schedule robustness metrics mainly develop new performance measures and compare them to the existing performance measures. Caulkins et al. (1993) argue that the DOT on-time performance metric is deficient in the sense that it does not take the differences across airports into account. They develop new metrics that consider airport difficulties (i.e., any airline schedule performance at an airport is normalized by on-time performances of all airlines that use that airport). Bratu and Barnhart (2005) propose two alternative flight-based delay metrics that can be more useful in reflecting passenger disruptions: the percentage of flights that are delayed more than 45 minutes and the percentage of canceled flights. In another study, Wu (2006) generates a schedule reliability index for flights which reflects both inherent and propagated delays. Using a schedule information of an anonymous airline, it is shown that the proposed sequential optimization algorithm, which re-allocates ground times between rotation legs through a simulation model, improves overall schedule reliability. Carey (1999) also develops new measures of reliability from analytical models which can be applicable to various forms of scheduled transportation. Our approach in this area of research is to distinguish the "intrinsic" delay attributable to the flight itself and the "propagated" delay from previous flights, which is then used to define different measures of schedule robustness.

3. Data and Model

The data for our analysis comes from several sources, however, the primary data sources are the data published by the Bureau of Transportation Statistics (BTS), the FAA, and aircraft manufacturers. In particular, we use the airline on-time performance and T-100 Domestic Market data[3] published by the BTS, Aircraft Registry and the Aviation System Performance Metrics (ASPM)[4] databases published by the FAA, and aircraft specific information available from aircraft manufacturer web sites. Our data set has approximately 21 million observations over a three year period from 2005-2007 (one observation for each commercial scheduled flight flown by a major carrier in the US).

To begin, we define the *block-time* of a flight. For every flight in our data set, we compute the *scheduled* block-time[5] as the difference between the scheduled arrival time of the flight and its scheduled departure time. The *actual* block-time is computed as the difference between the actual arrival time of the flight and its scheduled departure time. The block-time comprises of several components including taxi-out time, en route time, and taxi-in time. Each of these components is subject to different causes of delay and the total block-time delay is the sum of all individual component delays. Note that our definition of actual block-time is different from the traditional definition which measures the total elapsed time between the time an aircraft pushes back from its departure gate and arrives at its destination gate. This is important because the traditional

[3] Includes load factor information for origin-destination pairs.

[4] Measures airport capacities.

[5] We interchangeably refer to scheduled/actual block-time as a flight's quantity/demand throughout this paper.

definition of block-time ignores the impact of propagated delays and departure delays, while our definition does not.

An airline network is coupled by the resources that are shared across the network. In particular, any given aircraft for an airline typically flies multiple flights over the course of a day, and hence a delay on one flight can potentially spill-over, or propagate, from one flight to the next. Each flight is operated by a particular aircraft, which is identified uniquely by its *tail number*, and an *aircraft rotation* is a sequence of flights operated by a particular tail number. In our model, an aircraft's rotation begins with the first revenue flight after a major maintenance, or a lay-over of several hours at an airport, and ends with the last flight operated before the aircraft returns for its next maintenance or remains on the ground for several hours. Throughout our analysis, we assume that airline crews remain with the aircraft and delays caused due to unavailability or disruption of crew work schedules are insignificant. This assumption is partly justified due to the availability of reserve crews and crew schedules that follow the aircraft rotations for many airlines. We acknowledge that unavailability of crew schedule information publicly remains a limitation of our analysis. Similar to the block-time, the time duration between the next flight's scheduled departure time, on an aircraft rotation, and the earlier flight's scheduled arrival time is referred to as the scheduled *turn-around*[6] time.

Any flight delay is composed of the following two components, (i) that which is attributable (intrinsic) to the flight itself, i.e., *root delay* and (ii) that which has "spilled over" from a previous flight in the aircraft's rotation, i.e., *spill-over delay.* As mentioned earlier, spill-over delays, are primarily due to the coupling of resources. A conceptual representation of the coupling of flights on an aircraft rotation is presented in Figure 1.

Let i index over the set of flights, i.e., each record in our database. Now, for each flight i, we define the following:

Q_i : Scheduled block-time of the flight,
(= scheduled arrival time of flight i - scheduled departure time of flight i),
G_i : Scheduled ground/turn-around time of the flight
(= Scheduled departure time of flight i - scheduled arrival time of the previous flight $i-1$),
T_i : Expected time to turn the aircraft around from previous flight for flight i,
B_i : Buffer time available on ground for flight i
$(= G_i - T_i)$,

[6] We use the terms turn-around time and ground-time interchangeably throughout this paper.

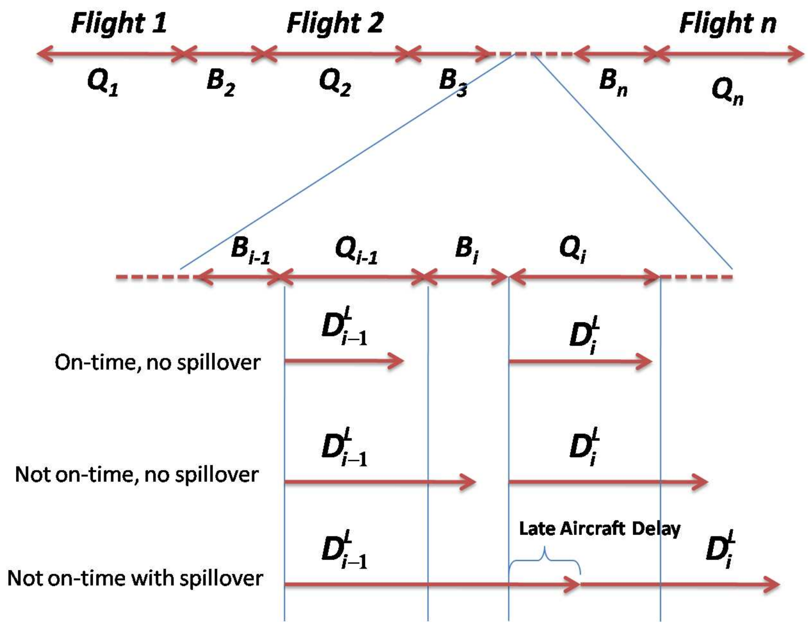

Figure 1 Illustration of propagated delay

D_i^L : Actual block-time of the flight, including the late aircraft delay or propagated delay L_i (= actual arrival time minus scheduled departure time of flight i),

L_i : Late aircraft delay/spill-over caused by the previous flight on the aircraft's rotation $(= [D_{i-1}^L - (Q_{i-1} + B_i)]^+)$,

$D_i = D_i^L - L_i$: Time to complete the flight excluding the late aircraft delay or propagated delay L_i (= truncated or intrinsic block-time).

We refer to the actual block-time of a flight, ignoring any spill-over delays, as the *intrinsic block-time*[7] of the flight. In order to compute G_i, from the on-time performance data set, we first sorted the data by airline, tail number and scheduled departure time so that all aircraft rotations are grouped together. Then, for each flight i, G_i is found by subtracting the scheduled arrival time of flight $i-1$ from the scheduled departure time of flight i. A snapshot of one such aircraft rotation flown on September 24, 2007 by Delta Airline's aircraft with tail number N980DL is shown in Table 3. Expected time to turn an aircraft was estimated by analyzing ground times at different airports for different types of aircraft for each airline. Note that expected times to turn an aircraft, T_i, show considerable differences across airports and across aircraft models. Average turn-around times of some airports and some aircraft models can be seen in Tables 1 and 2 respectively.

Finally, the B_i values were calculated by subtracting T_i from G_i for all flights except the first flight on the rotation. B_i value of the first flight of any rotation is assumed to be zero. Keeping

[7] We sometimes refer to the intrinsic block-time as the truncated block-time.

these definitions in mind, it is easy to see that there are three cases to consider between any two consecutive flights in an aircraft rotation: (*i*) previous flight is on-time so there is no spill-over to current flight, (*ii*) previous flight is not on-time but there is no spill-over to current flight since the spill-over can be absorbed by B_i, and (*iii*) previous flight is not on-time and B_i is not long enough to absorb the spill-over to current flight which creates the late aircraft delay L_i. The graphical representation of these three cases is shown in Figure 1.

Before proceeding forward, we cleaned up bad data and eliminated some flights with the following conditions: (*i*) scheduled departure time of an aircraft from an airport is earlier than its scheduled arrival of the previous flight to the same airport, (*ii*) same aircraft flying on the same route successively on the same day (duplicate records , e.g., Flight 1: IND-PHL, Flight 2: IND-PHL, and Flight 3: PHL-JFK), and (*iii*) an aircraft arrives to an airport and the next flight of the same aircraft is from a different airport in less than 5 hours. By excluding these bad observations (probably due to data entry errors), the total number of valid flights between 2005 and 2007 dropped down to 20,684,721 from 21,735,733 (See Tables 5, 6, and 7 in Appendix B for examples of bad data). In addition, we assumed that if an aircraft stays idle at an airport for more than 5 hours, the next flight of the same aircraft is the first flight of its next rotation.

Airport	**Avg. Turnaround Time**
New York JFK	42.36 min
Miami	38.17 min
Chicago O'Hare	36.44 min
Baltimore	31.73 min
Indianapolis	31.14 min
Chicago Midway	26.9 min
Houston William P. Hobby	24.02 min
Dallas Love Field	19.3 min

Table 1 Average turn-around times of some airports across all aircraft types

Model	**Num. of Seats**	**Avg. Turnaround Time**
Boeing 737-2	104-118	26.68 min
Boeing 757-2	200-234	41.05 min
Boeing 767-3	218-350	52.39 min
Airbus A-319	124-156	33.87 min
Airbus A-320	148-180	35.47 min
McDonnell Douglas DC-9	90-125	26.47 min

Table 2 Average turn-around times of some aircraft models across all airports

ORIGIN	**DEST**	**CRS DEP-TM**	**ACTUAL DEP-TM**	**CRS ARR-TM**	**ACTUAL ARR-TM**
BHM	ATL	6:00 AM	5:57 AM	7:50 AM	7:54 AM
ATL	IND	8:45 AM	8:40 AM	10:18 AM	10:14 AM
IND	ATL	10:58 AM	10:50 AM	12:33 PM	12:15 PM
ATL	DCA	1:20 PM	1:17 PM	2:59 PM	2:54 PM
DCA	ATL	4:00 PM	3:57 PM	5:53 PM	5:35 PM
ATL	IND	6:55 PM	6:52 PM	8:32 PM	8:22 PM
BLOCK-TM (Q_i)	**ACTUAL BLOCK-TM (D_i^L)**	**TURN-AROUND-TM (T_i)**	**BUFFER-TM (B_i)**	**GROUND-TM (G_i)**	**POSITION**
50 min	54 min	30 min	0 min	-	1
93 min	89 min	38 min	17 min	55 min	2
95 min	77 min	31 min	9 min	40 min	3
99 min	94 min	38 min	9 min	47 min	4
113 min	95 min	30 min	31 min	61 min	5
97 min	87 min	38 min	24 min	62 min	6

Table 3 A snapshot of aircraft rotation data

Next, we focus on describing two stochastic models that capture the coupling across an airline's network: (*i*) a stochastic model of the *intrinsic* block-time for a scheduled flight which captures factors such as airport capacity and schedule congestion, and (*ii*) a stochastic model of the actual block-time which includes delay propagation through the airline network. First, we build on the model developed in Deshpande and Arıkan (2009), where the authors show that the *intrinsic* block-time of a flight follows a log-Laplace distribution. Second, we develop a new delay propagation model that allows us to capture the ripple effect of flight delays through the network. To analyze the delay propagation in the network we focus our attention on aircraft rotations. The total block-time distribution of a flight is a convolution of the *intrinsic* block-time distribution of a flight and the propagated delay from a previous flight on the aircraft rotation. This convolution is not analytically tractable. We use a novel two-moment approximation to estimate the parameters of the total block-time distribution. The total block-time distribution is then used to characterize metrics of interest, such as the on-time arrival probability and the expected arrival delay of a flight.

These aforementioned stochastic models allow us to separate any flight delay into the two components mentioned earlier, i.e., root delay and spill-over delay. To this end, we define the two respective probabilistic metrics OTPF_i^I and OTPF_i^L as follows:

$\text{OTPF}_i^I = \Pr(D_i < Q_i)$: On-time probability without including the late aircraft delay L_i (i.e., OTP intrinsic to the flight),

$\text{OTPF}_i^L = \Pr\left(D_i^L < Q_i\right)$: On-time probability of flight i including the late aircraft delay L_i.

While delay propagation models have been studied in the past literature, it is important to note that most of these models are *deterministic*, in the sense that a stochastic root delay of a flight is propagated deterministically through the network. The key contribution of our modeling approach lies in constructing a stochastic model of delay propagation, which convolves a stochastic root delay with stochastic block-times of subsequent flights on an aircraft rotation to compute the propagated delay in the airline network. It is easy to show that the deterministic approach used in the prior literature overestimates the amount of delay propagated in the airline network and, hence, the stochastic approach is a better model in capturing network robustness.

3.1. A stochastic model of truncated/intrinsic block-time

In this sub-section we describe the model to estimate a flight's intrinsic block-time D_i. This model is explained in detail in Deshpande and Arıkan (2009). The essentials of this model are outlined below.

Deshpande and Arıkan (2009) identify several variables that affect the truncated block-time of a flight. These include the route flown by the flight, the air-carrier flying the flight, the origin and destination airports, congestion at the origin/destination airports at the time of flight departure/arrival, age of the aircraft, the number of passengers on a flight, and the month, day-of week, and arrival/departure time block of the flight. Deshpande and Arıkan (2009) show that the log-Laplace distribution provides a good-fit to model the truncated block-time of a flight. A unique measure of congestion is used to understand its impact on block-times. For each flight in the on-time performance data set, the arrival/departure congestion is calculated by counting the total number of flights scheduled to arrive/depart in an adjacent one-hour time block around the scheduled arrival/departure time at the destination/origin airport. This number is then normalized by dividing by that airport's arrival/departure capacity which was obtained from the ASPM database. This congestion measure is unique in the sense that it measures congestion for an *individual* flight and is not an average congestion measure which has been typically used in prior literature. Deshpande and Arıkan (2009) also provide a method to estimate the mean μ_i and variance $2b_i^2$ of the log-Laplace distribution for the truncated block-time of flight i. We represent $\widehat{\mu}_i$ as the best estimate of μ_i computed using our data sets.

The first estimator, $\widehat{\mu}_i$, can be found by regressing $\ln(D_i)$ on X_i, where X_i is the vector of variables that impact the truncated block-time distribution. The estimation model is given by

$$\begin{aligned}\ln(D_i) = {} & \beta_0 + \beta_1 * \text{Route}_i + \beta_2 * \text{Origin}_i + \beta_3 * \text{Destination}_i + \beta_4 * \text{Carrier}_i + \beta_5 * \text{Month}_i + \\ & \beta_6 * \text{Day-of-Week}_i + \beta_7 * \text{Departure-Time-Block}_i + \beta_8 * \text{Arrival-Time-Block}_i + \\ & + \beta_9 * \text{Arrival-Congestion}_i + \beta_{10} * \text{Departure-Congestion}_i + \\ & \beta_{11} * \text{Aircraft-Age}_i + \beta_{12} * \text{Average-Passengers}_i + \varepsilon_i \end{aligned} \tag{1}$$

The above regression model is run on 2005-2006 data consisting of 13,227,718 observations to obtain an estimate of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$. These estimates are then applied to 2007 data consisting of 6,736,521 flight observations to obtain $\widehat{\mu}_i$ as follows.

$$\widehat{\mu}_i = X_i^T \widehat{\boldsymbol{\beta}} \tag{2}$$

To estimate the variance $2b_i^2$ it is assumed that the variance of the block-time across all flights is heteroscedastic and the error terms can be split into three components:

$$\text{Total Error} = \text{Origin Airport Shock} + \text{Destination Airport Shock} + \text{Carrier Shock}.$$

For any flight i, in any month m in the given data, denoting the origin airport shock as $\widehat{\sigma}^2_{\text{org},m}$, the destination airport shock as $\widehat{\sigma}^2_{\text{dest},m}$, and the carrier specific shock as $\widehat{\sigma}^2_{i,c}$, the estimate of the variance of the truncated block-time, $2\widehat{b_i^2}$, is

$$2\widehat{b_i^2} = \widehat{\sigma}^2_{\text{org},m} + \widehat{\sigma}^2_{\text{dest},m} + \widehat{\sigma}^2_{i,c}. \tag{3}$$

See Deshpande and Arıkan (2009) for the methodology to obtain the estimates $\widehat{\sigma}^2_{\text{org},m}$, $\widehat{\sigma}^2_{\text{dest},m}$, and $\widehat{\sigma}^2_{i,c}$.

Next, in §3.2, using these truncated block-time distributions, we develop a stochastic model to estimate the total delay propagated on an aircraft rotation.

3.2. A stochastic model of actual block-time and delay propagation

Recollect that our definition of the actual block-time D_i^L includes delays that are propagated from previous flights. Our goal in this sub-section is to develop an approximate distribution for D_i^L. An aircraft rotation consists of a string of flights operated by the same tail number. Thus, if a flight $i-1$ on the aircraft rotation is delayed, it can end up affecting the next flight i on the aircraft rotation. This will happen if the delay of flight $i-1$ exceeds the ground buffer B_i of flight i. Thus, the spill-over/ propagated delay from flight $i-1$ to flight i is given by,

$$L_i = [D_{i-1}^L - (Q_{i-1} + B_i)]^+ \tag{4}$$

Thus, for any flight i in the rotation, we have

$$D_i^L = \overbrace{D_i}^{\text{no spill-over}} + \overbrace{[D_{i-1}^L - (Q_{i-1} + B_i)]^+}^{\text{spill-over}=L_i} \tag{5}$$

The first term in (5) captures a flight's own random block-time and the second term captures any spill-over (propagated[8]) delay incurred due to the previous flight in the aircraft rotation. It is

[8] We use the terms spill-over and propagated interchangeably throughout this paper.

noteworthy that the second term is the same as L_i. Moreover, we know that D_i follows a log-Laplace distribution.

However, the distribution of the actual block-time, D_i^L is not easy to characterize since it is the convolution of a log-Laplace distribution and the spill-over distribution. There does not exist a simple distribution that characterizes D_i^L. Hence, we make the simplifying assumption that D_i^L also follows a log-Laplace distribution with a mean $\tilde{\mu}_i$ and variance $\tilde{b}_i$. Later (see, §7.3 in Appendix C), we provide empirical evidence that such an assumption is reasonable. The challenge, however, is to estimate the mean $\tilde{\mu}_i$ and variance $\tilde{b}_i$ of the resulting (assumed) log-Laplace distribution. Next, in Proposition 1[9][10] we estimate $\tilde{\mu}_i$ and $\tilde{b}_i$ by matching the first two moments of equation (5).

PROPOSITION 1. *Matching the first two moments of the actual block-time distribution equation (5) gives:*

$$\tilde{b}_i^2 = (-1 - 2A) + \sqrt{(4A+1)(A+1)} \tag{6}$$

where $A = \frac{Var[D_i^L]}{\mathbb{E}[D_i^L]^2}$ *and*

$$\tilde{\mu}_i = \ln\left(\mathbb{E}[D_i^L](1 - \tilde{b}_i^2)\right). \tag{7}$$

There are several advantages of using the above two-moment approximation. First, the above approach provides a tractable analytical expression for the actual block-time D_i^L. Second, the above approximation can be applied recursively to a string of flights on an aircraft rotation to compute the distribution of each flight on an aircraft rotation. Thus, it easily captures the cumulative effect of prior accumulated propagated delays on a given flight i by keeping track of a single random variable D_{i-1}^L of the previous flight, instead of having to keep track of the truncated block-time distributions of all previous flights on the rotation. Third, the above computational procedure is extremely efficient. Evaluating the actual block-time distributions and the corresponding on-time arrival probability for each of the 7.4 million flights flown in 2007 took less than an hour on our desktop computer. This is very important, since suggesting policy changes requires estimating the impact of changes to the entire network consisting of millions of flights. This would be very difficult to do with other computational approaches such as simulation of the entire system. A discussion on the accuracy of the above approximation is provided in §7.3 in Appendix C.

We use the stochastic models described in this section to develop network robustness measures and service levels in the next section.

[9] See, §7.1 in Appendix A for the explicit representations of $\mathbb{E}[D_i^L]$ and $\text{Var}[D_i^L]$.

[10] $\tilde{\mu}_i$ and $\tilde{b}_i^2$ for any flight i can be computed using (6) and (7) recursively starting from the first flight on any rotation where $\tilde{\mu}_1 = \widehat{\mu}_1$ and $\tilde{b}_1^2 = \widehat{b}_1^2$.

4. Robustness and Service Reliability Measures for Airline Networks

DOT measures a flight to be on-time as long as the flight arrives at its destination no more than 15 minutes after its scheduled arrival time. The performance of airlines are typically ranked based on the aggregate (average) on-time performance of all the flights in their network. While such a measure, and ranking, provides a basis for comparison, several important issues remain unanswered. We list three important issues below:

1. The DOT aggregate measure does not distinguish between the nature of airline operations, network structures, and markets served by individual airlines.

2. The DOT metric focuses on each flight separately and does not distinguish between the root delay, attributable to the flight itself, and spill-over delay (propagated delay or late arrival delay) propagated through the network. Consequently, using the DOT aggregate metric for ranking does not allow us to clearly explain the negative externality imposed by a poorly operated flight on the rest of the network.

3. The DOT aggregate metric is not a passenger-centric metric in the sense that it does not differentiate between flights, in individual airline networks, which carry a large mix of connecting and local passengers.

In this section our goal is to develop measures that help us describe and quantify the aforementioned effects of (i) spill-over/propagated delay effects, and (ii) passenger service levels. Let $\overline{\text{OTPF}}_a^{DOT}$ represent the aggregate on-time performance computed, for any airline a's network, using the DOT definition of on-time performance, i.e., using the additional 15 minutes buffer beyond the scheduled arrival time of a flight. Recollect our definitions of OTPF_i^L and OTPF_i^I in §3 are different than those used by DOT since we do not use the 15 minute buffer beyond the scheduled arrival time. Let $\overline{\text{OTPF}}_a^L$ represent the aggregate on-time performance computed for any airline a's network excluding the 15 minute buffer, and $\overline{\text{OTPF}}_a^I$ represent the aggregate intrinsic (excluding late-aircraft delays) on-time performance computed for any airline a's network. Using the distributional parameters $\tilde{\mu}_i$ and $\tilde{b}_i$ developed in equations (7) and (6) respectively, we rewrite the expressions for the corresponding service metrics (represented by these probabilities) as

$$\text{OTPF}_i^L = \mathbb{L}\left(\frac{\ln(Q_i) - \tilde{\mu}_i - \widehat{\gamma}}{\tilde{b}_i}\right) \tag{8}$$

$$\text{OTPF}_i^{DOT} = \mathbb{L}\left(\frac{\ln(Q_i + 15) - \tilde{\mu}_i - \widehat{\gamma}}{\tilde{b}_i}\right) \tag{9}$$

$$\text{OTPF}_i^I = \mathbb{L}\left(\frac{\ln(Q_i) - \widehat{\mu}_i - \widehat{\gamma}}{\widehat{b}_i}\right). \tag{10}$$

where $\mathbb{L}$ represents the CDF of the Laplace distribution with parameters (0,1)[11] and $\widehat{\gamma}$ represents the median of all $\widehat{\varepsilon}_i$'s coming from (1). Let $\mathcal{A}$ represent the set of airlines in our data set and $\mathcal{F}_a$ denote the set of flights operated by a particular airline $a \in \mathcal{A}$. Let $\mathcal{R}_a$ denote the set of aircraft rotations belonging to airline $a \in \mathcal{A}$.

To this end, using equations (8) - (10), we can define network measures of on-time performance for each airline a as follows:

$$\overline{\text{OTPF}}_a^L = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^L}{|\mathcal{F}_a|} \tag{11}$$

$$\overline{\text{OTPF}}_a^{DOT} = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^{DOT}}{|\mathcal{F}_a|} \tag{12}$$

$$\overline{\text{OTPF}}_a^I = \frac{\sum_{i \in \mathcal{F}_a} \text{OTPF}_i^I}{|\mathcal{F}_a|} \tag{13}$$

To begin we quantify the inflation of on-time performance due to the use of the 15 minute buffer in the DOT definition of on-time performance. In Figures 2 and 3, we illustrate how airline rankings would vary if their rankings on $\overline{\text{OTPF}}_a^L$, i.e., without the use of the additional buffer of 15 minutes, were used instead of rankings using the 15 minute buffer $\overline{\text{OTPF}}_a^{DOT}$. As is readily

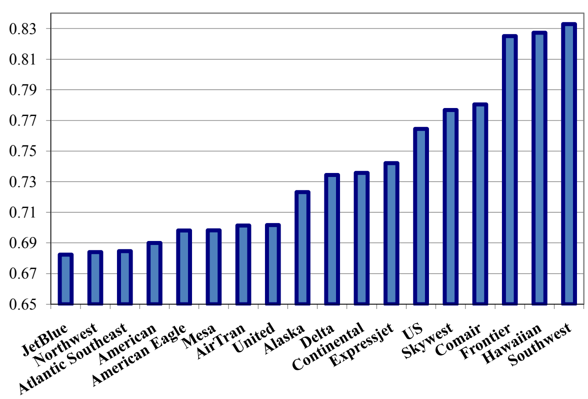

Figure 2 $\overline{\text{OTPF}}_a^{DOT}$ for different US airlines.

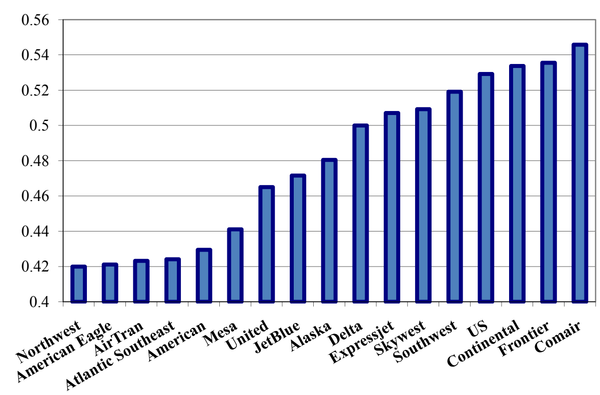

Figure 3 $\overline{\text{OTPF}}_a^L$ for different US airlines.

observable, some airlines are able to use the 15 minute buffer, available to them under the DOT measure, effectively. For example, the ranking for Southwest Airlines improves significantly under $\overline{\text{OTPF}}_a^{DOT}$. On the contrary, airlines such as JetBlue, improve their position, vis-à-vis other airlines,

[11] $L(x) = 0.5[1 + \text{sgn}(x)(1 - \exp(-|x|))]$

under $\overline{\text{OTPF}}_a^L$. Our observations suggest that providing an additional buffer does not address the issue of operational challenges faced by airlines, which are inherent to their network structures and markets served. The additional buffer provides some airlines a leeway in hiding their inefficiencies and improving their position in the DOT on-time performance rankings. Consequently, $\overline{\text{OTPF}}_a^L$ suffices as an on-time performance metric without providing some airlines undue advantage. In the remainder of the paper we compare all our measures to the $\overline{\text{OTPF}}_a^L$ measure.

We now define the following robustness measures for an airline flight network. Each of these measures quantify the delay externality imposed on a flight, or a resource using a set of flights, due to tight coupling of the flights within an airline's network.

4.1. The average drop in on-time arrivals due to spill-over/propagated delays

We begin by defining a robustness metric, $\overline{\text{OTPD}}_a$, that measures the average *network* impact (drop) on a flight's on-time performance due to spill-over/propagated delays (due to late arrivals of aircraft from its previous flight) for any airline $a \in \mathcal{A}$.

$$\overline{\text{OTPD}}_a = \overline{\text{OTPF}}_a^I - \overline{\text{OTPF}}_a^L \quad \forall\ a \in \mathcal{A}. \tag{14}$$

We illustrate the impact of spill-over/propagated delays (L_i's) in Figure 4. Earlier, in Figure 3 we plotted $\overline{\text{OTPF}}_a^L$ for each airline in our data set. In Figure 4, we plot $\overline{\text{OTPD}}_a$ to illustrate the impact of spill-over/propagated delays. Observe that the magnitude of spill-over delays, in

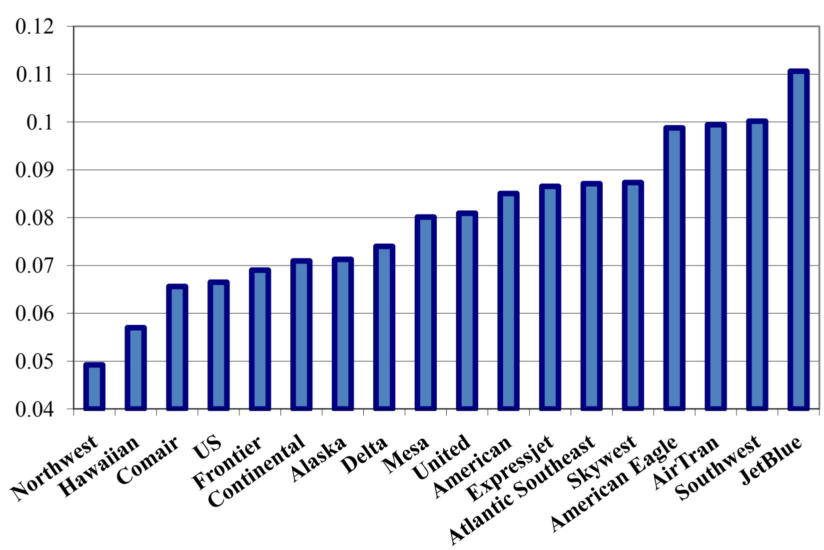

Figure 4 Average drop in on-time performance ($\overline{\text{OTPD}}_a$) due to spill over delays.

Figure 4, are substantial. Consequently, the probability of delay, intrinsic to the flight itself (i.e., root delay), is considerably lower. Thus, from an airline's perspective, the need of identifying and controlling spill-over delays, to improve operational performance, remains significant.

4.2. The fraction of expected delay attributable to spill-over/late-aircraft effects

Given that understanding the impact of spill-over delays is important, we now quantify the magnitude of the spill-over delay. For any airline $a \in \mathcal{A}$ and any flight $i \in \mathcal{F}_a$, this robustness metric, FEPD_i, measures the average fraction of total expected delay, observed on flight i, which is attributable to the spill-over effect from previous flights in the aircraft rotation. Essentially, it allows us to separate the average magnitude of the expected delay externality imposed on any flight i.

$$\text{FEPD}_i = \frac{\mathbb{E}(D_i^L - Q_i)^+ - \mathbb{E}(D_i - Q_i)^+}{\mathbb{E}(D_i^L - Q_i)^+}, \text{ where } i \in \mathcal{F}_a. \tag{15}$$

Using (15) we define the average fraction of expected propagated delay, attributable to spill-over effects, for each airline $a \in \mathcal{A}$. We denote this by $\overline{\text{FEPD}}_a$ for all $a \in \mathcal{A}$ and is computed as

$$\overline{\text{FEPD}}_a = \frac{\sum_{i \in \mathcal{F}_a} \left(\mathbb{E}(D_i^L - Q_i)^+ - \mathbb{E}(D_i - Q_i)^+ \right)}{\sum_{i \in \mathcal{F}_a} \mathbb{E}(D_i^L - Q_i)^+} \quad \forall\ a \in \mathcal{A}. \tag{16}$$

We plot $\overline{\text{FEPD}}_a$ in Figure 5 for all airlines in our data set. It is readily observable from Figure 5

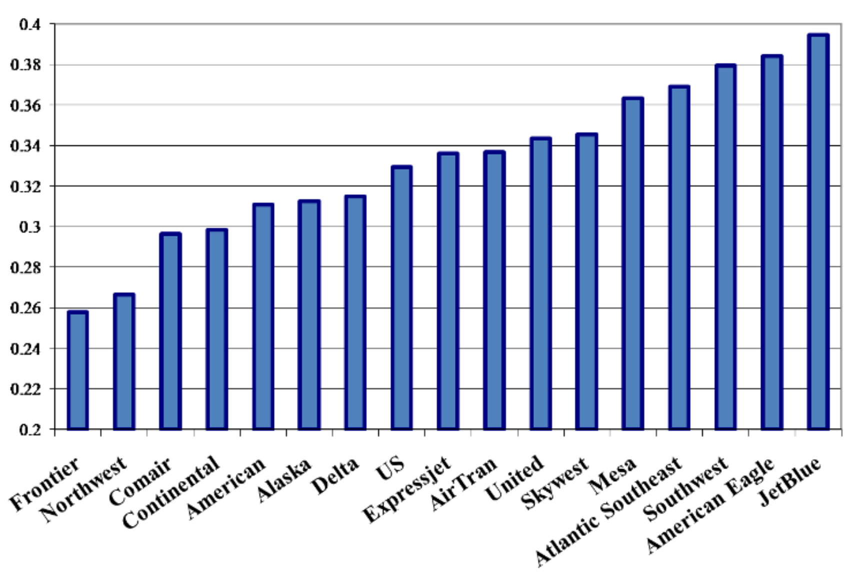

Figure 5 Fraction of expected delay due to spill over delays ($\overline{\text{FEPD}}_a$) for different US airlines.

that the fraction of average delay, attributable to spill-over effects, is substantial for most airlines. Consequently, from an airline's perspective, the need to identify and control flight operations to reduce the magnitude of the impact of spill-over delays remains significant.

4.3. Bottleneck flights

Recollect our definition of an aircraft rotation in §3. Let us denote an aircraft rotation by r. We first define the notion of a *bottleneck* flight within an aircraft rotation. Let $\mathrm{OTPF}^L_{i,j}$ denote the on-time probability of flight j, within a particular aircraft rotation, assuming that the aircraft rotation begins with flight i instead of the first flight in that rotation. Obviously, j must depart after i and with a slight abuse of notation we denote this by $j > i$. Consequently, $\mathrm{OTPF}^L_{1,j} = \mathrm{OTPF}^L_j$ and $\mathrm{OTPF}^L_{1,i} = \mathrm{OTPI}^L_i$. Now suppose $N(r)$ denotes the total number of flights in the aircraft rotation r (string of flights in rotation r), then the *network impact* of flight i, NI_i, is defined as the cumulative drop in on-time probability of subsequent flights in an aircraft rotation that is attributable to flight i, and is computed as follows:

$$\mathrm{NI}_i = \sum_{j=i+1}^{N(r)} \left(\mathrm{OTPF}^L_{i+1,j} - \mathrm{OTPF}^L_{i,j}\right) \tag{17}$$

A *bottleneck* flight, within a rotation aircraft r, is defined as the flight with the largest network impact. Let BNFL_r denote the position of the bottleneck flight within its rotation, i.e.,

$$\mathrm{BNFL}_r = \arg \max_{1 \le i < N(r); i \in r} \{\mathrm{NI}_i\}. \tag{18}$$

Using our data set, we compute the frequency with which a flight at a particular position in any flight rotation, across all rotations for all airlines, is the bottleneck flight, as defined by (18). We denote this frequency value as BNFL_p, where the index p denotes the position of a flight in the aircraft rotation. Anecdotal evidence (Ross and Swain 2007) suggests that a natural question, for airline managers, to ask is: Do these bottleneck flights correspond to earlier flights in the day? We count the bottleneck flights, across all airlines, within a particular departure time slot t and denote the corresponding frequency of occurrence as BNFL_t. Figure 6 plots the frequency of occurrence, BNFL_p, for different flight positions, p, within any aircraft rotation. As is readily observable, the first 4 flights in any rotation have a higher probability of being the bottleneck flights. In Figure 7, we plot BNFL_t for different departure times t. As we illustrate in Figure 7, our data set suggests that mid-day flights, 10:00 a.m. through 6:00 p.m., also have a high probability of being bottleneck flights. Consequently, early morning flights are not significantly more likely to be bottleneck flights. The above observations suggest that airlines must carefully identify bottleneck

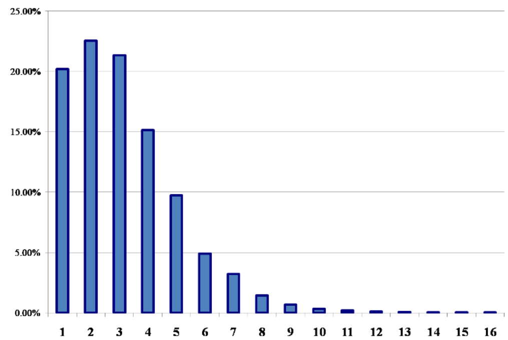

Figure 6 BNFL_p, computed across all airlines, for different flight positions.

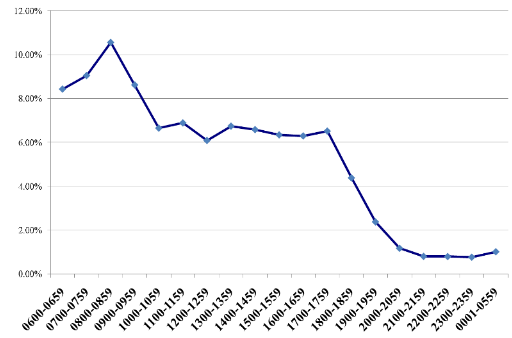

Figure 7 BNFL_t, computed across all airlines, for different departure times.

flights and monitor them to improve operational performance. Airlines can improve their network robustness by focusing their managerial attention on bottleneck flights and by allocating more resources to bottleneck flights in an aircraft rotation.

4.4. The expected passenger on-time probability

We define the *expected passenger on-time probability* as the probability that a passenger will complete a single/multi-flight itinerary by the scheduled arrival time of the last leg of the itinerary. Earlier we noted that, the DOT measure of on-time flight arrival $\overline{\text{OTPF}}_a^{DOT}$, is not a passenger-centric metric. Specifically, we mean that $\overline{\text{OTPF}}_a^{DOT}$ fails to measure the probability that passengers, with itineraries consisting of 2 or more flights, will make their connections on a particular airline's network. To quantify such a robustness measure, we first define the probability that a passenger would make a connection to a flight on the itinerary. Let i and j be two consecutive flights on a passenger itinerary such that $i,\ j \in \mathcal{F}_a$. Obviously, j must depart from the same station into which i arrives and j's scheduled departure must be after i's scheduled arrival time. With a slight abuse of notation we denote this by $j > i$. Let d_j^s denote the scheduled departure time of flight j and a_i^s denote the scheduled arrival of flight i. We define a *connection buffer* as the difference between the scheduled turn-around times between two consecutive flights on the passenger's itinerary and the minimum time required to travel between any two gates at the arrival airport of flight i, m_a. Let $\text{CB}_{i,j}$ denote the connection buffer corresponding to flight i, then,

$$\text{CB}_{i,j} = d_j^s - a_i^s - m_a. \tag{19}$$

The probability of connecting to flight j from flight i is $\text{OTPC}_{i,j}$, and is given by

$$\text{OTPC}_{i,j} = \Pr[D_i^L - Q_i < \text{CB}_{i,j}]. \tag{20}$$

Having noted that it is important to develop a measure to capture connection probability, we note that it is equally important to adjust this measure for the types of passenger itineraries served by a particular airline. Let, for any flight $i \in \mathcal{F}_a$, w_{ij} denote the average number of passengers connecting to flight j from flight i at the destination station of flight i and let w_{ii} denote the set of local passengers, i.e., who do not connect to any other flight at the destination station of flight i. For brevity, let $\mathcal{C}_i$ denote the set of flights that passengers traveling on flight i can connect to, i.e.,

$$\mathcal{C}_i = \left\{ j \in \mathcal{F}_a : d_j^s > a_i^s,\ \mathrm{CB}_{i,j} \geq 0 \text{ and } j \text{ departs from } i\text{'s arrival station} \right\}. \tag{21}$$

We now define the expected passenger on-time probability for every flight $i \in \mathcal{F}_a$, OTPP_i, as the probability that passengers traveling on flight i will make their connections (if they have a connection) and arrive at their final destination airport by the scheduled arrival time, appropriately weighted by the average number of passengers making those connections, i.e.,

$$\mathrm{OTPP}_i = \frac{w_{ii} \cdot \mathrm{OTPF}_i^L + \sum_{j \in \mathcal{C}_i} w_{ij} \cdot \mathrm{OTPC}_{i,j} \cdot \mathrm{OTPF}_j^L}{w_{ii} + \sum_{j \in \mathcal{C}_i} w_{ij}}. \tag{22}$$

Using (22) we define average passenger on-time probability for an airline $a \in \mathcal{A}$, $\overline{\mathrm{OTPP}}_a$ as

$$\overline{\mathrm{OTPP}}_a = \frac{\sum_{i \in \mathcal{F}_a} \left(w_{ii} \cdot \mathrm{OTPF}_i^L + \sum_{j \in \mathcal{C}_i} w_{ij} \cdot \mathrm{OTPC}_{i,j} \cdot \mathrm{OTPF}_j^L \right)}{\sum_{i \in \mathcal{F}_a} \left(w_{ii} + \sum_{j \in \mathcal{C}_i} w_{ij} \right)} \quad \forall\, a \in \mathcal{A}. \tag{23}$$

It is important to note that $\overline{\mathrm{OTPP}}_a$ also serves to address the issue that $\overline{\mathrm{OTPF}}_a^{DOT}$ does not distinguish between airlines carrying different loads of connecting and local passengers.

Ranking airlines using this adjusted metric accounts for the inherent differences in the markets served by individual airlines. In Figure 8 we plot the values of $\overline{\mathrm{OTPP}}_a$ and $\overline{\mathrm{OTPF}}_a^L$ for all the airlines in our data set. As is readily observable $\overline{\mathrm{OTPF}}_a^L > \overline{\mathrm{OTPP}}_a$ for most airlines. While the observation is not surprising per se, the magnitude of the difference is more important, since it changes the relative rankings within the airlines if we compare them based on $\overline{\mathrm{OTPP}}_a$. We plot the magnitude of this difference between these two measures in Figure 9. For example, observe the difference for Northwest Airlines. Among the major airlines, Northwest shows the biggest drop based on passenger on-time arrival probability. The ranking for Southwest Airlines, too, is lower when measured on connection probabilities, while the ranking for Continental Airlines improves significantly once passenger connections are taken into account. Our observations suggest that $\overline{\mathrm{OTPP}}_a$ is an important measure of passenger service reliability.

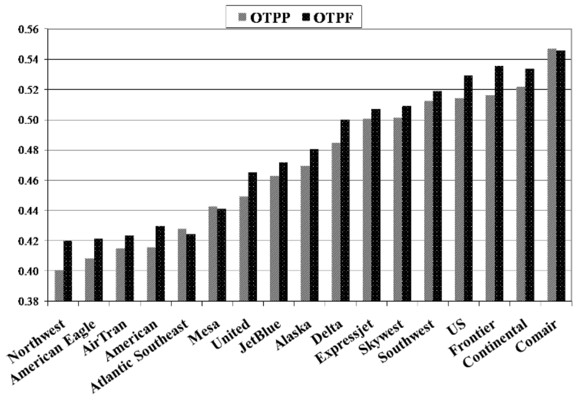

Figure 8 $\overline{\text{OTPP}}_a$ and $\overline{\text{OTPF}}_a^L$.

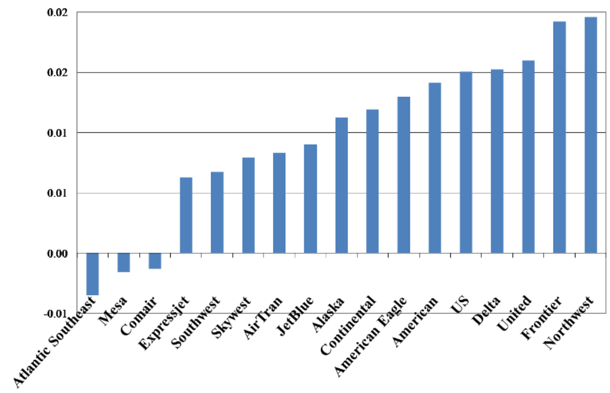

Figure 9 $\overline{\text{OTPF}}_a^L - \overline{\text{OTPP}}_a$.

5. Analysis and policy implications

The goal in this section is to compare major airlines and airports in the U.S. based on the robustness measures developed in §4 and to develop policy recommendations for improving the overall operational performance of the domestic air transportation system. We begin by analyzing airline on-time performance and related policy implications.

5.1. Airline on-time performance analysis

On-time performance is a combination of spill-over delays and intrinsic flight delays. Earlier, in §4 we argued that it is important to separate the effect of spill-over delays and showed how we could separate their impacts. From an airline's perspective it is important to control both of these effects to improve the overall on-time performance. To capture the impact of such spill-over delays, in sections 4.1 and 4.2 we developed two robustness measures $\overline{\text{OTPD}}_a$ and $\overline{\text{FEPD}}_a$ respectively. These measures allow us to rank airlines based on the average spill-over delays observed within their networks, i.e., how "tightly coupled" are these airline networks. By "tightly coupled" we mean how much slack is available, in the form of block- and ground-time buffers within the airline network, to absorb spill-over delays.

Thus, the first issue airlines face is the allocation of ground- and block-time buffer within aircraft rotations. Different airlines, depending on their network structures and the nature of service provided, allocate these buffers differentially. In Figures 10 and 11 we show the average ground-time buffer and block-time buffer for different airlines in our data set. The trade-offs for allocating block-time and ground-time buffers are complex. Increasing block-time buffer can increase planned costs substantially and reduce resource utilization, while decreasing block-time could directly impact the flight's on-time performance, i.e., OTPF_i^L. Increasing ground-time buffer is relatively less expensive

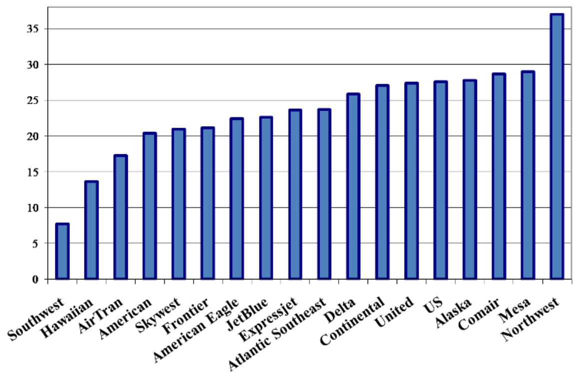

Figure 10 Average ground-time buffer allocation.

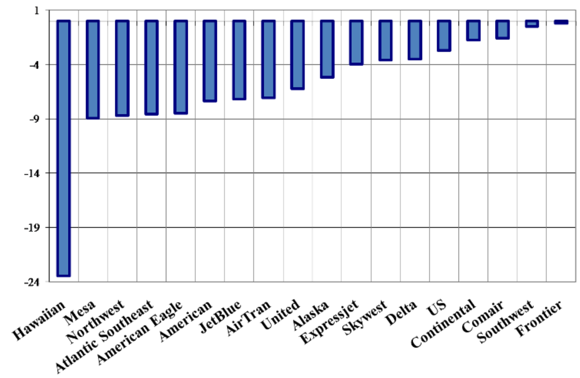

Figure 11 Average block-time buffer allocation.

but decreases resource utilization. Moreover, allocation of these buffers also depends on several other factors pertaining to the particular airline's flight schedule (network). For example, consider the case of Southwest Airlines. Southwest Airlines allocates low ground-time buffers yet maintains sufficient block-time buffers. Southwest Airlines is predominantly a point-to-point carrier, i.e., does not sell a large fraction of passenger itineraries with connections, and mostly operates to and from lesser congested airports. This allows Southwest Airlines to tightly control its ground-time buffers to maintain high resource (such as aircraft and ground equipment) utilization. Allocating a slightly higher buffer on block-times allows Southwest Airlines to maintain a high OTPF_a^L. At the other extreme, consider the example of Northwest Airlines. Northwest Airlines maintains a much larger ground-time buffer yet does not allocate enough buffer on its block-time. Plausible reasons are the hub-and-spoke network, congested hub airports, and larger fraction of connecting passenger itineraries at hub airports.

The second important issue concerning airlines is the identification of bottleneck flights. Earlier, in §4.3 we developed a metric to identify a bottleneck flight by isolating the impact of a particular flight in propagating delays through the network. As we showed in Figures 6 and 7, bottleneck flights are spread throughout the day, i.e., need not correspond to the earlier flights in the day. Ross and Swain (2007) indicates that airline planners focus on the "first" flights of the day to prevent propagated delays. While that is a good strategy, focusing on mid-day flights is essential, too.

From a policy perspective, we make several recommendations based on our analysis. It seems that airlines carry very little block time buffers for their flights, with all airlines averaging negative block-time buffers. This implies that, on the average, the scheduled block-time allocated for a flight is not sufficient to cover the actual block-time. We speculate that this scheduling behavior is partly in response to how on-time performance is measured and reported by the DOT, i.e., by adding a 15 minute buffer to the scheduled arrival time. If this is the case, an immediate conclusion is to modify the measurement of on-time performance of flights so that on-time truly means "on-time", i.e., a flight should be considered delayed if it arrives past its scheduled arrival time.

Second, our stochastic model is better suited to isolate the impact of spill-over/propagated delay on schedule robustness. While recent research has provided models on properly allocating slack in schedules to make them robust, such models have been primarily deterministic. Our preliminary analysis shows that if a stochastic model is used to allocate ground and block-time buffers, it does a better job of making the schedule robust while minimizing the amount of buffer that is needed. Sohoni et al. (2009) discuss an optimization model that incorporates block-time uncertainty in building robust schedules.

Finally, our model also provides guidelines to airlines on which bottleneck flights they should focus on in order to minimize the network impact. We constructed a network impact metric for each flight using our stochastic model. In order to make their schedules more robust, airlines should allocate larger resources (crew, block-times, etc.) to flights that have large network impact.

5.2. Airport congestion analysis

Airports are a critical resource in air transportation networks and impact airline operations substantially. To understand their impact on the average propagated delays through the entire airline network, we compute the average values of overall on-time performance of flights, and intrinsic on-time performance of flights, within every departure time window t. We represent these by $\overline{\text{OTPF}}_t^L$ and $\overline{\text{OTPF}}_t^I$ respectively, computed across all airlines a in our data set, where t represents a particular departure time window. These plots are shown in Figures 12 and 13 respectively. As is

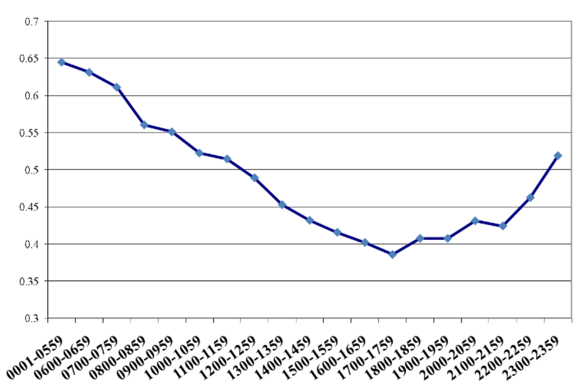

Figure 12 $\overline{\text{OTPF}}_t^L$.

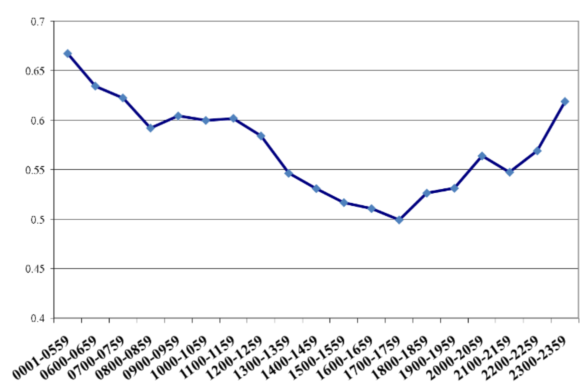

Figure 13 $\overline{\text{OTPF}}_t^I$.

readily observable from Figure 12 the average on-time performance steadily decreases throughout the day until 6:00 p.m. and begins to improve thereafter. However, if we remove the impact of spill-over delays (Figure 13) the performance of flights improves between 9:00 a.m. and 1:00 p.m. These observations suggest two important issues (*i*) spill-over delays get worse throughout the day, and (*ii*) even after adjusting for spill-over delays, flights continue to perform badly as the day progresses. Figure 14 shows the average drop in on-time performance, across all the airlines in our data set, for different scheduled departure time blocks. We denote this average drop by $\overline{\text{OTPD}}_t$. It

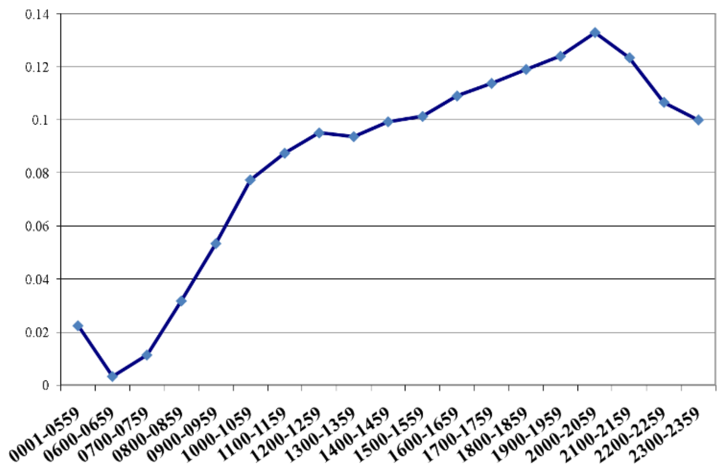

Figure 14 $\overline{\text{OTPD}}_t$ for different scheduled departure times.

is evident that flight performance dips as the day progresses. Some of the reasons contributing to these delays are (*i*) airport and airspace congestion, and (*ii*) scheduling practices of most airlines, i.e., airlines tend to schedule their flights close to each other to compete for market share. In this section we focus on the impact airport congestion. To this end, we rank airports in our data set on two dimensions, (*i*) largest network impact (volume), and (*ii*) largest network impact per flight.

As is readily observable from Figures 15 and 16 the top five airports based on total network impact are ORD (Chicago O'Hare), ATL (Atlanta Hartsfield-Jackson), DFW (Dallas Fort Worth), PHX (Phoenix Sky Harbor), and EWR (Newark Liberty), while the top five airports which have the largest network impact per flight flown are JFK (New York JFK), HNL (Honolulu), ORD (Chicago O'Hare), EWR (Newark Liberty), and SEA (Seattle Tacoma). From a resource allocation perspective, and improving overall operational performance of the flight networks of all airlines, elevating the congestion at these airports is essential.

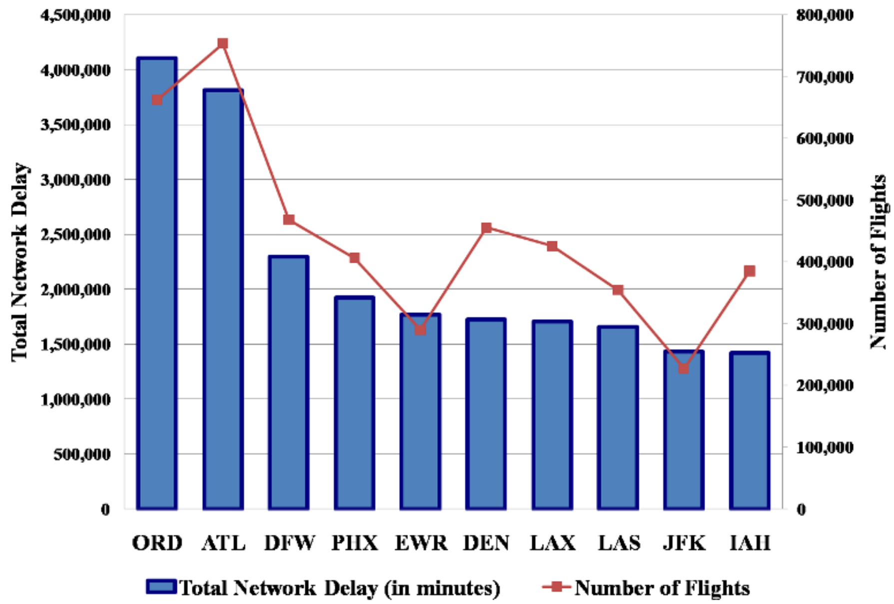

Figure 15 Top 10 U.S. airports with largest total network delay.

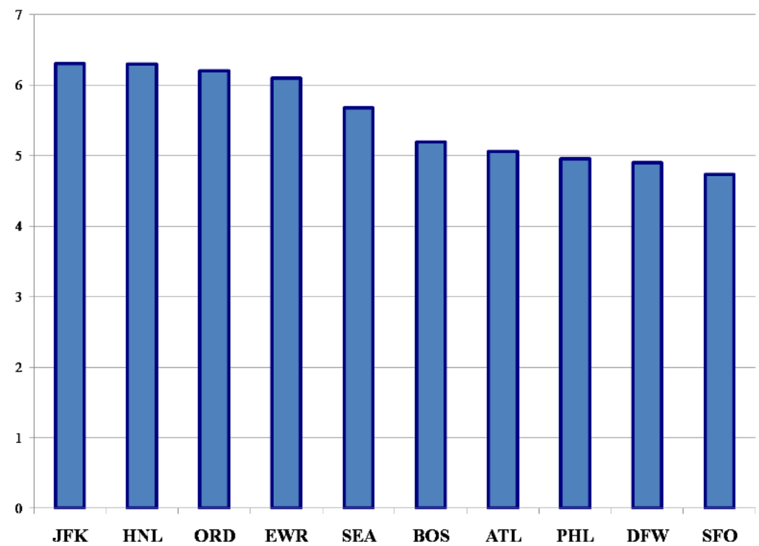

Figure 16 Top 10 U.S. airports with largest per flight network delay.

To conduct an analysis on the impact of airport congestion on the airline networks, we increased the airport capacities of top five airports ORD, ATL, DFW, PHX, EWR, and JFK[12] by 15%, separately one-by-one, and then measured the network impact of increasing these capacities. First, by increasing the capacity at any airport, all flights arriving and departing from that airport experience reduced delays due to lower congestion at that airport. In addition to this direct impact,

[12] JFK has the highest impact per flight flown but not in the top 5 airports list based on total network impact.

our model also captures the indirect impact of increasing the airport capacity, i.e., reduced spill-overs on subsequent flights flown by the aircraft on its subsequent route. In Table 4, we provide the total network impact reduction of increasing the capacity by 15% at these six airports.

Airport	**Current Total Network Impact (minutes)**	**Total Network Impact After 15% Capacity Increase (minutes)**	**Network Impact Reduction (minutes)**	**Network Impact Reduction (%)**
ORD	4,105,041	3,663,916	441,125	10.74
ATL	3,809,529	3,381,949	427,580	11.22
DFW	2,293,861	2,086,537	207,323	9.03
EWR	1,764,704	1,595,101	169,603	9.61
PHX	1,921,950	1,737,380	184,570	9.60
JFK	1,430,190	1,296,270	133,570	9.36

Table 4 Change in the network impact of 6 airports with 15% capacity increase

As can be seen, the network impact is different at different airports. The total network impact reduction in flight delays is largest by adding capacity at Chicago, O'Hare airport, while in terms of percentage reduction in network delays, Atlanta Hartsfield-Jackson airport displays the largest reduction. Complete results are shown in Table 4.

While we do not include the cost of adding capacity at each airport, our model provides a new tool for evaluating investments in either capacity expansions at each airport, or in evaluating the benefits of new technologies (e.g., modernized landing and tracking technologies, En Route Automation Modernization program (ERAM), and etc.) assessed by the FAA under its Air Traffic Control (ATC) and Next Generation (NextGen) modernization program. Since, our model is at a micro (flight) level, the benefits of such capacity expansion proposals on network delays can be easily evaluated (ATC Modernization and NextGen: Near-Term Achievable Goals 2009).

5.3. Improving schedule reliability for passengers

As mentioned earlier, the DOT measures a flight to be on-time as long as the flight arrives at its destination no more than 15 minutes after its scheduled arrival time. This on-time performance metric is computed at an aggregate level across all flights within an airline's network. No specific importance is given to passenger connections and load-factors on these flights, i.e., all flights are treated equally within the network. To address this issue, earlier in §4.4 we developed the service robustness metric $\overline{\text{OTPP}}_a$. In this section we study policy implications for airlines and other authorities to improve upon the DOT metric.

To do so we analyze the dependence of $\overline{\text{OTPP}}_a$ on the departure time during the day. We compute the average $\overline{\text{OTPP}}_a$ across all the flights within a particular departure time window, for all the airlines in our data set, and denote the value of $\overline{\text{OTPP}}_t$, where t denotes a particular departure time window. Figure 17 shows our results. As is readily observable from Figure 17, the probability

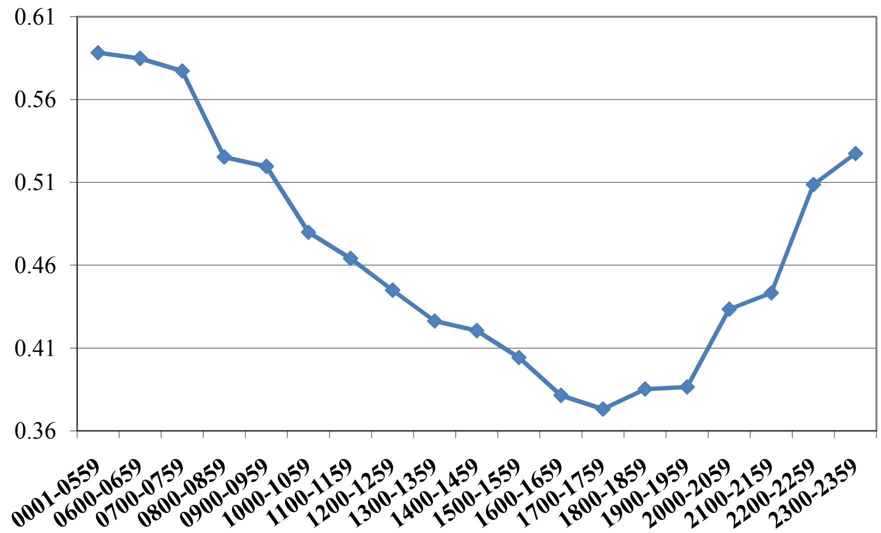

Figure 17 $\overline{\text{OTPP}}_t$ for different departure time windows.

of making connections dips substantially up until 6:00 p.m. and only starts to improve later in the evening. This observations has several implications for planners. How should airlines restructure their networks to provide higher reliability? This ties back to the two issues raised earlier: (i) allowing better allocation of ground-buffer and block-times to increase reliability, and (ii) reducing congestion at airports by changing scheduled departure times of flights or augmenting capacity.

Our recommendation is that performance metrics should be modified to include passenger on-time arrival probability, and not just flight on-time arrival probability. Passenger on-time probability is driven by airline network structures, as well as connecting itineraries taken by passengers. Since a sample of this data is also publicly available, our proposed metric is reasonably easy to compute.

Another important insight of this analysis is that airlines should focus on passenger on-time arrival probability in order to measure the robustness of their schedules, and not just flight on-time arrival probability. We provide a stochastic model to compute the passenger on-time arrival probability based on airline network structures and connecting itineraries chosen by passengers. Airlines can use this model to build robust schedules from the passengers perspective.

6. Concluding Remarks

Flight delays have been a growing issue and they have reached an all-time high in recent years, with the airlines' on-time performance at its worst level in 2007 since 1995. A recent report by the Joint Economic Committee of the US Congress (Schumer and Maloney 2008) has estimated that the total cost to the US economy due to flight delays was as much as $41 billion in 2007.

The goal of this was paper was to build stochastic models and use publicly available data to examine the impact of airline network structures and schedules on the reliability of the air-travel infrastructure. Our analysis was focused on answering the following questions: Which network based passenger-centric metrics could be used by the FAA to measure on-time performance and schedule robustness? Which are the bottleneck airports in the US air-travel infrastructure (i.e., airports that cause most delay propagation)? How would increasing airport capacity at these airports alleviate delay propagation? Which airlines have the least robust schedules? How could these schedules be made more robust? Which flight in an aircraft rotation is a bottleneck flight (and, hence, deserves managerial attention)?

Flight delays are typically attributed to two factors: (*i*) The randomness in the intrinsic block-time for a scheduled flight (which is the actual block-time excluding propagated delays), and (*ii*) the propagation of this randomness through the air-travel network and infrastructure. Our stochastic model captures both of these factors that cause travel delays.

Using this stochastic model, we construct several metrics that measure the robustness of airline networks. In particular, these metrics are: (*i*) average drop in on-time arrival due to spillover/propagated delays, (*ii*) the fraction of total expected delay attributable to spill-over/propagated delays, (*iii*) the network impact of a flight delay, i.e., the cumulative impact of a flight delay on all subsequent flights, and (*iv*) passenger on time arrival probability.

Using these metrics, we make several policy recommendations. Our analysis suggests that the DOT on-time metric should be modified so that on-time really means "on-time", i.e., the 15 minute buffer provided in the on-time metric should be eliminated. This would potentially result in the elimination of negative block-time buffers currently used by all airlines. We also recommend that performance metrics should be modified to include passenger on-time arrival probability, and not just flight on-time arrival probability. Passenger on-time probability is driven by airline network structures, as well as connecting itineraries taken by passengers. Since a sample of this data is also publicly available, we propose a metric that computes passenger on-time arrival probability that is easy to compute.

From the airlines perspective, our stochastic model is better suited to isolate the impact of spill-over/propagated delay on flight schedule robustness. While recent research has provided models on properly allocating slack in schedules to make them robust, such models have been primarily deterministic. Our preliminary analysis shows that if a stochastic model is used to allocate ground and block-time buffers, it does a better job of making the schedule robust while minimizing the amount of buffer that is needed. By providing a stochastic model to compute the passenger on-time arrival probability based on airline network structures and connecting itineraries chosen by passengers, we provide a useful tool to airlines to build robust schedules from passengers perspective. Our model also provides guidelines to airlines on which bottleneck flights they should focus on in order to minimize the total network impact. In order to make their schedules more robust, airlines should allocate larger resources (e.g., crew, block-times, etc.) to flights that have a large network impact.

We also provide an analysis of alleviating the impact of congestion by augmenting airport capacities at different airports. This can be implemented through a variety of ways such as adding runways, using modernized landing techniques, better equipment, etc. Our analysis shows that alleviating the impact of congestion at Chicago O'Hare and Atlanta Hartsfield-Jackson airports would provide the largest network benefit to the domestic air-travel system.

Overall, the contribution of this paper is two-fold. First, we develop stochastic models, using empirical data, to analyze the propagation of delays through air-transportation networks. A key feature of our analysis is that we utilize data available on the entire US domestic network consisting of all airports and all major airlines, instead of focusing on a few airports or a few airlines. Another key feature of our model is that we conduct analysis at a micro (individual flight) level, instead of an aggregate level. Such a model is useful in evaluating the impact of changes made to one part of the system on the entire network. Also, our stochastic models allow us to develop three important robustness measures for airline networks. Second, our analysis enables us to make policy recommendations regarding managing bottleneck resources in the air-travel infrastructure, which if addressed, could lead to a significant improvement in air-travel reliability.

References

AhmadBeygi, S., A. Cohn, Y. Guan, P. Belobaba. 2008a. Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management* **14** 221–236.

AhmadBeygi, S., A. Cohn, M. Lapp. 2008b. Decreasing airline delay propagation by re-allocating scheduled slack. *Working paper.* University of Michigan, MI.

ATC Modernization and NextGen: Near-Term Achievable Goals. 2009. 111th Congress Subcommittee on Aviation Hearing.

Ball, M., C. Barnhart, G. Nemhauser, A. Odoni. 2007. *Air transportation: Irregular operations and control. C. Barnhart, G. Laporte, eds.*, *Handbooks in Operations Research and Management Science: Transportation*, vol. 14, chap. 1. North-Holland, 1–61.

Barnhart, C., A. Cohn. 2004. Airline schedule planning: Accomplishments and opportunities. *Manufacturing & Service Operations Management* **6**(1) 3–22.

Bratu, S., C. Barnhart. 2005. An analysis of passenger delays using flight operations and passenger booking data. *Air Traffic Control Quarterly* **13**(1) 1–27.

Carey, M. 1999. Ex ante heuristic measures of schedule reliability. *Transportation Research Part B: Methodological* **33** 473–494.

Caulkins, J.P., A. Barnett, P.D. Larkey, Y. Yuan, J. Goranson. 1993. The on-time machines: Some analyses of airline punctuality. *Operations Research* **41**(4) 710–720.

Deshpande, V., M. Arıkan. 2009. The impact of airline flight schedules on flight delays: An empirical study. *Working paper.* Purdue University, IN.

Hebert, J.E., D.C. Dietz. 1997. Modeling and analysis of an airport departure process. *Journal of Aircraft* **34**(1) 43–47.

Lan, S. 2003. Planning for robust airline operations: Optimizing aircraft routings and flight departure times to achieve minimum passenger disruptions. Ph.D. thesis, Massachusetts Institute of Technology.

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

Lapp, M., S. AhmadBeygi, A. Cohn, O. Tsimhoni. 2008. A recursion-based approach to simulating airline schedule robustness. S. J. Mason, R. Hill, L. Moench, O. Rose, eds., *Proceedings of the 2008 Winter Simulation Conference*.

McCartney, S. 2007, August 13. Frequent flying: Small jets, more trips worsen airport delays. The Wall Street Journal, pp. A1.

McCartney, S. 2007, May 29. Why flights are getting longer. The Wall Street Journal, pp. D1, D5.

McCartney, S. 2010, February 4. Why a six-hour flight now takes seven. The Wall Street Journal, pp. D1.

Mueller, E.R., G.B. Chatterji. 2002. Analysis of aircraft arrival and departure delay characteristics. Tech. rep., AIAA Aircraft Technology, Integration, and Operations (ATIO).

Ross, A., A. Swain. 2007. Fighting flight delays. *OR/MS Today* **34**(2).

Schumer, C. E., C. B. Maloney. 2008. Your flight has been delayed again: Flight delays cost passengers, airlines, and the us economy billions. *The US Senate Joint Economic Committee* .

Sherry, L., D. Wang, G. Donohue. 2007. Air travel consumer protection: Metric for passenger on-time performance. *Transportation Research Record: Journal of the Transportation Research Board* **2007** 22–27.

Shumsky, R. A. 1995. Dynamic statistical models for the prediction of aircraft take-off times. Ph.D. thesis, Massachusetts Institute of Technology.

Sohoni, M., Y-C. Lee, D. Klabjan. 2009. Robust airline scheduling under block bime uncertainty. *Working paper.* Indian School of Business, Hyderabad, India.

The MITRE Corporation. 2007. Capacity needs in the national airspace system: An analysis of airports and metropolitan area demand and operational capacity in the future. Tech. rep., Federal Aviation Administration.

Tien, S., B. Subramanian, M. Ball. 2009. Constructing a passenger trip delay metric: An aggregate-level approach. *Working paper.* University of Maryland, MD.

Tu, Y., M. Ball, W.S. Jank. 2008. Estimating flight departure delay distributions-a statistical approach with long-term trend and short-term pattern. *Journal of the American Statistical Association* **103**(481) 112–125.

Wang, D. 2007. Methods for analysis of passenger trip performance in a complex networked transportation system. Ph.D. thesis, George Mason University.

Willemain, T. R., H. Ma, N. V. Yakovchuk, W. A. Child. 2003. Factors influencing estimated time en route. Tech. rep., The National Center of Excellence for Aviation Operations Research (NEXTOR).

Wu, C.L. 2006. Improving airline network robustness and operational reliability by sequential optimisation algorithms. *Networks and Spatial Economics* **6**(3) 235–251.

7. Appendices

7.1. Appendix A: Proofs

Proof of Proposition 1.

Deshpande and Arıkan (2009) showed that $\log(D_i) \sim \text{Laplace}\left(\widehat{\mu}_i, 2\widehat{b}_i^2\right)$, so $D_i \sim \text{log-Laplace}\left(\frac{e^{\widehat{\mu}_i}}{1-\widehat{b}_i^2}, e^{2\widehat{\mu}_i}\left(\frac{1}{1-4\widehat{b}_i^2} - \frac{1}{(1-\widehat{b}_i^2)^2}\right)\right)$. Given (5), we have

$$\mathbb{E}[D_i^L] \quad = \mathbb{E}[D_i] + \mathbb{E}[D_{i-1}^L - (Q_{i-1} + B_i)]^+, \tag{1a}$$

$$\text{Var}[D_i^L] = \text{Var}[D_i] + \text{Var}[D_{i-1}^L - (Q_{i-1} + B_i)]^+. \tag{1b}$$

We assume that D_{i-1}^L follows a log-Laplace distribution with parameters $\tilde{\mu}_{i-1}$ and $\tilde{b}_{i-1}$, then $D_i^L \sim \text{log-Laplace}\left(\frac{e^{\tilde{\mu}_i}}{1-\tilde{b}_i^2}, e^{2\tilde{\mu}_i}\left(\frac{1}{1-4\tilde{b}_i^2} - \frac{1}{(1-\tilde{b}_i^2)^2}\right)\right)$. Let $\tilde{f}_{i-1}(x)$ denote the probability distribution function of D_{i-1}^L and define $\tilde{Q}_{i-1} = Q_{i-1} + B_i$. Then,

$$\mathbb{E}[D_{i-1}^L - (Q_{i-1} + B_i)]^+ = \mathbb{E}[D_{i-1}^L - \tilde{Q}_{i-1}]^+ = \int_{\tilde{Q}_{i-1}}^{\infty} (x - \tilde{Q}_{i-1}) \tilde{f}_{i-1}(x) dx \tag{1c}$$

and

$$\int_{\tilde{Q}_{i-1}}^{\infty} (x - \tilde{Q}_{i-1}) \tilde{f}_{i-1}(x) dx = \begin{cases} \frac{e^{\tilde{\mu}_{i-1}}}{1-\tilde{b}_{i-1}^2} - \tilde{Q}_{i-1} + \frac{\tilde{b}_{i-1} e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1+\tilde{b}_{i-1}}{\tilde{b}_{i-1}}}}{2(1+\tilde{b}_{i-1})}, & \text{if } e^{\tilde{\mu}_{i-1}} > \tilde{Q}_{i-1} \\ \frac{\tilde{b}_{i-1} e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{\tilde{b}_{i-1}-1}{\tilde{b}_{i-1}}}}{2(1-\tilde{b}_{i-1})}, & \text{otherwise.} \end{cases}$$

So, we can rewrite (1a) as follows:

$$\mathbb{E}[D_i^L] = \frac{e^{\widehat{\mu}_i}}{1-\widehat{b}_i^2} + \begin{cases} \frac{e^{\tilde{\mu}_{i-1}}}{1-\tilde{b}_{i-1}^2} - \tilde{Q}_{i-1} + \frac{\tilde{b}_{i-1} e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1+\tilde{b}_{i-1}}{\tilde{b}_{i-1}}}}{2(1+\tilde{b}_{i-1})}, & \text{if } e^{\tilde{\mu}_{i-1}} > \tilde{Q}_{i-1} \\ \frac{\tilde{b}_{i-1} e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{\tilde{b}_{i-1}-1}{\tilde{b}_{i-1}}}}{2(1-\tilde{b}_{i-1})}, & \text{otherwise.} \end{cases}$$

Let $K = \mathbb{E}[D_{i-1}^L - \tilde{Q}_{i-1}]^+$, then

$$\text{Var}[D_{i-1}^L - (Q_{i-1} + B_i)]^+ = \text{Var}[D_{i-1}^L - \tilde{Q}_{i-1}]^+ = \mathbb{E}\left[\left(D_{i-1}^L - \tilde{Q}_{i-1} - K\right)^2\right]^+. \tag{1d}$$

Expanding the last term in (1d), we obtain

$$\begin{aligned} \text{Var}[D_{i-1}^L - \tilde{Q}_{i-1}]^+ &= \mathbb{E}\left[\left(D_{i-1}^L - \tilde{Q}_{i-1}\right)^2 - 2\left(D_{i-1}^L - \tilde{Q}_{i-1}\right) K + K^2\right]^+ \\ &= \int_{\tilde{Q}_{i-1}}^{\infty} \left((x - \tilde{Q}_{i-1})^2 - 2K(x - \tilde{Q}_{i-1}) + K^2\right) \tilde{f}_{i-1}(x) dx \end{aligned}$$

$$= \int_{\tilde{Q}_{i-1}}^{\infty} (x-\tilde{Q}_{i-1})^2 \tilde{f}_{i-1}(x)dx - 2K^2 + K^2 Pr(D_{i-1}^L \geq \tilde{Q}_{i-1})$$

since $\int_{\tilde{Q}_{i-1}}^{\infty}(x-\tilde{Q}_{i-1})\tilde{f}_{i-1}(x)dx = \mathbb{E}[D_{i-1}^L - \tilde{Q}_{i-1}]^+ = K$ and $\int_{\tilde{Q}_{i-1}}^{\infty} \tilde{f}_{i-1}(x)dx = Pr(D_{i-1}^L \geq \tilde{Q}_{i-1})$.

Note that

$$Pr(D_{i-1}^L \geq \tilde{Q}_{i-1}) = \begin{cases} 1 - \frac{e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1}{\tilde{b}_{i-1}}}}{2}, & \text{if } e^{\tilde{\mu}_{i-1}} > \tilde{Q}_{i-1} \\ \frac{e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{-1}{\tilde{b}_{i-1}}}}{2}, & \text{otherwise} \end{cases}$$

and

$$\int_{\tilde{Q}_{i-1}}^{\infty}(x-\tilde{Q}_{i-1})^2\tilde{f}_{i-1}(x)dx = \begin{cases} \frac{e^{2\tilde{\mu}_{i-1}}}{1-4\tilde{b}_{i-1}^2} + \tilde{Q}_{i-1} - \frac{\tilde{b}_{i-1}^2 e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1+2\tilde{b}_{i-1}}{\tilde{b}_{i-1}}}}{(1+2\tilde{b}_{i-1})(1+\tilde{b}_{i-1})}, & \text{if } e^{\tilde{\mu}_{i-1}} > \tilde{Q}_{i-1} \\ \frac{\tilde{b}_{i-1}^2 e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{2\tilde{b}_{i-1}-1}{\tilde{b}_{i-1}}}}{(2-\tilde{b}_{i-1}-1)(\tilde{b}_{i-1}-1)}, & \text{otherwise.} \end{cases}$$

Therefore, we can rewrite (1b) as follows:

$$\begin{aligned} \text{Var}[D_i^L] &= e^{2\widehat{\mu}_i}\left(\frac{1}{1-4\widehat{b}_i^2} - \frac{1}{(1-\widehat{b}_i^2)^2}\right) - 2K^2 \\ &+ \begin{cases} \frac{e^{2\tilde{\mu}_{i-1}}}{1-4\tilde{b}_{i-1}^2} + \tilde{Q}_{i-1} - \frac{\tilde{b}_{i-1}^2 e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1+2\tilde{b}_{i-1}}{\tilde{b}_{i-1}}}}{(1+2\tilde{b}_{i-1})(1+\tilde{b}_{i-1})} + K^2\left(1 - \frac{e^{\frac{-\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{1}{\tilde{b}_{i-1}}}}{2}\right), & \text{if } e^{\tilde{\mu}_{i-1}} > \tilde{Q}_{i-1} \\ \frac{\tilde{b}_{i-1}^2 e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{2\tilde{b}_{i-1}-1}{\tilde{b}_{i-1}}}}{(2-\tilde{b}_{i-1}-1)(\tilde{b}_{i-1}-1)} + K^2\left(\frac{e^{\frac{\tilde{\mu}_{i-1}}{\tilde{b}_{i-1}}} \tilde{Q}_{i-1}^{\frac{-1}{\tilde{b}_{i-1}}}}{2}\right), & \text{otherwise.} \end{cases} \end{aligned}$$

To develop analytic expressions for $\tilde{\mu}_i$ and $\tilde{b}_i$, we perform a second-order matching. To do so, we solve the following expressions for the mean and variance together.

$$\mathbb{E}[D_i^L] = \left(\frac{e^{\tilde{\mu}_i}}{1-\tilde{b}_i^2}\right), \tag{1e}$$

$$\text{Var}[D_i^L] = e^{2\tilde{\mu}_i}\left(\frac{1}{1-4\tilde{b}_i^2} - \frac{1}{(1-\tilde{b}_i^2)^2}\right). \tag{1f}$$

Let $A = \frac{\text{Var}[D_i^L]}{\mathbb{E}[D_i^L]^2} = \left(\frac{(1-\tilde{b}_i^2)^2}{1-4\tilde{b}_i^2} - 1\right)$ and $S = \tilde{b}_i^2$. Then,

$$\begin{aligned} A + 1 &= \frac{(1-S)^2}{1-4S} \\ (A+1)(1-4S) &= 1 - 2S + S^2 \\ 0 &= S^2 + 2S(1+2A) - A. \end{aligned} \tag{1g}$$

By solving the quadratic equation (1g) we obtain:

$$S = \tilde{b}_i^2 = -(1+2A) + \sqrt{(4A+1)(A+1)}. \tag{1h}$$

Using (1e) and (1h) we obtain

$$\tilde{\mu}_i = \ln\left(\mathbb{E}[D_i^L](1-\tilde{b}_i^2)\right).$$

♦

7.2. Appendix B: Bad Data Examples

ORIGIN	DEST	DATE	ACTUAL DEP-TM	ACTUAL ARR-TM	ACTUAL-DIFF	POSITION
ATL	LAS	1/30/2005	8:36	13:25	35:48	1
LAS	ATL	1/30/2005	14:20	21:00	0:55	2
ATL	FLL	1/30/2005	16:15	17:52	0:00	.
ATL	FLL	1/30/2005	22:33	0:15	1:33	3
FLL	ATL	1/31/2005	6:50	8:54	6:35	1

Table 5 A duplicate record

ORIGIN	DEST	DATE	ACTUAL DEP-TM	ACTUAL ARR-TM	ACTUAL-DIFF	POSITION
VPS	ATL	1/15/2005	6:10 AM	8:40 AM	.	1
ATL	CHA	1/15/2005	1:35 PM	2:10 PM	4:55	2
CHA	ATL	1/15/2005	3:00 PM	3:55 PM	0:50	3
PFN	ATL	1/15/2005	6:20 PM	8:21 PM	.	.
ATL	VPS	1/15/2005	10:00 PM	10:20 PM	6:05	1

Table 6 Wrong origin or destination

ORIGIN	DEST	DATE	ACTUAL DEP-TM	ACTUAL ARR-TM	ACTUAL-DIFF	POSITION
ORD	IAH	2/09/2007	13:00	15:21	0:55	1
IAH	ORD	2/09/2007	16:27	.	.	2

Table 7 Missing data

7.3. Appendix C: Accuracy of the Second Moment Approximation

The stochastic model discussed in §3.2 is tested using a simulation model. We use a randomly selected Southwest Airlines rotation with 6 consecutive flights flown in January 2007. For each flight i in this rotation, 60,000 truncated block-times are randomly generated from a log-Laplace distribution with parameters $\widehat{\mu}_i$ and $\widehat{b}_i$. Considering the scheduled arrival times, each simulated flight in the rotation is assigned an indicator variable of 0 or 1 depending on whether the flight arrived on time or not (i.e., 1 represented an on-time and 0 represented a delayed flight). For each of the 6 flights in the rotation, simulated on-time averages are calculated across 60,000 replications using the on-time indicators. Then, these simulated on-time averages (i.e., simulated on-time probabilities) are compared to the estimated on-time probabilities which are calculated by using $\tilde{\mu}_i$ and $\tilde{b}_i$ obtained from our two-moment approximation. The maximum difference between simulated on-time probabilities and estimated on-time probabilities is equal to 0.038 among these 6 flights. The absolute value of the average difference is 0.012. The simulated on-time probabilities turn out to be very close (i.e., within 4% range) to the estimated on-time probabilities. Table 8 displays the details of the simulation study explained above. Similar results were obtained by repeating this experiment across different rotations for different airlines in our data-set.

Position	$\widehat{\mu}_i$	$\widehat{b}_i$	$\tilde{\mu}_i$	$\tilde{b}_i$	**On-time Prob. (Estimated)**	**On-time Prob. (Simulated)**	**Absolute Difference**
1	4.53	0.10	4.51	0.10	0.687	0.684	0.003
2	4.02	0.11	4.03	0.13	0.703	0.726	0.023
3	4.04	0.10	4.05	0.12	0.357	0.396	0.039
4	5.12	0.10	5.10	0.10	0.789	0.779	0.010
5	5.08	0.10	5.07	0.11	0.513	0.540	0.027
6	5.15	0.10	5.17	0.11	0.690	0.691	0.001

Table 8 Results of the simulation study