The Potential of Demand Management as a Short-Term Means of Relieving Airport Congestion

By Terence P. C. Fan and Amedeo R. Odoni
Massachusetts Institute of Technology

Abstract

Rapid air traffic growth combined with limited airport airside capacities have led to ever-increasing delays. When the demand for air travel and the general economy recover from the current crisis, it is conceivable that air traffic congestion may again come to the fore. Given the rather bleak prospects for significant near-term increases in airport runway system capacity, it is likely that demand for runway access may need to be carefully managed to keep flight delays under control. While various approaches to demand management have been suggested in the research literature, few studies to date have provided quantitative evidence on two major questions: the magnitude of the impact that demand management may have; and the extent to which the current weight-based landing fee systems under-price airside access to busy airports. In this paper, we address in quantitative terms these issues, using as case studies for illustration, New York's LaGuardia, Boston's Logan International and Austin's Bergstrom International airports. We furher propose a framework for developing future demand management policies.

Introduction

Rapid growth in demand combined with limited airport airside capacities have contributed to increasing incidence of delays at many major commercial airports worldwide, at least until September, 2001. Faced with the prospect of limited near-term expansion of runway capacity, policy- makers in both the United States and Europe were considering the need for carefully managing demand for access to busy airports to keep flight delays under control. New York's LaGuardia Airport (LGA), where delays reached nearly catastrophic proportions between September 2000 and January 2001, provides a dramatic case in point. The ongoing work of the authors at LGA and at Boston's Logan International Airport (BOS) will serve as the basis for much of the material and illustrations presented in this paper.

Since 1969, four airports in the United States, LGA and JFK International in New York, O'Hare in Chicago (ORD) and Reagan (National) in Washington (DCA) have been declared "high density rule" (HDR) airports and have been operated with controls that restrict access to slot holders only. Approximately 140 airports outside the United States, including practically all of the busiest ones in Europe, have similarly been designated as "fully coordinated" with access limited to slot holders. The administrative slot allocation practices developed by the International Air Transport Association (IATA) for allocating slots at fully coordinated airports have come under increasing attack in both Europe and the United States, as anti-competitive and running contrary to air transport deregulation and liberalization policies.

In 2000, the Wendell H. Ford Aviation Investment and Reform Act for the Twenty- first Century, or "Air 21", provided for immediate slot exemption at LGA, JFK and ORD for regional aircraft with 70 or fewer seats operating to small communities. Air 21 also calls for the elimination of the slot system at these airports within five years. While requests for new flights under this exemption have so far stayed at a manageable level for JFK and ORD (Moorman, 2000), requests for new flights at LGA at one point totalled to more than 600 additional movements per day, compared to the 1,100 operations a day scheduled up to Summer 2000. In September 2000, LGA had a 15% increase in the number of operations compared to September 1999 (Fiorino, 2001), contributing to a 60% increase in nationwide air traffic control delays, according to the FAA (Woodberry, 2000) and making LGA by far the most congested airport in the US – and probably the world.

In response, the Port Authority of New York and New Jersey (PANYNJ), which manages and operates LGA, announced on September 19 a moratorium on additional flights there, starting in October. In November 2000, the Federal Aviation Administration (FAA) announced that a lottery would be used to reduce the number of flights at LGA, effective on January 31, 2001, until a more permanent method can be established by September 2001 (Kennedy, 2000). Under the FAA plan, an average of 75 commercial airline operations per hour would be allowed at LGA until September 2001. The lottery, combined with administrative prioritizing, was used to determine which of the newly added flights could continue to operate at LGA until then. This, in effect, constitutes a temporary demand management action designed to keep delays to a tolerable level until a more permanent (and less arbitrary than the lottery) demand management system is put in place.

Demand management, in the air transport context, refers to the collection of strategic administrative and economic policies designed to ensure that demand for access to some element of the ATM system is kept at a manageable level. This is distinct from air traffic flow management (ATFM) which operates on a more tactical, day-to-day basis.

While various approaches to demand management have been suggested in the research literature, few studies to date have provided quantitative evidence on two major questions: the magnitude of the impact that demand management may have; and the extent to which the current weight-based landing fee systems under-price airside access to busy airports. In this paper, we address in quantitative terms these issues, using tools drawn from queuing theory. After briefly describing our methodology and tools, we discuss, first, the question of the relative impacts of reductions in the *total* demand at an airport vs. a shifting in the *distribution* of demand by time of day. We then demonstrate, how at congested airports, such as LGA and BOS, the current system of assessing weight-based landing fees *contributes* to congestion by greatly under-pricing access to scarce economic resources – a busy airport's runways. By contrast, these same weight-based landing fees are appropriate for un-congested airports, as a means for cost recovery that takes ability-to-pay in consideration. Overall, the analysis suggests that, *in the short-run, well-designed demand management schemes can be far more effective than any other types of alternatives,* in relieving instances of serious actual or pending congestion. These observations lead to a proposed framework for assessing demand management alternatives and developing future policies in this direction.

Methodology

The focus of the paper is on airside congestion, with an operation's (landing or take-off) delay defined as the waiting time for access to an airport's runway system, absent other potential constraints such as en route or terminal area airspace congestion or "bottlenecks" on the taxiways or aprons. To obtain estimates of such delays, one has the choice between using computer-based, numerical queuing models or a simulation tool. We have chosen the former, as our objective is to conduct a policy-oriented study that needs to examine numerous cases at multiple airports and to obtain estimates of several statistical measures, including estimates of the probability distribution of delays. To this purpose, an advanced dynamic and stochastic queuing model has been used. The model, developed and improved over the years by successive researchers (Koopman, 1972; Kivestu, 1976; Malone, 1995; Malone and Odoni, 2001; Stamatopoulos, 2000) models the dynamic behavior of a queuing system over time by solving numerically and iteratively a large set of first-order differential equations, known as the Chapman-Kolmogorov equations, that describe the system.

In this model, the time-varying demand for runway access, obtained from published flight schedules, is approximated as a non-homogeneous Poisson process, while the time-varying runway capacities are approximated by k-th order Erlang probability density functions (see, e.g., Larson and Odoni, 1981). Starting with initial conditions at time t=0, the model solves the equations describing the evolution of queues and computes the probabilities of having 0, 1, 2, 3, ... aircraft in queue at different time intervals. The model outputs various statistics about the queue, including the average queue length, average waiting time, total delay, fraction of flights delayed by more than a user-specified amount of time, the probability of having flight delays of a specified magnitude at a particular time t, etc. The model runs on PCs and takes less than a second to compute a dynamic 24-hour delay profile at a large airport such as LGA.

Inputs

To apply the numerical queuing model to various airports, two sets of information inputs are needed. The first is the *demand* for runway access as represented by the list of flight operations scheduled by airlines on a representative weekday. For LGA, the schedules for Monday, November 13, 2000, just prior to the moratorium of flight addition, and for August 7 , 2001 (after implementing schedule changes subsequent to the FAA slot lottery), were selected from the *Official Airline Guides* (OAG). For BOS, the typical schedules for the summer of 1998; for Austin (AUS), the schedules for November, 2000 were used. It is assumed for simplicity that all of the scheduled flights will operate, when in fact a small percentage may be cancelled on any given day due to a variety of operational concerns. It is further assumed that the scheduled departure and arrival times will not be affected by congestion delays. While seemingly over-simplifying, this last assumption may still be reasonable, since i) airlines do design some "slack" in their schedules, and ii) significant runway congestion in a given period is likely to increase the probability of congestion in the subsequent period (aircraft arriving late on an inbound flight is likely to face congestion delay on departure on the outbound flight in any case).

The second set of information relates to the *supply* of runway capacity in flight movements per hour for the selected airports. For simplicity, the approximate sustainable capacities under visual flight rules (VFR) are used. Moreover, since there is little public information on cargo, general aviation and other non-scheduled operations at LGA and BOS, the capacities used for these two airports reflect a reasonable estimate of the capacity available for scheduled passenger operations, at 75 and 115 operations per hour respectively. As for AUS, the sustainable VFR capacity is divided proportionally into different categories of flight operations depending on the number of total operations recorded in 1999, and a capacity of 54 operations per hour (54% of the approximate VFR capacity) is used. As we shall demonstrate, the precise airport runway capacity is an important factor in causing delays at congested airports. It is assumed that the mix of departure and arrival operations will not significantly affect the airport runway capacity.

Results - Potential Delay Reduction through Demand Management

In this section, we quantitatively demonstrate the relative impacts on delays of demand management approaches aimed at i) reducing the total demand for runway capacity at a severely congested airport, and ii) shifting the distribution of demand by time of day.

Effect of a small reduction in demand at a congested airport

Figure 1 shows the combined take-off and landing profiles by hour-of-day on a typical weekday at LGA in November, 2000 and in August, 2001. Between these two months, the total number of scheduled airline operations decreased by about 10% (from 1,348/day to 1,205/day) primarily as a result of the lottery results that became effective on January 31, 2001. The peaks and troughs in demand are still visible in both schedules, yet the overall demand has moved closer to the 75-operations-per-hour capacity.

Figure 2 compares average delay per flight, using the flight schedules in November, 2000, and August, 2001. Assuming that no flights are cancelled, each flight scheduled to depart or arrive at the evening peak period between 8 pm and 10 pm in November can expect to be delayed for 1 hour and 20 minutes, even if the VFR capacity is maintained throughout the day. In contrast, flights scheduled to operate during the same period in August, are only delayed for an average of 20 minutes per flight. This represents an 80% reduction in average delays during peak evening hours. Figure 3 shows the total delay in aircraft-hours during the day, suffered by operations scheduled in each hour. Similar to average delays, the total delay during evening peak hours rose beyond 140 aircraft-hours in November, compared with about 25 aircraft-hours in August for the same time period. The area under the August aircraft delays in Figure 3, representing the total delays for the entire day, totaled 210 aircraft-hours, compared with 1,160 aircraft-hours for the November schedule.

While Figures 2 and 3 demonstrate the enormous reduction potential in flight delays when excess demand is reduced to a manageable level, it is equally important for the airport concerned to accurately determine its true capacity in handling aircraft operations. For a congested airport with demand for runway capacity close to or above its supply, the amount of delay is extremely sensitive to the precise capacity number used. Figures 4 and 5 show how the average and total delays in August, 2001 at LGA vary, if the capacity is at 5 operations per hour above and below the 75 operations per hour capacity used in the simulation. For instance, reducing the 75 operations-per-hour capacity by 5 per hour increases the total delays between 6 pm and 9 pm from 73 aircraft-hours to almost 200; increasing the capacity by 5 per hour halves the total delays to about 30 aircraft-hours.

The relationship between delay and the demand-to-capacity ratio can be further illustrated by comparing LGA with BOS and AUS. Figure 6 shows the number of scheduled airline operations per hour as a percent of the respective approximate, sustainable VFR capacity for these airports. Note that while demand is at or above 100% of the VFR capacity at LGA for several periods of time during the day, demand is about equal to capacity at BOS for only the evening peak period. At AUS, the demand barely reaches 40% of its VFR capacity. As a result, the amount of average and total runway congestion delays (in Figures 7 and 8 respectively) at BOS and AUS are noticeably less than in LGA, with BOS showing large average delays only during the evening peak.

Effect of shifting the time-of-day distribution of demand

For LGA, flight delays can conceivably be further reduced by leveling demand peaks, after the overall demand has been reduced to the level experienced in August, 2001. Figure 9 shows what the flight operations profile at LGA would look like in the extreme case in which demand is evenly distributed throughout the period between 7 am and 10 pm (72 to 73 operations/hour during this period). As shown in Figures 10 and 11 respectively, the average and total delays resulting from the de-peaking of flights are reduced by a further 40% during peak evening hours and 20% during the morning peak hours. Compared with the actual August schedules, this reduced the total delays on the typical weekday from 211 aircraft-hours to 168 aircraft-hours, representing a 20% reduction.

Note that even though the scheduled operations for the day never reached or exceeded the capacity of the airport (75 operations/hour), the average and total delays continued to increase until the demand drops abruptly after 10 pm. This is a consequence of the probabilistic effects that the queuing model accounts for and a distinguishing feature from deterministic models.

Results - Pricing of Airport Runway Capacity

Vickrey (1969) noted that optimal use of a congested transportation facility cannot be achieved unless each user pays for the marginal delay costs that (s)he imposes on all other users. From the hour-by-hour delay profiles generated by the queuing model, the marginal delay attributable to an additional flight operation in a given period of time can be estimated. This, combined with an average per-hour cost of operations, can be used to estimate the amount of marginal congestion cost caused by an additional flight. This estimate can then be compared with the current airport charge. A similar procedure was first demonstrated by Carlin and Park (1970).

Figure 12 shows the marginal delay at the three sample airports. The marginal delay graph for LGA's November schedule rises sharply for flight additions after 6 am and then tapers off almost linearly for flights added after 9 am. This pattern is characteristic of a saturated server system, where the servicing of an additional user (flight) would simply impart a delay equal to the service time on all subsequent users. It was this phenomenon that spurred the declaration of a moratorium on additional flights in September, 2000. In comparison, the marginal delay profile for the August schedules at LGA is a lot more restrained, peaking sustainably at around 4 aircraft-hours of additional delay per additional flight operation for most of the day. This almost constant marginal delay increase is representative of an airport with demand extremely close to but not above its capacity for most of the day. For BOS, the marginal delay is just under 2 aircraft-hours during the evening peak period, while it is practically zero throughout the day for AUS.

To arrive at an order-of-magnitude estimate of the marginal cost of congestion, an average per-hour flight operating cost, as reported on the Department of Transportation Form 41, has been used. This reported cost includes the costs of cockpit crew, fuel and oil, direct and contract maintenance, possession and insurance, but does not include the cost of inflight servicing and the cost of lost passenger time. Further, this estimate is taken for the "average" aircraft given the scheduled airline fleet mix.

From the published airline schedules, an average fleet size at LGA is 102 in seating capacity and 52,000 kg in maximum take-off weight, corresponding roughly to a DC-9-30 or a small 737. Using an estimate of $1,600/hour operating cost for an aircraft of this size, the marginal delay costs can be computed from the marginal delay graph in Figure 12. The resulting marginal delay cost curve for the August 2001 schedule shown in Figure 13 can then be compared with the current airport charge for each flight operation (the current charge is levied on each take-off and landing combination). As shown in Figure 13, from 8 am to 8 pm, the marginal delay cost caused by an extra flight operation is 10 to 20 times as large as the average airport charge (landing fee) levied (PANYNJ, 2001). In other words, runway access at LGA is severely under-priced for most of the day and this, in part, contributes to the observed flight delays.

Implications for Demand Management in the U.S.

As illustrated in the preceding section, even a small reduction in or redistribution of the total demand for runway access can lead to significant reductions in flight delays. This has substantial implications for policy directions on demand management in the U.S. In this section, we outline a tentative framework for discussing policy alternatives on demand management at different airports in the U.S., based on i) airports' demand-capacity relationship and ii) the typology of airport users.

Level of total demand versus capacity

Based on the illustration and discussion above, Figure 14 illustrates the notional amount of flight delay reductions that may be achieved through a range of measures at airports with different demand-to-capacity ratios. On the far right is the case where the demand for airport runway access is close to or above the maximum (sustainable VFR) capacity all day, like LGA. For airports in this category, a significant fraction of delay can be eliminated by reducing the *total* demand for the runway access. A mere leveling of the demand at such airports without reducing total demand to a level at or below the sustainable capacity will not be particularly effective. Reducing *total* demand of course requires relatively strong demand management actions, such as the imposition of a flat surcharge on landing fees for the greatest part of the typical day.

Airports like BOS fit in a middle category, where the demand is close to or above maximum capacity for only some periods of the day. In these cases, a mere leveling of peak-period demand will indeed lead to a sizable reduction in delays. Policy actions, such as mild forms of demand management, encouraging the shift of peak-period demand to non-peak periods – without significantly reducing the total demand – may be sufficient. Further, the delay reductions achieved through better management of demand and capacity during or after severe weather (when capacity falls sharply) may be as large and possibly much larger than delay reductions resulting from shifting demand or reducing total demand.

Finally, for airports like AUS with low demand-to-capacity ratios throughout the day, demand management measures are obviously not appropriate. For those cases where the airport demand is close to or above capacity under severe weather, efforts focusing on the efficient use of capacity during short periods of time on an infrequent basis may be most productive in reducing flight delays.

Typology of high-volume users

As mentioned earlier, optimal use of a congested transportation facility is achieved only if each user pays for ("internalize") the marginal costs (s)he imposes on all other users. For airports like LGA and BOS, where there is a large number of high-volume users (or low industry concentration) operating non-homogeneous, point-to-point services, the extent of such internalization is limited, absent a congestion-related charge. Economic demand management measures can be very effective in moving toward a more efficient operating point in such cases.

In contrast, at airports like Chicago O'Hare, Dallas-Fort Worth and Minneapolis/St. Paul, dominated by only one or two high-volume users (airlines with hubs there) operating connecting services, the high-volume users already internalize a substantial portion of the marginal costs. Economic demand management measures may therefore not be as effective in these environments.

This observation, combined with those in the previous section, lead to a tentative airport classification grid for the purpose of designing appropriate demand management measures (see Figure 15). It should be clear at this point that a single set of demand management measures may not suit the needs of all airports.

Concluding Thoughts

Using a numerical queuing model, we provided quantitative evidence on the magnitude of the impact that demand management measures may have. In particular, we demonstrated the relative impact of reducing total demand on flight delays versus shifting the distribution of demand by time of day. We further demonstrated the extent to which the current weight-based landing fee systems under-price airside access to busy airports in the U.S. Based on the empirical findings, we went on to describe a tentative framework that may be used to help evaluate potential demand management measures for different airports in the U.S. Here, we discussed how different demand-to-capacity relationships, as well as the number of high-volume users at an airport, may influence the relative attractiveness of alternative demand management measures. It is clear that any single set of demand management policy initiatives will not work equally well at airports with different characteristics.

References

Carlin, A. and Park, R., 1970. Marginal Cost Pricing of Airport Runway Capacity. *American Economic Review*, 310-319.

Fiorino, F., 2001. LaGuardia Crowned Airport Delay King. *Aviation Week & Space Technology*, 154 (6), 5 February, 49-50.

Kennedy, R., 2000. U.S. Plans Lottery System to Reduce Flights at La Guardia Airport. *New York Times*, 9 November, A19.

Kivestu, P., 1976. *Alternative Methods of Investigating the Time Dependent M/G/k queue*. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Koopman, B., 1972. Air Terminal Queues under Time-Dependent Conditions. *Operations Research*, 20, 1089-1114.

Larson, R. and Odoni, A., 1981. *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, NJ.

Malone, K. 1995. *Dynamic Queuing Systems: Behavior and Approximations for Individual Queues and for Networks*, Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Malone, K. and Odoni, A., 2001. *The Approximate Network Delays Model,* Working Paper, Operations Research Center, MIT.

Moorman, R., 2000. Gridlock at LaGuardia. *Air Transport World*, 37 (8), August, 74-75.

Official Airline Guides, various issues in 2000 and 2001, North American editions.

PANYNJ (Port Authority of New York and New Jersey), 2001. *Schedule of Charges for Air Terminals*, Issued by the Aviation Department, January, 7.

Stamatopoulos, Miltos A., 2000. *A Decision Support System for Airport Planning*, Doctoral dissertation, Athens University of Economis and Business, Athens, Greece.

U.S. House of Representatives, 2000. *Subcommittee on Aviation Hearing on the Slot Lottery at LaGuardia Airport*. 5 December, 1-3.

Vickrey, W., 1969. Congestion Theory and Transport Investment. *American Economic Review*, 59 (2), May, 251-260.

Woodberry, W. Jr., 2000. LaGuardia is Facing Air Gridlock. *New York Times*

Figures referred to in paper

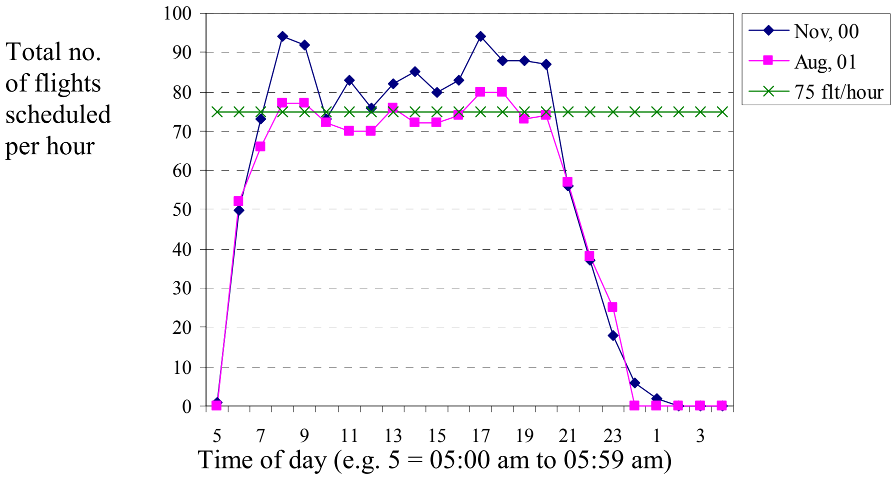

Figure 1. Flight operations at LaGuardia before and after slot lottery

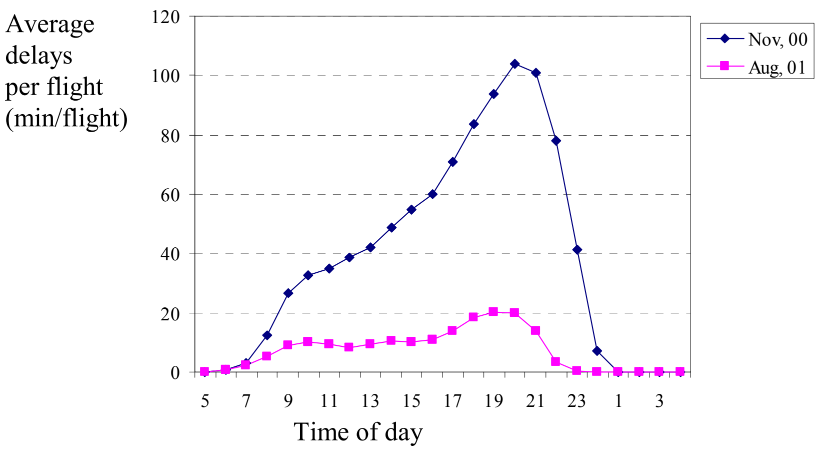

Figure 2. Average flight delays at LaGuardia before and after slot lottery

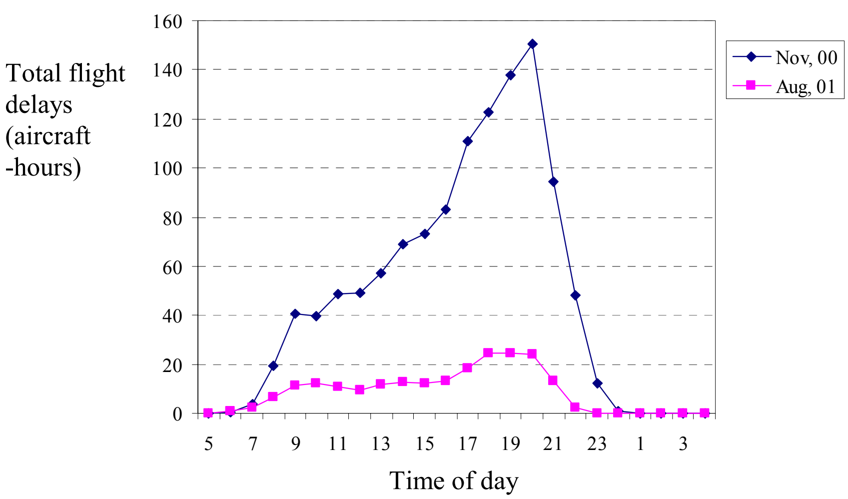

Figure 3. Total flight delays at LaGuardia before and after slot lottery

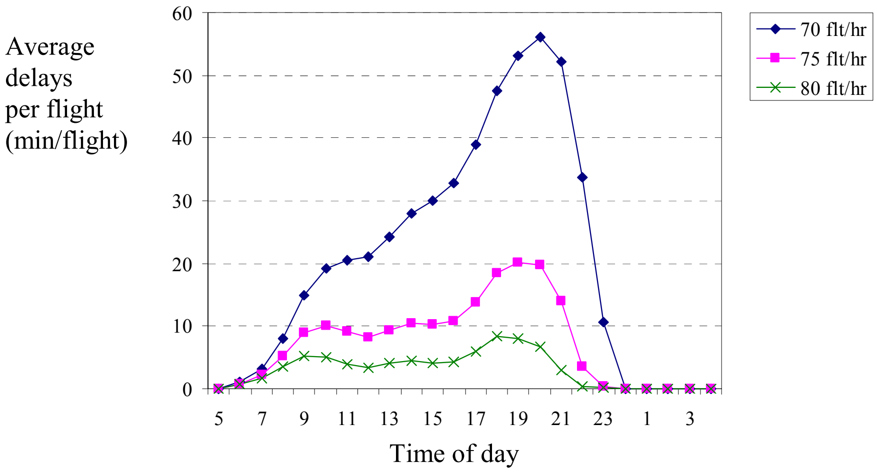

Figure 4. Average flight delays at LaGuardia under different capacities

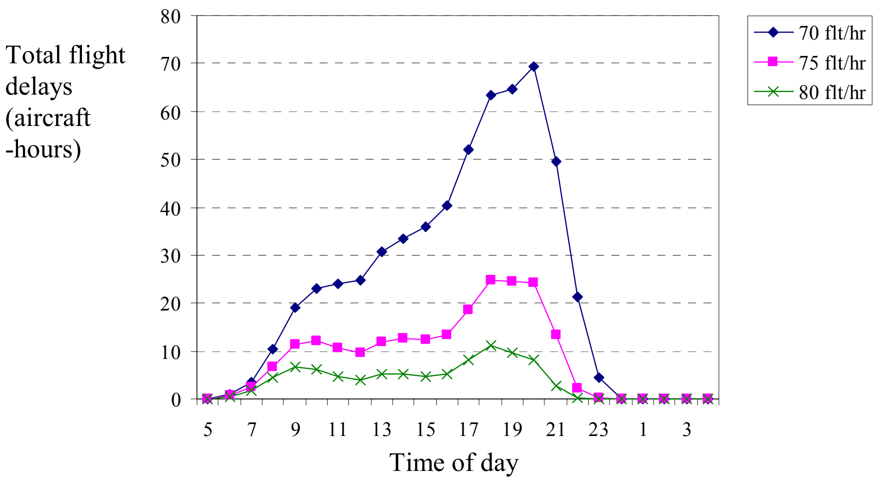

Figure 5. Total flight delays at LaGuardia under different capacities

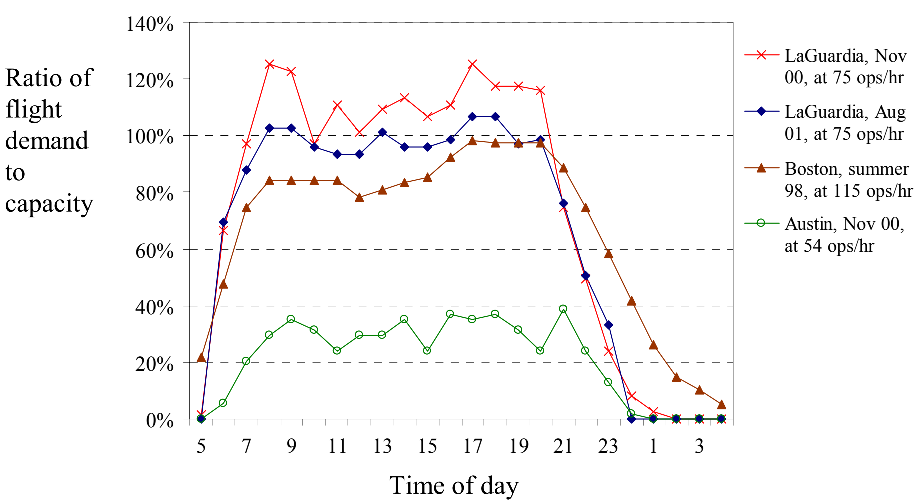

Figure 6. Demand profiles by time of day at sample airports

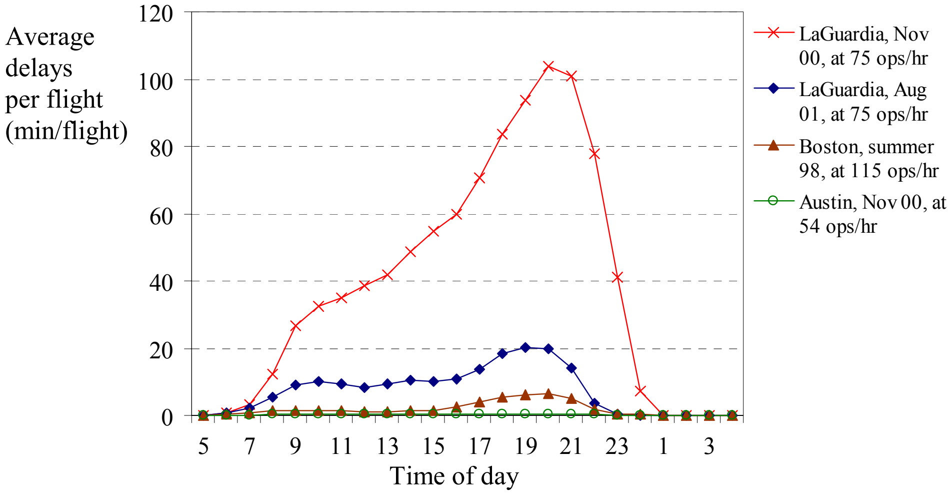

Figure 7. Average flight delays at sample airports

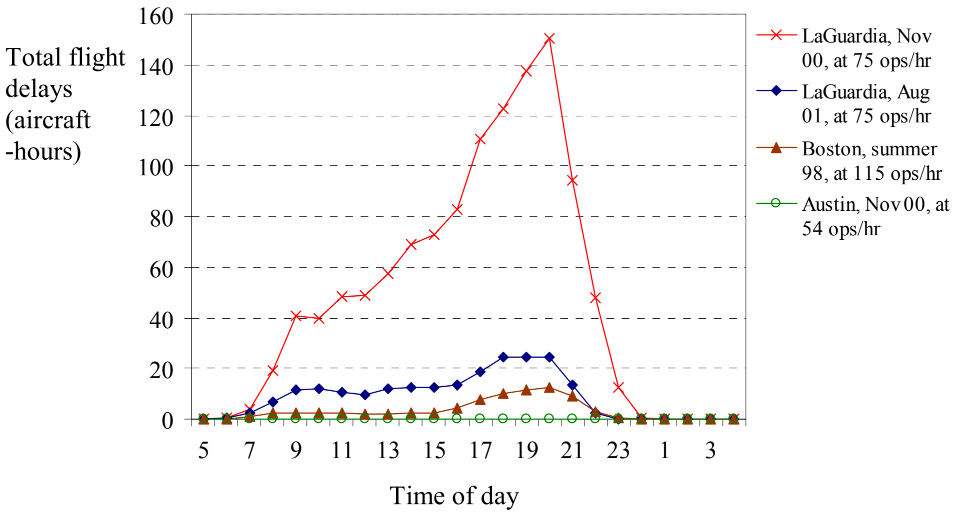

Figure 8. Total flight delays at sample airports

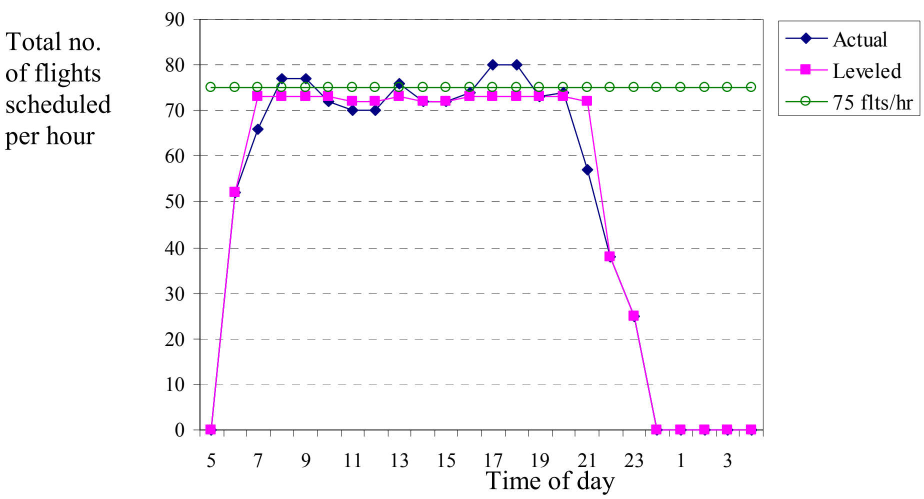

Figure 9. Leveling the hourly distribution of flights at LaGuardia from August schedule

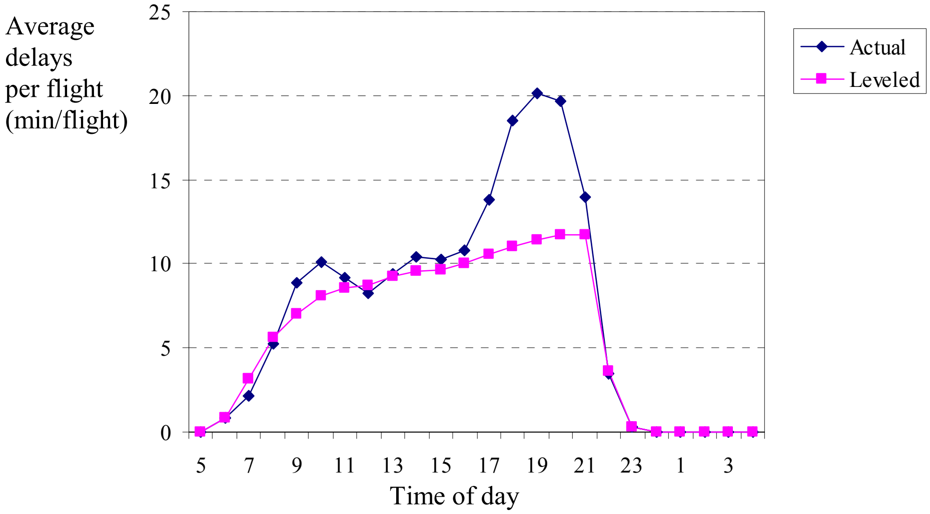

Figure 10. Average flight delays for leveled distribution of flights from August schedule

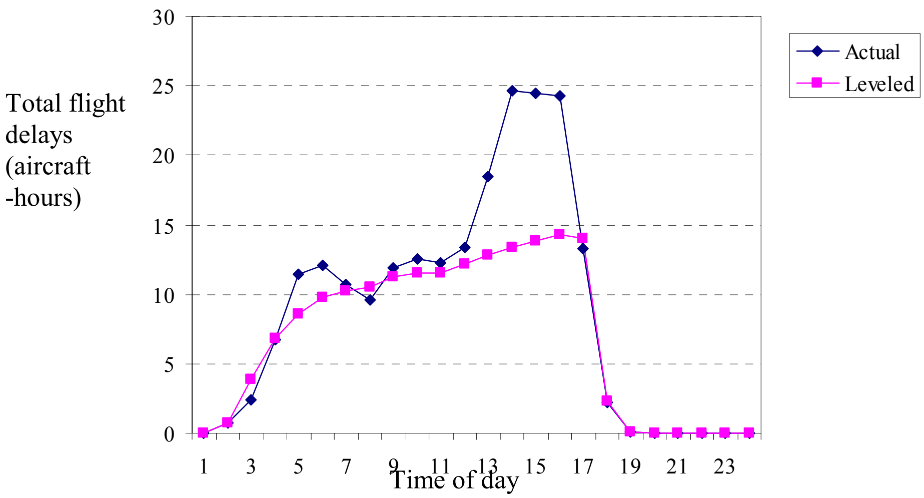

Figure 11. Total flight delays for leveled distribution of flights from August schedule

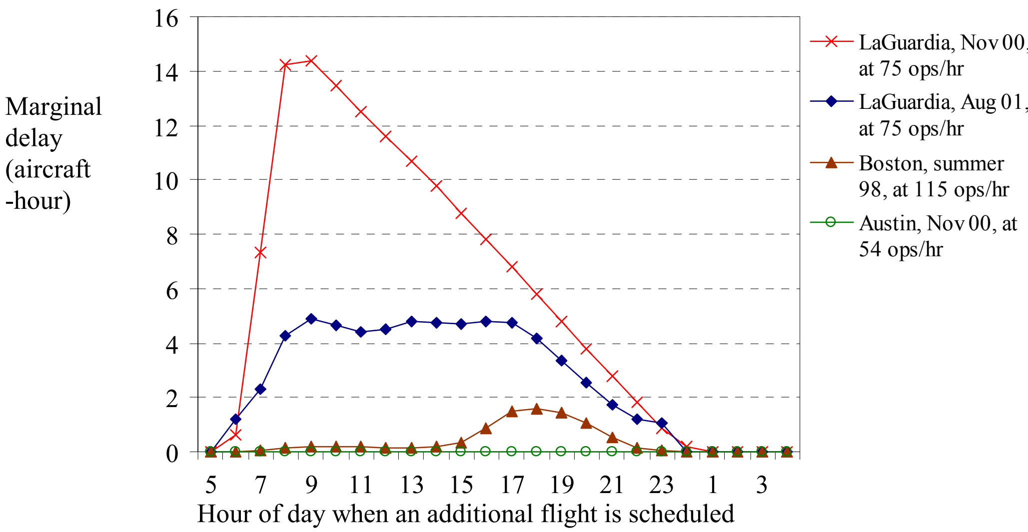

Figure 12. Marginal flight delays (incremental delays from adding one more flight)

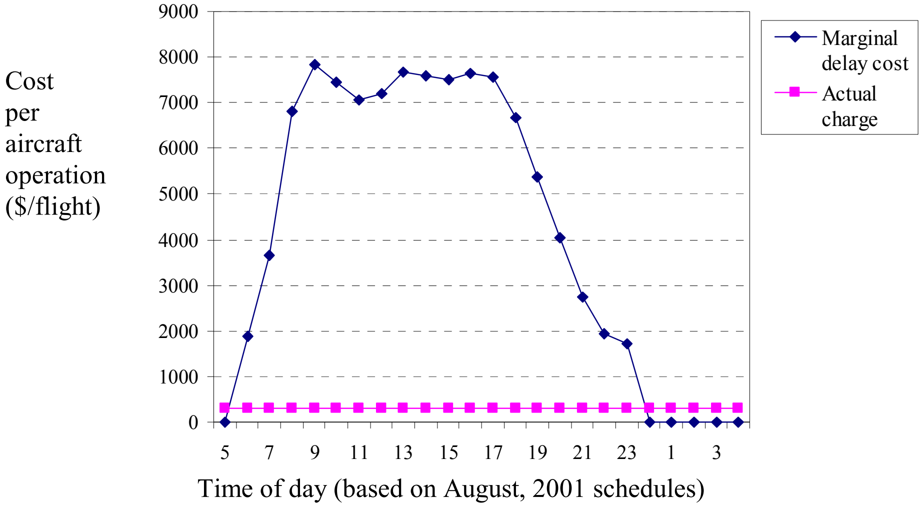

Figure 13. Comparison between marginal delay costs and actual charges at LaGuardia

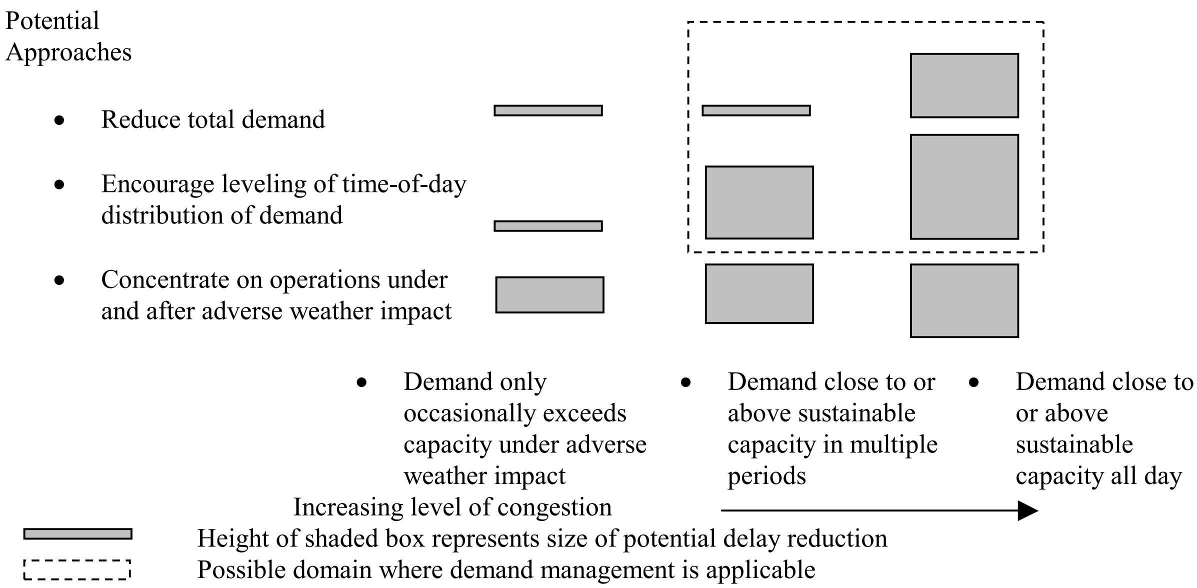

Figure 14. Notional illustration of delay reduction for different airports

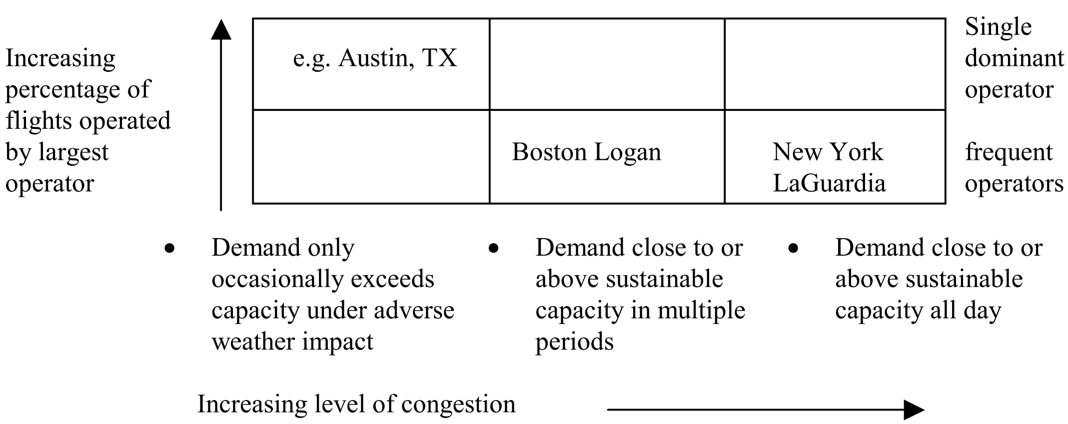

Figure 14. Typology of high-volume users and tentative positions of sample airports

Authors' Biographies

Terence P. C. Fan is currently a doctoral candidate in Transportation at Massachusetts Institute of Technology (MIT), after having worked with transportation and logistics companies in North America and Europe. His research interests include the efficient management of transportation capacity and companies, as well as emerging industry structure in the airline industry. He has won the Best Paper Awards from both the Canadian and U.S. Transportation Research Forum Student Paper Contest. Terence also holds a Master of Science in Aeronautics and Astronautics and another one in Technology and Policy at MIT, as well as a Bachelor of Applied Science in Mechanical Engineering (with Honours) from the University of British Columbia in Canada.

Amedeo R. Odoni is the T. Wilson Professor of Aeronautics and Astronautics and of Civil and Environmental Engineering at the Massachusetts Institute of Technology, and co-director of both the National Center of Excellence in Aviation Operations Research (NEXTOR) as well as the Global Airline Industry Center at MIT. He has received the INFORMS Prize for Lifetime Contributions to Transportation Science and several teaching awards and is the author or co-author of five books and numerous scientific papers.

Electronic mail contact:
Terence Fan: tpfan@mit.edu
Amedeo Odoni: arodoni@mit.edu