

An Assessment of the Impact of Demand Management Strategies for Efficient Allocation of Airport Capacity

Abstract

Airport demand management strategies have the potential to mitigate congestion and delays. However, the extent to which the delays can be reduced using such strategies is not clear. In this paper, we develop a bound on the minimum possible level of delays that can be achieved using these strategies. We solve the aggregated timetable development and fleet assignment problem to minimize the system-wide delays assuming a single monopolistic carrier that satisfies all the passenger demand in the US and maintains the same level-of-service as achieved with the current revenue-management practices of individual carriers. The problem is formulated as a large-scale, integer linear programming model and solved using linear programming relaxation and heuristics. A network delay simulator is used for calculating the delays under different capacity scenarios. The results indicate the large inefficiencies in the usage of airport infrastructure in the domestic US caused by competitive airline scheduling decisions.

1. Introduction

Departure time and fares are vital aspects of an airline's ability to attract passengers and market share (Belobaba, 2009). In addition to these two factors, on-time performance and service reliability play a critical role in maintaining and improving an airline's profitability (Bratu and Barnhart, 2005). The attractiveness of an airline, especially to the high-yield business passengers, is dependent on service reliability (Belobaba and Simpson, 1982). Low reliability of service adversely affects airline revenues. Not only are flight delays known to drive away high-yield passengers (Flint, 2000), but poor on-time performance is also found to have a negative impact on airfares (Januszewski, 2002). Airlines engage in revenue management practices in different ways, such as providing adequate seating capacity at desired time periods; creating schedules with sufficient slack to improve on-time performance and schedule reliability; and engaging in differential pricing and market segmentation. This paper addresses the first two of these three issues at a system-wide level and assesses the industry-wide impact of the interplay between the same.

Growing congestion at major US airports translates into several billion dollars of delay costs each year. Total flight delays rose sharply during much of the current decade. Although the current economic recession has led to airline schedule reductions and consequently resulted in delay reduction over the last couple of years, large delays are expected to return as soon as the economic crisis subsides (Tomer and Puentes, 2009). Aircraft delays result in passenger delays and discomfort, as well as additional fuel consumption and green house gas emissions. Various studies have estimated the total economic impacts of delays and the numbers differ considerably depending on the assumptions and scope of each study (ATA, 2008; Schumer and Maloney,

2008; Sherry and Donohue, 2008). A recently concluded study commissioned by the Federal Aviation Administration (FAA) provides a comprehensive assessment of the costs and impacts of flight delays in the US (Ball et al., 2010). It has estimated that the total cost of US air transportation delays in 2007 was \$32.9 billion. The delays are imposing a tremendous cost on the overall economy.

According to Bureau of Transportation Statistics (BTS, 2008), delays to approximately 50% of the delayed flights in 2007 were due to the National Aviation System (NAS). The leading cause of NAS delays was weather, which resulted in 65.60% of such delays, while volume was the second most important cause responsible for 18.91% of the NAS delays. Volume related delays are due to scheduling more operations than available capacity, while weather related delays are due to capacity reduction in adverse weather conditions causing the realized capacity to drop below demand. Both these types of delays are due to scheduling more operations than the realized capacity and hence can be termed as delays due to demand-capacity mismatch. Demand-capacity mismatch thus contributed to 84.51% of the NAS delays in 2007, making it a primary cause of flight delays in United States.

Capacity enhancement and demand reduction are the two natural ways of alleviating this demand-capacity mismatch. Capacity enhancement measures require a longer time horizon and more investment compared to demand management strategies. Demand management can reduce delays by reducing the demand for airport resources without affecting the number of passengers being transported. However, the extent to which the delays can be reduced using demand management is not clear. In this research, we assess the maximum possible impact of these demand management strategies. All the analysis is based on extensive amounts of publicly available data on airline schedules, passenger flows and airport capacities. Detailed flight and passenger flow information was obtained from the Bureau of Transportation Statistics website (BTS, 2008). Actual realized airport capacity values for one entire year were used for calculation of flight delays.

Before proceeding further, let us define some terminology that will be used frequently in this paper. In all our models, a market is defined as a passenger origin and destination pair. A segment is defined as an origin and destination pair for non-stop flights. A path is defined as a sequence of segments along which a passenger is transported from origin to destination. A flight leg is defined as a combination of origin, destination, departure time and arrival time of a non-stop flight. An itinerary is a sequence of flight legs along which a passenger is transported from origin to destination. We will refer to the actual network of flights operated by multiple airlines in the US domestic markets in 2007 as the *existing network*. Also we will refer to our delay minimizing network as the *single airline (or SA) network*.

The rest of this paper is organized as follows. Section 2 provides a review of demand management strategies and explains the motivation behind solving the single airline scheduling problem. Section 3 briefly reviews the literature on airline scheduling and highlights the

important differences between the previous research and the problem at hand. Section 4 provides a detailed problem statement for this study. Section 5 describes the modelling framework. Section 6 outlines the set of algorithms used to solve the problem. Section 7 provides details of data sources and implementation. A summary of results is provided in section 8. Finally, we conclude with a discussion of the main contributions and the directions for future research in section 9.

2. Airport Demand Management

The passenger demand for air travel in the US domestic markets has been increasing rapidly over the last couple of decades and is expected to double within the next 10 to 15 years (Ball et al., 2007b). Delays are expected to outpace the demand considerably with each 1% increase in airport traffic expected to bring about a 5% increase in delays (Schaefer et al., 2005). In spite of the increase in passenger demand, the average aircraft size in terms of the seating capacity in the US domestic flights is shrinking rapidly. According to the analysis by Bonnefoy and Hansman (2008), the average aircraft size has shrunk by more than 30% over the 17 year period from 1991 to 2007. This roughly means that more flights are necessary to transport the same number of passengers. This phenomenon of proliferation of smaller aircraft is closely connected with frequency competition among competing airlines. Higher frequency shares are believed to be associated with disproportionately higher market shares because the two follow what is commonly known as an S-shaped relationship (Belobaba, 2009). Therefore, there is a tendency among the competing airlines to match non-stop frequencies in key markets to retain market share. As a result, airlines tend to favour large numbers of flights using small aircraft rather than a smaller number of flights using larger aircraft. Runway capacity is usually the most restrictive element at major airports and it is also the predominant cause of the most extreme instances of delays (Ball et al., 2007b). From the runway capacity perspective, takeoff (and landing) of a small aircraft requires almost the same time as that of a large aircraft. Hence frequency competition between airlines is responsible for aggravating airport congestion and delays.

Demand management strategies refer to any administrative or economic regulation that restricts airport access to users. It should be noted that these strategies refer to managing the demand for flight arrival and departure slots at an airport to meet a given level of passenger demand, and not to managing passenger demand itself. Few of the most congested US airports, such as Kennedy (JFK), Newark (EWR) and Laganardia (LGA) Airports in New York City area, O'Hare (ORD) Airport at Chicago and Reagan (DCA) Airport at Washington DC, have been slot controlled in one way or the other for a long time. Current slot allocation strategies, based on administrative controls, are inefficient because of large barriers to market entry (DEL, 2001) and use-it-or-lose-it rules that encourage over-scheduling practices (Harsha, 2008). In response to these shortcomings, market-based mechanisms such as slot auctions and congestion pricing have been proposed as efficient means of reducing congestion. There is extensive literature (Ball et al., 2007a; Ball, Donohue, and Hoffman, 2006; Daniel, 1992; DEL, 2001; Fan and Odoni, 2001;

Grether, Isaac, and Plott, 1979; 1989) suggesting that market-based approaches, if designed properly, allocate scarce resources efficiently and promote fair competition. Harsha (2008) shows that market-based mechanisms can lead to airline schedule changes that reduce the demand for runway capacity, without reducing the number of passengers transported. This is achieved by better utilization of capacity in off-peak hours and by greater usage of larger aircraft.

In theory, these pricing and auction mechanisms should bring the demand and supply into balance by removing the inefficiencies in the system. However, the extent to which the system-wide delays can be reduced by these mechanisms is still unclear. On one hand, restricting airport utilization to a very low level can practically ensure the absence of congestion related delays, but this could mean that the airport is highly underutilized and all the passenger demand might not be satisfied. On the other hand, scheduling a very large number of operations can satisfy all the passenger demand but the delays could reach unacceptable levels. An important question is what minimum level of airport utilization and delays needs to be permitted in order to satisfy all the passenger demand.

In this research, we measure the extent to which airport capacity in the US domestic air transportation network is being inefficiently utilized. The aim is to build a schedule that would minimize the delays in the absence of frequency competition. In order to obviate the effects of competition, we assume a single airline that will satisfy all the passenger demand without compromising the level-of-service for passengers. A network delay simulator (Odoni and Pyrgiotis, 2009; 2010) is used to estimate the delays for the resulting network. The delay values for the single airline network are compared with those for the existing network under various realistic capacity reduction scenarios. All the days in an entire year are divided into 5 categories based on the total duration of capacity reduction on that day across all the busy airports. One representative day from each category is chosen for delay calculations. These delay estimates serve as theoretical lower bounds on the system delays when airport capacity is allocated most efficiently. The maximum possible delay reduction will indicate the maximum potential impact of implementing efficient demand management techniques. If insignificant, then passenger demand has already reached a level where large delays are inevitable and capacity enhancement is the only realistic means of delay reduction. On the other hand, if the results suggest substantial delay reduction under the single airline case, then the existing level of passenger demand can be efficiently served using the existing infrastructure with much lower delays and there is ample opportunity for congestion mitigation using demand management strategies.

3. Airline Schedule Development

The airline schedule development process includes decisions regarding daily frequency, departure times, aircraft sizes and crew schedules. Due to the enormous size and complexity of the airline schedule development process, the problem is typically broken down into four stages: 1) timetable development; 2) fleet assignment; 3) maintenance routing; and 4) crew scheduling (Barnhart, 2009). The task of deciding the set of flight legs to be operated along with the

corresponding origin, destination and departure time for each leg is called timetable development. Although the entire timetable development problem can be modelled as an optimization problem, practitioners typically focus on incremental changes to existing schedules. The fleet assignment problem involves a profit maximizing assignment of fleet types to flight legs. Hane et al. (1995) proposed a leg-based fleet assignment model which assumes independent leg demand and average fares. Jacobs, Johnson, and Smith (1999) and Barnhart, Knicker, and Lohatepanont (2002) proposed itinerary-based fleet assignment models that produce significant profit improvement over the leg-based models. Maintenance routing is the assignment of specific aircraft to individual flight legs while satisfying the periodic aircraft maintenance requirements. The maintenance routing problem is typically solved as a feasibility problem or as a through-revenue maximization problem (Barnhart, 2009). The problem of assigning a cost minimizing combination of pilot and cabin crews to each flight leg is called crew scheduling. Because of the complicated duty rules and pay structure for airline crews, crew pairing has long been regarded as a challenging problem. Crew pairing is modelled as a set partitioning problem and solved using techniques such as column generation and branch-and-price (Barnhart et al., 1998b; Desrosiers et al., 1995; Klabjan, Johnson, and Nemhauser, 2001).

The aforementioned sequential solution process may result in suboptimal solutions. Many researchers have tried to integrate some of the stages into simultaneous optimization problems. Rexing et al. (2000) proposed joint models for flight retiming within time windows and fleet assignment. Lohatepanont and Barnhart (2004) present an integrated model for incremental schedule development and itinerary-based fleet assignment. Clarke et al. (1996) and Barnhart et al. (1998a) have proposed models to incorporate the effect of maintenance routing while making the fleet assignment decisions. There is a large body of literature on the integration of maintenance routing and crew scheduling problems including Cohn and Barnhart (2003), Cordeau et al. (2001) and Klabjan et al. (2002).

All the models mentioned above aim to produce precise schedules that maximize planned profit. Models developed for the purpose of this study are different in several important ways. Rather than producing an operable schedule, the main purpose is to obtain a bound on delays. Because of the complex and stochastic relationship between schedule and delays, any such bound will have to be approximate. Therefore there is no point in developing a very precise schedule. Moreover, the problem of single airline schedule development is even larger in size than the schedule development problem for any existing airline, which itself is solved sequentially due to tractability issues. Therefore in this study, we use aggregate models that are sufficient for our purposes while maintaining tractability. Instead of profit maximization, the objective is delay minimization subject to satisfaction of demand and level-of-service requirements. Therefore, only the relevant decisions such as timetable development and fleet assignment are included in the problem. The output of our models is a flight schedule with departure times and fleet types corresponding to each flight. We do not solve the maintenance routing and crew scheduling problems in this research. Harsha (2008) has proposed an aggregated, integrated airline

scheduling and fleet assignment model (AIASFAM) to help airlines place a bid in a slot auction. The itinerary-based version of this model is an extension of the Barnhart, Knicker, and Lohatepanont (2002) model, with more aggregate time-line discretization for computational tractability. The models presented in this study share some characteristics with the AIASFAM model.

4. Problem Statement

In order to obtain a lower bound on airport congestion, we assume the existence of a single monopolistic airline. The problem at hand is to design a schedule for this single airline with the objective of minimizing airport congestion, while satisfying the entire passenger demand and maintaining a comparable level-of-service. An important modelling consideration is how to capture the passenger demand satisfaction requirements for every market and every time period of the day. The single carrier must be able to transport all passengers who are currently transported by existing airlines, from their respective origins to their respective destinations. To model this, we divide the day into four time intervals and ensure that all the passengers who are currently transported during a particular interval continue to be transported during the same interval in the new schedule. We define the level-of-service (as perceived by air passengers) as the number of stops in an itinerary. Almost 97.6% of all US domestic passengers travelled on non-stop and one-stop itineraries in 2007 (BTS, 2008). Hence, in the single airline scheduling model, we assume that all the passengers must be transported on itineraries with at most one stop.

Service frequency is another important criterion of level-of-service in the current competitive environment. Therefore, in our model, we require that the single airline provide at least the same daily frequency on each non-stop segment as the *effective frequency* provided by the existing carriers. Cohas, Belobaba, and Simpson (2005) propose a model of effective frequency available to air passengers faced with a choice between multiple competing carriers. When more than one airline operates in a market, the effective frequency depends on how closely the schedules are matched. For example, consider two competing carriers, each offering n flights per day. If one airline schedules flights at a time when the other airline does not offer service, then the effective frequency increases. However, if one airline schedules all its flights close to the departure times of flights by the other airlines, then the number of different options to the passengers does not increase above n . Thus, the important criterion in deciding the effective frequency is the closeness of competing airline schedules. We calculate the effective non-stop frequency for a segment as the total number of non-stop flights offered by all carriers as long as the flight departure times are not within less than one hour of each other. If the departure times of two flights are separated by less than one hour, then we assume the two flights to be equivalent to a single flight. The minimum frequency to be provided on each non-stop segment by the single airline must be greater than or equal to the effective frequency currently provided by all the existing carriers on that segment. This constraint ensures that the passengers experience the same

or higher effective frequency in each market. This constraint combined with the time-of-the-day demand satisfaction criteria also ensures that there is negligible shift in the passengers' arrival and departure times in comparison to the desired values of the same.

Before designing a schedule, decisions must be taken regarding the network structure. Network design involves decisions about network type i.e. hub-and-spoke or point-to-point, choice of hubs, choice of non-stop segments, choice of allowable airports for passenger connections. One possible approach would be to include all these decisions into our single-airline optimization problem. For the problem size under consideration, that would lead to an integer optimization problem involving over one-hundred million variables. Instead, we solve the problem sequentially in three stages.

5. Modelling Framework

Figure 1 provides a schematic description of the overall modelling framework. The first stage is the Network Design (ND) stage, which involves decisions about the number and location of hubs, candidates for non-stop routes and allowable airports for passenger connections. The network structures of existing airlines were used as a guideline for our network design stage. Many of the major airlines in US domestic market today have a set of 4 or 5 major hubs. The direct flights are allowed to bypass the hub for a few important markets with large demand. Our selection of hubs was made based on qualitative criteria including the number of operations in the existing network, available capacity, geographic location and weather. Atlanta (ATL), Denver (DEN), Dallas/Fort Worth (DFW) and Chicago O'Hare (ORD) were chosen because they are in the top five US airports in terms of both existing capacity as well as the number of operations in the existing network. None of the airports in the New York area were chosen because of their low capacities. Los Angeles (LAX) was not chosen because of its geographically extreme location in the continental US. Phoenix (PHX) was chosen because of large number of operations in the existing network and the maximum capacity being available for a large fraction of the time due to good weather conditions. Our choices of non-stop segments bypassing the hub were made based on the market demand corresponding to the non-stop segments. Any market with a daily demand of at least 250 passengers was included as a candidate for non-stop flights. We allow passengers to connect only at the hubs.

The second stage involves the daily Frequency Planning and Fleet Assignment (FPFA) problem. A delay minimizing schedule should have fewer flights per day and better distribution of flight timings to avoid clustering of demand near peak hours. Obtaining a good feasible solution is the main aim of our FPFA stage. A good solution will keep the number of flights to a minimum, so that airport usage is minimized. We tried using a variety of formulations with different objective functions for this stage. Our initial modelling efforts for this stage showed that there are multiple optimal solutions that minimize the total number of flights but differ in terms of the amount of slack in the seating capacity. In order to produce an efficient schedule it is important to choose the most appropriate aircraft size for each segment so as to avoid excessive seating capacity. We

achieved this by choosing cost coefficients (denoted by $c_{s,k}$ for segment s and fleet k) such that the overall cost increases with increasing seating capacity and cost per seat decreases with increasing seating capacity. These costs are consistent with those that the airlines report through Form 41 financial reports (BTS, 2008). Satisfaction of the daily demand and the minimum daily frequency requirement are the two main constraints. Output of this second stage includes the daily frequency of service on each segment and fleet types assigned to each segment. Constraints (1) through (5) provide integer programming formulation for the FPFA problem.

FPFA Formulation:

Notation:

K = Set of fleet types

S = Set of segments

P = Set of paths

M = Set of markets

$c_{s,k}$ = Operating cost of fleet type k on segment s , $s \in S$ and $k \in K$

C_k = Seating capacity of fleet type k , $k \in K$

D_m = Daily demand in market m , $m \in M$

f_s = Minimum daily frequency to be provided for segment s , $s \in S$

$P(m)$ = Set of paths associated with market m , $m \in M$

$\delta_s^p = \begin{cases} 1, & \text{if path } p \text{ contains segment } s \\ 0, & \text{otherwise} \end{cases} \quad p \in P \text{ and } s \in S$

Decision variables:

$x_{s,k}$ = No. of flights of fleet type k on segment s per day, $s \in S$ and $k \in K$

y_p = No. of passengers on path p per day, $p \in P$

Formulation:

$$\text{Minimize } \sum_{s \in S} \sum_{k \in K} c_{s,k} * x_{s,k}$$

Subject to:

$$\sum_{p \in P(m)} y_p = D_m, \forall m \in M \quad (1)$$

$$\sum_{k \in K} x_{s,k} * C_k \geq \sum_{p \in P} \delta_s^p * y_p, \forall s \in S \quad (2)$$

$$\sum_{k \in K} x_{s,k} \geq f_s, \forall s \in S \quad (3)$$

$$x_{s,k} \in \mathbb{Z}^+, \forall s \in S \text{ and } k \in K \quad (4)$$

$$y_p \in \mathbb{Z}^+, \forall p \in P \quad (5)$$

Constraint (1) ensures that the total daily demand for each market is satisfied. Constraint (2) ensures that the total number of seats on each segment is sufficient for carrying all the passengers whose paths contain that segment. Constraint (3) enforces that the daily frequency of service in each market is at least equal to the effective frequency currently provided by the existing carriers in that market. Constraints (4) and (5) restrict the allowable values for the number of passengers and number of flights to non-negative integers. Alternatively, the number of passengers could be modelled as continuous variables without significant impact on solution quality. The integrality of variables corresponding to the number of flights, however, is critical to obtaining meaningful solutions.

The third stage involves actual Timetable Development (TD). Similar to the approach adopted by Harsha (2008), the departure and arrival times are aggregated to the nearest hour to keep the number of decision variables low. Given the daily frequencies and fleet assignment for each segment, output of this stage produces the scheduled set of flight legs. Constraints (6) through (11) provide an integer programming formulation of the TD model.

The utilization ratio is defined as the ratio of demand to capacity of a server, which in this case is an airport. Queuing theory suggests that the average flight delay is an increasing and convex function of the utilization ratio (Larsen and Odoni, 2007). Considering the tremendous size of the problem at hand, using a non-linear objective function would make the problem intractable. Total delay is a nonlinear and stochastic function of the number of scheduled flights. Therefore, we aim to minimize the maximum utilization ratio as a surrogate objective function for the scheduling problem. Due to the convex relationship between the utilization ratio and delays, the effect of the maximum utilization ratio on total delay in a queuing network is disproportionately high. The objective function in the TD formulation is to minimize the maximum utilization ratio across all busy airports across all airport-time period (ATP) pairs. The duration of each ATP is 1 hour in this case. The hourly utilization ratio is the ratio of the sum of all flight frequencies corresponding to that ATP to the hourly capacity of the airport. Thus, the maximum utilization ratio is a deterministic and linear function of the flight frequencies. Constraint (6) enforces the satisfaction of demand for each market-time period (MTP) pair. Constraint (7) ensures that the total number of seats for each flight leg is at least equal to the total number of passengers whose itineraries contain that flight leg. Constraint (8) ensures that the minimum daily frequency

requirement is satisfied. Constraint (9) relates the maximum utilization ratio to the operations in each ATP. Constraints (10) and (11) restrict the possible values for the number of passengers and the frequencies to non-negative integers.

TD Formulation:

Notation:

A = Set of airports

I = Set of itineraries

F = Set of flight legs

$L(m)$ = Set of MTPs associated with market m , $m \in M$

D_l = Demand in MTP l , $l \in L(m)$ and $m \in M$

C_f = Seating capacity for fleet type assigned to flight leg f , $f \in F$

f_s = Minimum daily frequency to be provided for segment s , $s \in S$

HC_t = Hourly capacity (i.e. maximum total number of operations) for ATP t , $t \in T(a)$ and $a \in A$

$T(a)$ = Set of ATPs associated with airport a , $a \in A$

$I(l)$ = Set of itineraries associated with MTP l , $l \in L(m)$ and $m \in M$

$F(s)$ = Set of flight legs associated with segment s , $s \in S$

$$\delta_f^i = \begin{cases} 1, & \text{if itinerary } i \text{ contains flight leg } f \\ 0, & \text{otherwise} \end{cases} \quad i \in I \text{ and } f \in F$$

$$\gamma_f^t = \begin{cases} 1, & \text{if ATP } t \text{ is utilized by flight leg } f \\ 0, & \text{otherwise} \end{cases} \quad t \in T \text{ and } f \in F$$

Decision variables:

x_f = Frequency for flight leg f , $f \in F$

y_i = No. of passengers on itinerary i , $i \in I$

r_{max} = Maximum utilization ratio for airport hourly capacities

Formulation:

Minimize r_{max}

Subject to:

$$\sum_{i \in I(l)} y_i = D_l, \forall l \in L(m), m \in M \quad (6)$$

$$x_f * C_f \geq \sum_{i \in I} \delta_f^i * y_i, \forall f \in F \quad (7)$$

$$\sum_{f \in F(s)} x_f \geq f_s, \forall s \in S \quad (8)$$

$$\sum_{f \in F} \gamma_f^t * x_f \leq r_{max} * HC_t \forall t \in T(a), a \in A \quad (9)$$

$$x_f \in \mathbb{Z}^+, \forall f \in F \quad (10)$$

$$y_i \in \mathbb{Z}^+, \forall i \in I \quad (11)$$

6. Solution Algorithm

As mentioned earlier, due to large problem size, obtaining an exact solution is difficult. Additionally, because of the aggregate nature of our analysis, approximate solution methods are sufficient. We solve the FPFA linear programming (LP) relaxation and round up the resulting solution to the nearest integer values greater than or equal to the LP optimal solution. Due to the nature of the constraints in FPFA formulation, none of the constraints is violated if segment frequencies are increased.

Solution to the FPFA problem involves determining daily frequency values, which are relatively large integers. The impact of rounding up is comparatively small. But for the TD problem, the solutions are highly fractional because the hourly frequencies are much smaller than daily frequency values. Therefore, the rounding up procedure worsens the objective function dramatically. Much of the solution's non-integrality stems from the markets in which demand is extremely low per day. Therefore, the LP solution has very small fractions of flight legs serving small markets, and the solution is not of sufficient quality for our purposes. Therefore, we solved the TD problem in two steps. The TD solution procedure is described schematically in Figure 2. In the first step, the TD LP relaxation was solved for a smaller sub-problem involving all the markets with a daily demand of at least 250 passengers. These constituted over 60% of the total demand. These markets include all the candidates for non-stop service bypassing the hub. This LP solution was rounded upward to the nearest integers. Due to the nature of constraints in the TD formulation, an increase in value of any x variable in a feasible solution does not affect feasibility. Moreover, most of flights in these important markets serve as connecting flights for smaller markets. Therefore the additional seating capacity made available due to flight rounding is very likely to be utilized to carry passengers in remaining smaller markets. The remaining problem was solved using a greedy heuristic, as depicted in Figures 3 and 4.

In the first step of the heuristic, as described in Figure 3, additional non-stop flights are scheduled to satisfy the demand in markets with a daily demand of less than 250 passengers and with at least one endpoint at a hub (the *Hub Markets*). The markets are processed one after the other in decreasing order of demand. Additional flights are scheduled such that the maximum utilization ratio across all the affected ATPs is minimized at each step. Scheduling an additional non-stop flight increases the utilization ratio for the origin airport during the departure hour and also increases the utilization ratio of the destination airport during the arrival hour. Among all the departure time choices for a MTP, the one which minimizes this maximum utilization ratio is chosen.

In the second step of the heuristic, as described in Figure 4, additional flight legs are scheduled to satisfy the demand for markets with a daily demand of less than 250 passengers, where neither endpoint is a hub (the *Non-hub Markets*). The demand for these small, non-hub markets has to be satisfied by one-stop itineraries. The algorithm processes the markets in decreasing order of demand and schedules additional flights on first or second or both legs of an itinerary so as to minimize the maximum utilization ratio among all the affected ATPs.

The optimization algorithm ignores the aircraft flow balance constraints, which is an important component of the airline scheduling procedure. The purpose of this study is not to come up with a schedule that can be operated using actual aircraft fleets- our goal is not to design a monopolistic airline schedule for the US- but rather to find a lower bound on the levels of delays. However, to assess the potential impact of aircraft balance requirements, after completing the optimization process, we performed a post-processing step wherein additional flights were added to the optimal unbalanced schedule in order to balance it. Fortunately, because the passenger demands typically are balanced in both directions of a market, the resulting schedule is not too far from a balanced schedule. In our algorithm, we add one flight at a time from an airport with a surplus of a particular aircraft type to an airport with a deficit of the same aircraft type. The departure time of each additional flight is chosen to greedily minimize the maximum utilization ratio of the affected ATPs at each step. The results presented in the next section provide the statistics on both the balanced and unbalanced single airline (SA) networks, and compare these statistics with the existing network.

Flight delays are estimated using the Approximate Network Delay (AND) model, described by Odoni and Pyrgiotis (2009; 2010) as follows: AND is a stochastic and dynamic queuing model. It has two main components, a queuing engine (QE) and a delay propagation algorithm (DPA). AND treats a network of airports as a set of interconnected queuing systems where each airport is modelled as an $M(t)/E_k(t)/I$ queuing system. The queuing system is characterized by a non-stationary Poisson arrival process, time-dependent k^{th} -order Erlang service-time distribution and a single server with infinite queuing capacity. AND iterates between QE and DPA by tracking each flight in the network and updating the airport demand profiles based on revised flight arrival and departure times.

7. Data Sources and Implementation Details

Schedules of major US domestic airlines were obtained from the Airline On-Time Performance Database provided by the Bureau of Transportation Statistics (BTS, 2008). A 10% sample of the passengers carried by each airline per quarter is provided in the Airline Origin and Destination Survey (DB1B) available on the same website. An estimate of the total passenger demand per OD market is obtained by multiplying the DB1B passenger number by 10. The top 20 airlines in the US, which constitute over 95% of the total traffic, are considered for the analysis. Demand for each market per day is divided into 4 time periods that roughly correspond to morning peak period, off-peak period, evening peak period and another off-peak period. Given the daily demands, the Decision Window Model by the Boeing Airplane Company (BAC, 1997) is used to calculate the demand in each MTP. Apart from these 20 airlines, there are several other flights such as cargo, general aviation and international flights which are not included in our analysis. We assume that these remaining flights continue to be operated as they currently are. We simply add those operations to the total operations at each airport for calculating the utilization ratios.

The Federal Aviation Administration has published benchmark capacity values for the 35 most congested US airports (FAA, 2004). For each one of those airports, the report provides 3 values of capacity along with the corresponding probabilities of their realization, based on weather, wind and other conditions. The capacity is measured as the maximum number of operations (takeoffs and landings) possible at an airport per hour. Based on the capacity distributions for these airports, we use capacity values for each airport that correspond to the maximum capacity that is available for at least 95% of the time at that airport. We use these capacities only for calculating the utilization ratios. The delay calculation is based on the actual realized capacity values under each weather scenario.

The analysis is carried out for Tuesday, 16th October 2007. Because this analysis is aggregate in nature, we use only three generic fleet types. We call them Wide Body (WB), Narrow Body (NB) and Regional Jets (RJ). The seating capacity values for these three fleet types are chosen to be the average seating capacities of the corresponding aircraft in the 2007 fleet of the seven largest US airlines.

Realized hourly capacity data for the 35 busiest US airports was compiled for the entire year from April 2007 to March 2008 in order to generate realistic capacity reduction scenarios. Based on the total duration of capacity reduction across the 35 airports, we divided the 366 days in the year into 5 categories, namely, very good, good, normal, bad and very bad, each containing approximately the same number of days. A typical (median) day belonging to each category was chosen and delays were computed using the corresponding realized capacity values.

8. Results

For our single airline network, the maximum utilization ratio over all airports over all ATPs was found to be 90.77%. In comparison, the maximum utilization ratio in the existing network was 160%. Table 1 provides the cumulative distribution of utilization ratios across the 35 busiest airports, across 24 hours of the day. The first column contains the percent utilization ratios and the next three columns show the number of times that value was exceeded in the existing network, unbalanced single airline (SA) network and balanced single airline (SA) network respectively. Each combination of airport and hour is counted as one observation. In case of the single airline networks, compared to the existing network, there is a substantial reduction in the hourly utilization ratios at the busy airports, which means that there would be substantially lower delays when bad weather reduces airport capacity.

Utilization Ratio	Existing Network	Unbalanced Single Airline Network	Balanced Single Airline Network
>150%	1	0	0
> 140%	3	0	0
> 130%	8	0	0
> 120%	19	0	0
> 110%	28	0	0
> 100%	55	0	0
> 90%	76	1	1
> 80%	133	60	83
> 70%	196	131	151
> 60%	275	196	210
> 50%	350	268	292

Table 1: Comparison of Airport Utilization Ratios

Table 2 compares some important metrics for the unbalanced and balanced single airline networks to those for the existing network. US domestic flights in October 2007 had an average load factor of 78.45% (BTS, 2008). The load factor for the single airline network is lower, reflecting that the single airline network design problem had less flexibility in selecting the most appropriate aircraft for each flight leg due to aggregation of aircraft sizes into 3 broad categories. We define block hours as the difference between the scheduled arrival time and scheduled departure time of a flight. The unbalanced single airline network requires 26% fewer flights and 8% lower total block hours than the existing network. For the year 2007, the DB1B Market database (BTS, 2008) shows that approximately 31.98% of the passengers travelling on major US airlines were connecting passengers, which is very close to that for the single airline network.

Metric	Existing Network	Unbalanced Single Airline Network	Balanced Single Airline Network
No. of Flights	20,539	15,231	16,872
Fraction of Connecting Passengers	31.98%	33.77%	33.77%
Load Factor	78.45%	71.77%	64.51%
Total No. of Block Hours	44,103	40,756	45,739

Table 2: Network Performance Metrics

In the single airline network, operations are more evenly spread over the day at the congested airports. We compared the means and standard deviations of hourly utilization ratios from 6 am to midnight at the 35 busiest airports in the US. For the existing network, the average values of the mean and the coefficient of variation (i.e., the ratio of standard deviation to mean) are 36.7% and 56.4% respectively, while for the balanced single airline network they equal 31.5% and 47.8% respectively. So the single airline schedule has not only lower average utilization but also lower variation in the number of operations scheduled. This effect is especially strong for the more congested airports. There are 9 airports, Atlanta (ATL), Washington Reagan (DCA), Newark (EWR), New York Kennedy (JFK), Los Angeles (LAX), New York Laguardia (LGA), Chicago O’Hare (ORD), Seattle-Tacoma (SEA) and San Francisco (SFO), which have a mean utilization of at least 70% between 6 am and midnight in the existing network. For these most congested airports, the mean utilization ratio decreases from 86.5% for the existing network to 68.7% for the balanced single airline network and the coefficient of variation averaged across these airports decreases from 29.6% for the existing network to 16.5% for the balanced single airline network. To illustrate, Figure 5 shows the distribution of the utilization ratio over a day at John F. Kennedy International Airport, New York (JFK), one of the busiest airports in US. This efficient utilization of off-peak hour capacity is one of the factors that contribute to lowering the congestion levels in the single airline network.

The other important factor that contributes to lower congestion is an increase in average aircraft size. Table 3 illustrates the number of flights being operated by each fleet type. The numbers in the parentheses indicate the percentage of all flights corresponding to that particular fleet type. The strong shift towards wide body aircraft can be observed in both the unbalanced and balanced networks. Most of the increase in the number of wide bodies comes from a decrease in usage of narrow bodies.

Fleet Type	Wide Body	Narrow Body	Regional Jets
Existing Network	363 (1.77%)	12788 (62.26%)	7388 (35.97%)
Unbalanced Single Airline Network	3285 (21.57%)	6058 (39.77%)	5888 (38.66%)

Balanced Single Airline Network	3561 (21.11%)	6775 (40.16%)	6536 (38.74%)
---------------------------------	---------------	---------------	---------------

Table 3: Aircraft Size Distributions

Table 4 compares the total aircraft delay under 5 different capacity scenarios. The delay values are computed excluding any propagated delay due to late arriving aircraft. In each capacity scenario, the single airline network produces substantially lower delays, with the delay reduction ranging from 53% to 88%. As the congestion worsens, the absolute, as well as percentage, delay reduction increases. Given the way the scenarios are chosen, each chosen scenario can be considered equally likely. Thus, across different scenarios an average delay reduction of 81.72% can be achieved in the absence of competition. This provides an estimate of the inefficiencies due to the competitive scheduling practices of carriers using the existing airport infrastructure. The results suggest that congestion related delays could be reduced to less than one fifth of the existing level if there was no competition. In other words, competition is responsible for more than a 400% worsening of congestion related delays. Most of the improvement is due to efficient utilization of airport capacity during off-peak hours and a strong shift towards larger aircraft.

Scenario	Existing Network Delay (aircraft-min)	Balanced Single Airline Network Delay (aircraft-min)	Reduction
Very Good	7495.05	3552.97	52.60%
Good	14682.30	4090.06	72.14%
Normal	27998.76	5940.40	78.78%
Bad	35081.44	6289.88	82.07%
Very Bad	64026.52	7421.76	88.41%
Average	29856.82	5459.02	81.72%

Table 4: Total Flight Delay under Various Weather Scenarios

In Figure 6, we show the total number of operations at the 35 busiest airports under the existing network and under the balanced single airline (SA) network. The airports are arranged from left to right in decreasing order of daily utilization ratios in the existing network. The number of operations in the SA network is lower than in the existing network in all but 7 out of these 35 airports, with the total number of operations being 13.9% lower. This effect is especially strong in the cases of the leftmost four airports, ORD, EWR, JFK and LGA, which are the top four airports in terms of the daily utilization ratio, as well as average flight delays. The total number of operations at these 4 airports is 29.6% lower in the SA network than in the existing network. The DFW, DEN and PHX airports, which have comparatively lower utilization ratios and a lot of excess capacity in the existing network, are the only airports with significant increases in the number of operations. These were also chosen as 3 of the 5 hubs in the single airline network

because of these characteristics. The overall increase in operations at these three airports is 62.9%. Finally, operations are substantially reduced at airports such as Salt Lake City (SLC), Cleveland (CLE), Cincinnati (CVG) and Memphis (MEM), which already have low utilization ratios in the existing network. This apparent anomaly stems from the fact that these are the hub airports for at least one major existing network carrier, but none of them has a strong local market with high passenger demand. These existing hubs tend to rely on connecting passenger flow, in the absence of which the number of operations is expected to be lower.

9. Conclusion

In this paper, we calculated a lower bound on system-wide delays that can be achieved in the absence of competition while satisfying all passenger demand and maintaining a comparable level-of-service to that achieved with the current revenue-management practices of individual carriers. Aggregate integrated integer programming models of timetable development and fleet assignment were developed and solved using a heuristic-based approach. Results show that there is significant room for improvement in congestion even with the existing airport infrastructure. Passenger demand has not yet reached a level where delays as high as present are unavoidable. Given the available capacity, efficient administrative controls and/or market-based mechanisms can potentially lead to substantial reductions in airport congestion and delays.

These results provide critical insights into the nature and causes of aviation delays, allowing better planning and utilization of aviation infrastructure. In particular, the results help differentiate between the delays caused due to insufficient capacity and delays caused due to inefficient utilization of capacity. The models and solution methods presented here can also be used for analyzing the best-case delay levels under different future scenarios, with different levels of demand and capacity.

The critical next step is to devise intelligent mechanisms and incentives that will result in airlines gradually migrating their schedules from those in place today towards the delay-minimizing schedules presented in this paper, in the presence of market competition. Vaze and Barnhart (2010) have demonstrated that, under frequency competition between carriers, simple changes to existing administrative slot controls at a congested airport can lead to improved schedule reliability and delay reduction while maintaining a comparable level-of-service through the provision of adequate frequency and seating capacity. Moreover, they also result in a considerable increase in airline profits. This is achieved through reductions in total allocated airport capacity and aircraft upgauges. Market-based slot pricing and auctioning mechanisms, mentioned earlier, are expected to achieve similar improvements in schedule reliability and delays. However, due to the associated monetary payments, the overall impact on airline profits needs to be evaluated carefully while accounting for the effects of airline schedule competition.

References

- ATA (2008). Cost of Delays. Obtained through the Internet: <http://www.airlines.org/economics/specialtopics/ATC+Delay+Cost.htm/>, [accessed 10/12/2008].
- Ball, M.O., Ausubel, L.M., Bernardino, F., Donohue, G., Hansen, M. and Hoffman, K. (2007a) 'Market based alternatives for managing congestion at New York's LaGuardia airport'. *Paper Presented at the AirNeth Annual Conference. April 11-13, 2007*. The Hague, Netherlands.
- Ball, M.O., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A. and Zou, B. (2010) *Total Delay Impact Study Report*. Federal Aviation Administration, Washington D.C.
- Ball, M.O., Barnhart, C., Nemhauser, G. and Odoni, A. (2007b) 'Air transportation: irregular operations and control', *Handbook in OR & MS*, Vol. 14, No. 1, pp.23-38.
- Ball, M.O., Donohue, G. and Hoffman, K. (2006) 'Auctions for the safe, efficient and equitable allocation of airspace resources', *In: Cramton, P., Shoham, Y. and Steinberg, R. (eds.), Combinatorial Auctions*, (pp.507-538), Cambridge: MIT Press.
- Barnhart, C. (2009) 'Airline schedule optimization', *In: Belobaba, P., Odoni, A. and Barnhart, C. (eds.), The Global Airline Industry*, (pp.183-212), West Sussex: Wiley.
- Barnhart, C., Boland, N., Clarke, L., Johnson, E.L., Nemhauser, G., and Sheno, R. (1998a) 'Flight string models for aircraft fleet and routing', *Transportation Science*, Vol. 32, No. 3, pp.208-220.
- Barnhart, C., Johnson, E.L., Nemhauser, G., Savelsbergh, M. and Vance, P. (1998b) 'Branch-and-price: column generation for solving huge integer programs', *Operations Research*, Vol. 46, No. 3, pp.316-329.
- Barnhart, C., Knicker, T., and Lohatepanont, M. (2002) 'Itinerary-based airline fleet assignment', *Transportation Science*, Vol. 36, No. 2, pp.199-217.
- Belobaba, P. (2009) 'Overview of airline economics, markets and demand', *In: Belobaba, P., Odoni, A. and Barnhart, C. (eds.), The Global Airline Industry*, (pp.47-72), West Sussex: Wiley.
- Belobaba, P. and Simpson, R. (1982) 'Air transportation economics'. Working Paper, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.
- BAC (1997) *Decision Window Path Preference Methodology Description*, Seattle, WA.
- Bonnefoy, P.A. and Hansman, R. J. (2008) 'Overview of recent forces and trends in the airline industry and challenges for future operations'. *Paper Presented at the Lufthansa-European Business School Conference. 2008*. London, United Kingdom.

- Bratu, S. and Barnhart, C. (2005) 'An analysis of passenger delays using flight operations and passenger booking data'. *Air Traffic Control Quarterly*, Vol. 13, No. 1, pp.1-28.
- BTS (2008). Airline On-Time Statistics and Delay Causes. Obtained through the Internet: <http://www.bts.gov/>, [accessed 15/12/2008].
- Clarke, L., Hane, C., Johnson, E.L. and Nemhauser, G. (1996) 'Maintenance and crew considerations in fleet assignment', *Transportation Science*, Vol. 30, No. 3, pp.249-260.
- Cohas, F.J., Belobaba, P.P. and Simpson R.W. (1995) 'Competitive fare and frequency effects in airport market share modelling', *Journal of Air Transport Management*, Vol. 2, No. 1, pp.33-45.
- Cohn, A. and Barnhart, C. (2003) 'Improving crew scheduling by incorporating key maintenance routing decisions', *Operations Research*, Vol. 51, No. 3, pp.387-396.
- Cordeau, J., Stojkovic, G., Soumis, F. and Desrosiers, J. (2001) 'Benders decomposition for simultaneous aircraft routing and crew scheduling', *Transportation Science*, Vol. 35, No. 4, pp.375-388.
- Daniel, J. (1992) *Peak-Load-Congestion Pricing and Optimal Capacity of Large Hub Airports: with Application to the Minneapolis St.Paul Airport*. PhD thesis. Massachusetts Institute of Technology, Cambridge MA.
- Desrosiers, J., Dumas, Y., Solomon, M.M. and Soumis, F. (1995) 'Time constrained routing and scheduling', *Handbook in OR & MS Network Routing*, pp.35-139.
- DEL (2001). Auctioning Airport Slots. Obtained through the Internet: <http://www.dotecon.com/publications/slotauactp.pdf/>, [accessed 05/01/2010].
- Fan, T.P. and Odoni, A.R. (2001) 'The potential of demand management as a short-term means of relieving airport congestion'. *Paper Presented at the EUROCONTROL-FAA Air Traffic Management R&D Review Seminar. December, 2001*. Santa Fe NM.
- FAA (2004). Airport Capacity Benchmark Report. Obtained through the Internet: http://www.faa.gov/about/office_org/headquarters_offices/ato/publications/bench/, [accessed 15/12/2008].
- Flint, P. (2000) 'Back on schedule'. *Air Transport World*, pp.47-51.
- Grether, D.M., Isaac, R.M. and Plott, C.R. (1979) *Alternative methods of allocating airport slots: performance and evaluation*. Technical report, Polinomix Research Laboratories, Pasadena CA.
- Grether, D.M., Isaac, R.M. and Plott, C.R. (1989) *The Allocation of Scarce Resources: Experimental Economics and the Problem of Allocating Airport Slots*, Boulder: Westview Press.

- Hane, C.A., Barnhart, C., Johnson, E.L., Marsten, R.E., Nemhauser, G.L., and Sigismondi, G. (1995) 'The fleet assignment problem: solving a large-scale integer program', *Mathematical Programming*, Vol. 70, No. 1, pp.211-232.
- Harsha, P. (2008) *Mitigating Airport Congestion: Market Mechanisms and Airline Response Model*. PhD thesis. Massachusetts Institute of Technology, Cambridge MA.
- Jacobs, T.L., Johnson, E.L., and Smith, B.C. (1999) 'O&D FAM: incorporating passenger flows into the fleet assignment process'. *Paper Presented at the 39th Annual AGIFORS Symposium. October 3-8, 1999*. New Orleans LA.
- Januszewski, S. (2002) 'The effect of air traffic delays on airline prices'. Working Paper, Department of Economics, Massachusetts Institute of Technology, Cambridge MA.
- Klabjan, D., Johnson, E.L. and Nemhauser, G.L. (2001) 'Solving large airline crew scheduling problems: random pairing generation and strong branching', *Computational Optimization and Applications*, Vol. 20, No. 1, pp.73-91.
- Klabjan, D., Johnson, E.L., Nemhauser, G.L., Gelman, E., and Ramaswamy, S. (2002) 'Airline crew scheduling with time windows and plane count constraints', *Transportation Science*, Vol. 36, No. 3, pp.337-348.
- Larsen, R.C. and Odoni, A.R. (2007) *Urban Operations Research*, 2nd Ed., Belmont: Dynamic Ideas.
- Lohatepanont, M. and Barnhart, C. (2004) 'Airline schedule planning: integrated models and algorithms for schedule design and fleet assignment', *Transportation Science*, Vol. 38, No. 1, pp.19-32.
- Odoni, A. and Pyrgiotis, N. (2009) 'A stochastic and dynamic queuing model of a network of major airports'. *Paper Presented at the INFORMS annual meeting. October 11-14, 2009*. San Diego CA.
- Odoni, A. and Pyrgiotis, N. (2010) 'Modelling delay propagation within an airport network'. *Proceedings of World Conference on Transport Research. July 11-15, 2010*. Lisbon, Portugal.
- Rexing, B., Barnhart, C., Knicker, T., Jarrah, A., and Krishnamurthy, N. (2000) 'Airline fleet assignment with time windows', *Transportation Science*, Vol. 34, No. 1, pp.1-20.
- Schaefer, A.J., Johnson, E.L., Kleywegt A.J. and Nemhauser, G.L. (2005) 'Airline crew scheduling under uncertainty', *Transportation Science*, Vol. 39, No. 3, pp.340-348.
- Schumer, C.E. and Maloney, C.B. (2008) *Your flight has been delayed again: flight delays cost passengers, airlines, and the US economy billions*. The US Senate Joint Economic Committee Report, Washington DC.

Sherry, L. and Donohue, G. (2008) *U.S. airline passenger trip delay report*. Centre for Air Transportation Systems Research, George Mason University, Fairfax, VA.

Tomer, A. and Puentes, R. (2009) *Expect delays: an analysis of air travel trends in the United States*. Brookings Metropolitan Infrastructure Initiative Series, Washington DC.

Vaze, V. and Barnhart, C. (2010) 'Efficient utilization of airport capacity under frequency competition'. *Proceedings of World Conference on Transport Research. July 11-15, 2010*. Lisbon, Portugal.

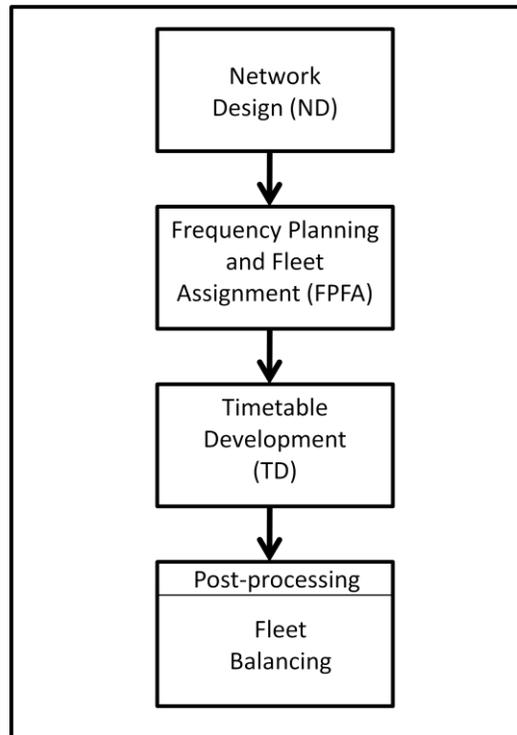


Figure 1: Modelling Framework

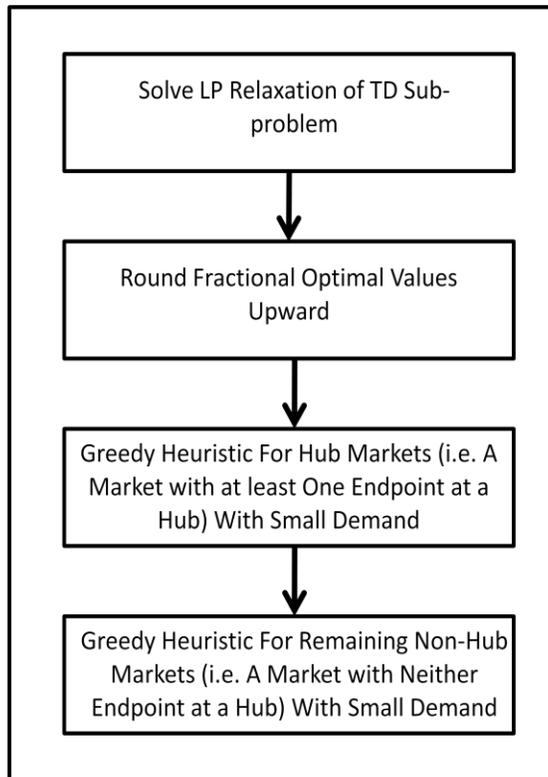


Figure 2: Timetable Development Algorithm

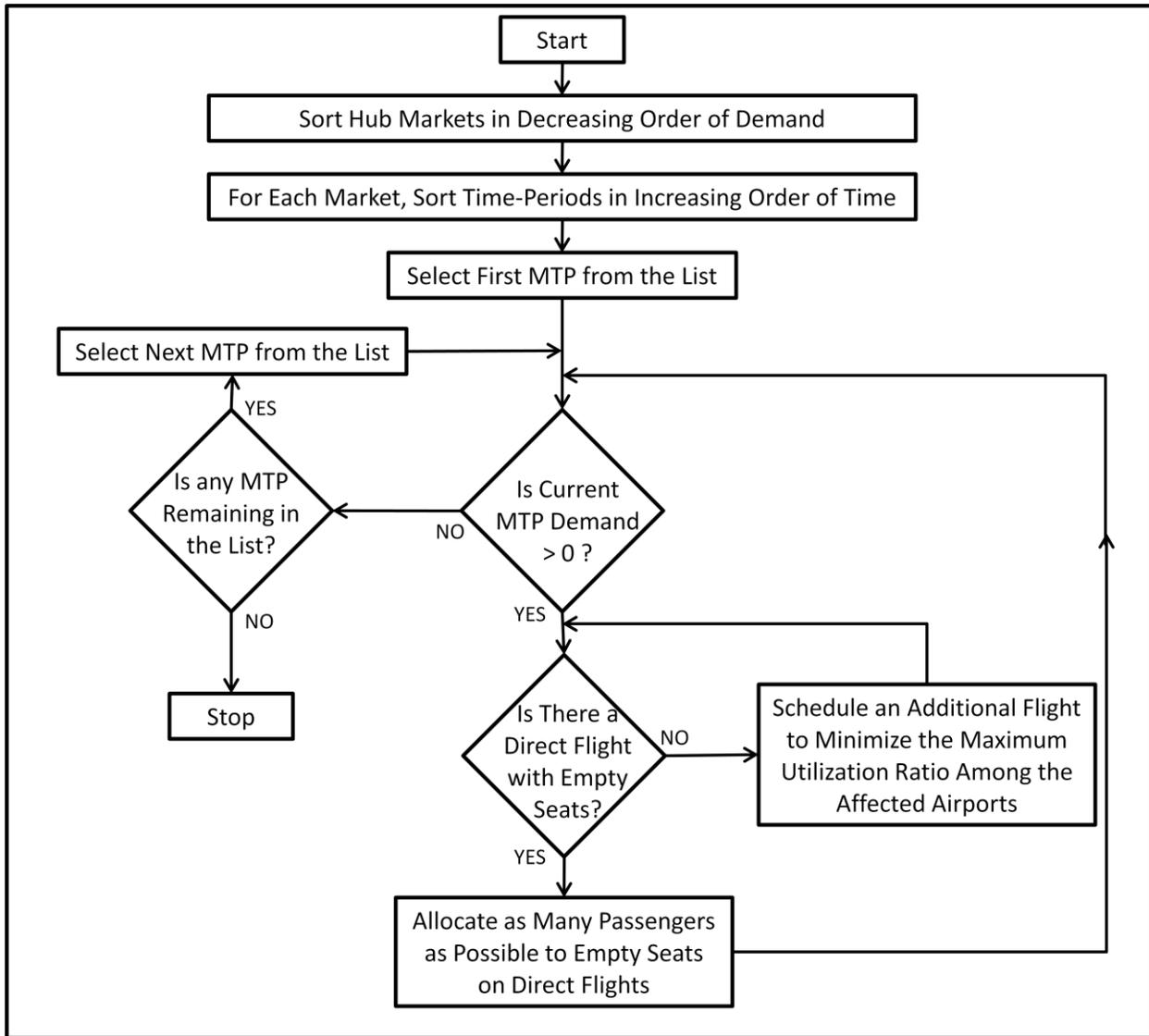


Figure 3: Greedy Heuristic for Hub Markets with Small Demand

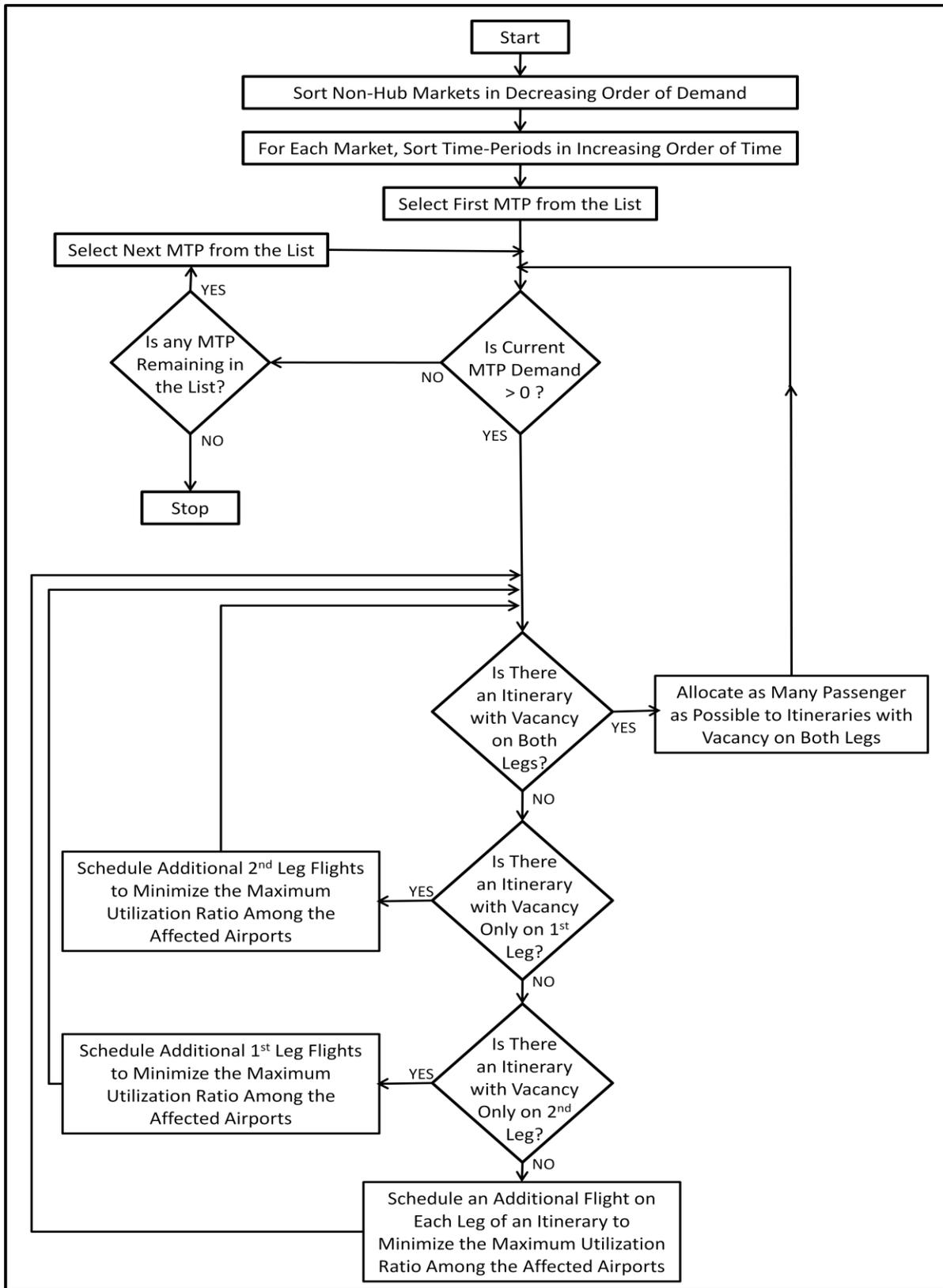


Figure 4: Greedy Heuristic for Non-hub Markets with Small Demand

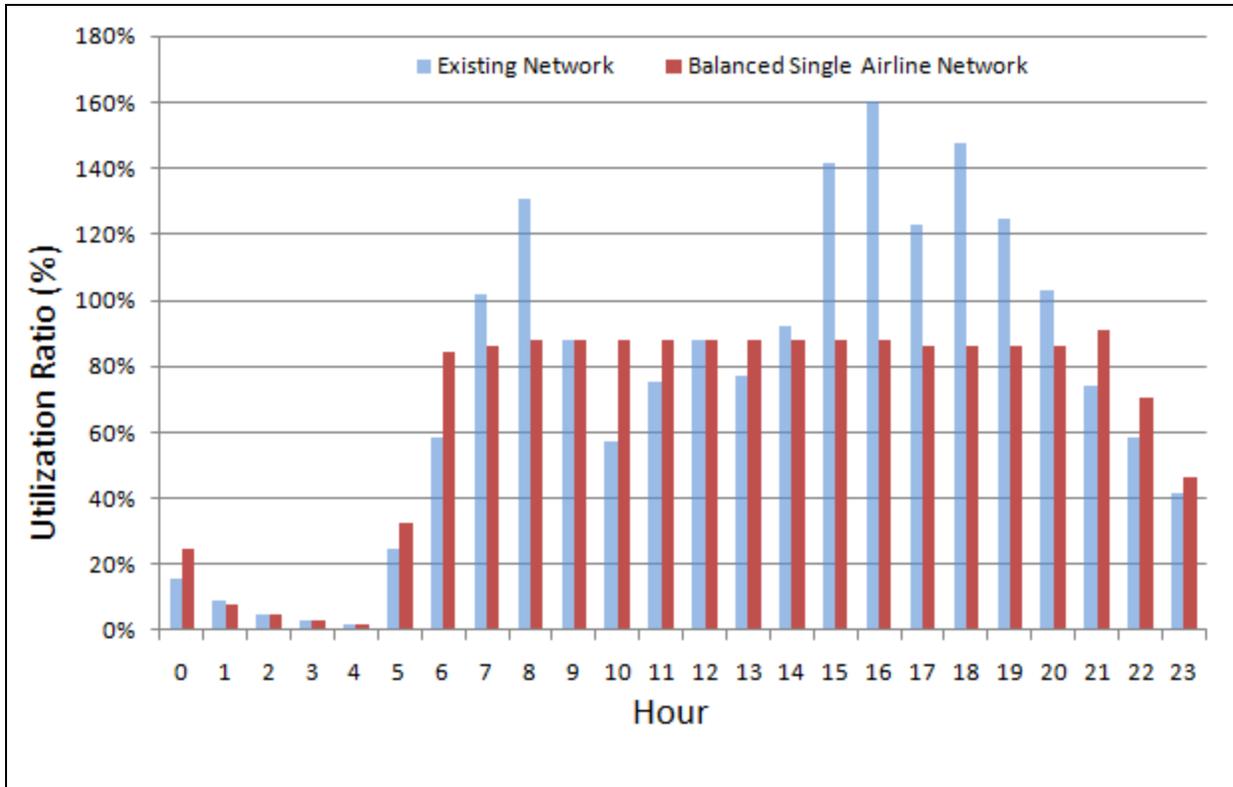


Figure 5: Distribution of Utilization Ratio over a Day at JFK Airport

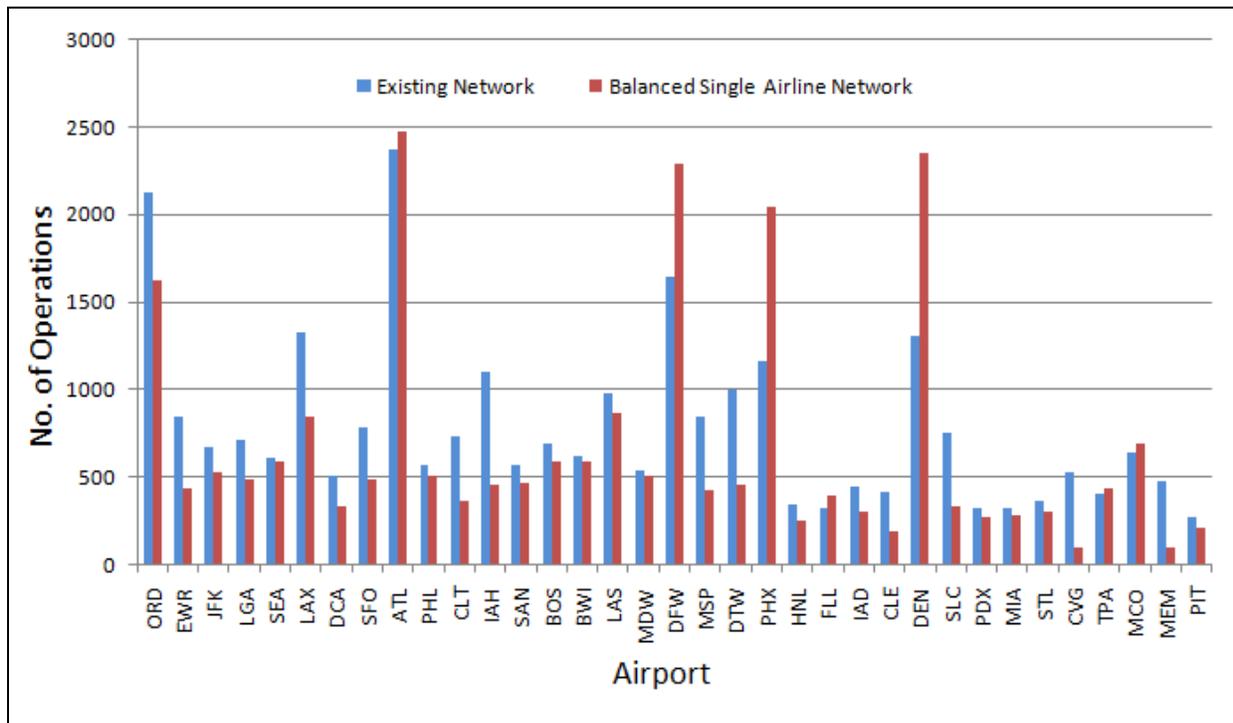


Figure 6: Number of Operations by Airport