

Airport Capacity: Representation, Estimation, Optimization

Eugene P. Gilbo

Abstract—A major goal of air traffic management is to strategically control the flow of traffic so that the demand at an airport meets but does not exceed the operational capacity. This paper considers the major aspects of airport operational capacities relevant to the strategic management of air traffic. A representation of airport capacity that properly reflects an airport's operational limits is discussed. A method is presented for estimating practical airport capacities under various operational conditions. A technique is proposed for optimizing the available airport capacity to best satisfy the expected traffic demand. The optimization is achieved by considering arrival and departure operations as interdependent processes and by strategically allocating the airport capacity between arrivals and departures. The underlying mathematical model is presented, as well as numerical examples illustrating the benefits when solving airport congestion problems.

I. INTRODUCTION

THE restricted capacities of the National Airspace System (NAS) and growing amounts of air traffic increase the potential for congestion both in the air and on the ground, which in turn may substantially increase delays. Problems arise whenever demand exceeds the available capacity at some element in the NAS. In these situations, the role of air traffic management becomes especially significant.

The most important and restrictive NAS component is the airport. The Federal Aviation Administration (FAA) has identified certain major airports as *pacing* airports, so called because the traffic throughput at these airports paces the flow of traffic through the NAS as a whole. A pacing airport is identified by two characteristics: it has a high volume of traffic and the traffic volume frequently exceeds the operational capacity of the airport.

The FAA's Traffic Management Branch closely monitors traffic at the pacing airports and implements strategic programs to manage situations where the demand significantly exceeds the capacity. A strategic program is typically a *ground delay program*, where Controlled Departure Times (CDT's) are assigned to flights departing in the next two to four hours. The CDT's are computed to achieve a prescribed arrival acceptance rate, which reflects the arrival capacity of the overloaded airport.

The accurate and reliable prediction of airport capacity and demand is crucial to the effectiveness of the strategic traffic

management programs. There are existing methods and tools for predicting air traffic demand [15]. However, the problem of predicting airport capacity is less well resolved.

The determination of an airport capacity is complex. Airport capacity depends on many factors, such as meteorological conditions, runway configurations, arrival/departure ratio, and fleet (aircraft type) mix. Furthermore, the *practical capacity* for the purposes of strategic traffic management may be affected by airspace factors (e.g., arrival fix loading, sector loading) as well as human factors (e.g., controller workload).

The vast majority of publications on analysis and optimization of air traffic flow management treats the airport capacities as given, constant parameters (see [1], [2], [17], [18], and [21]). Usually, the airport capacity is defined by two constants: one for arrival capacity and another for departure capacity. The constants can vary for different weather conditions and runway configurations, but they remain constant throughout the time those conditions exist.

The engineered performance standards (EPS) developed by the FAA give more realistic information on airport capacities. The EPS values vary not only by runway configuration and weather, but also by arrival/departure ratio. Three operating conditions are generally given: departure priority (75% or more departures), equal priority (50% arrivals and 50% departures), and arrival priority (75% or more arrivals). For some airports the EPS show only one pair of arrival and departure capacity values for each runway configuration. However, even in the best cases, the EPS data do not cover the entire range of arrival/departure ratios.

The most complete information on airport capacities under various arrival/departure ratios can be represented by a functional relationship between arrival and departure capacities. The character of the relationship was studied extensively (see [3], [10], [11], [13], [14], [16], and [19]).

In one of these studies [19], the analytical model called the FAA Airfield Capacity Model, was developed. This model is capable of determining the relationship between arrival and departure capacities. The MITRE Corporation appears to be the first to apply this relationship in the NASPAC (National Airspace System Performance Analysis Capability) simulation model, where arrival and departure slots can be assigned in response to peak demands [5]. In this paper, a similar representation of airport capacity is used to estimate the capacity and to formulate a new approach to the operational optimization of airport capacity.

The work presented here is being conducted under the FAA's Advanced Traffic Management System (ATMS) pro-

Manuscript received April 1, 1993; revised June 11, 1993. This work was supported by the FAA and was conducted within the ATMS project at the Volpe National Transportation Systems Center.

The author is with Unisys Corporation, 55 Broadway, Cambridge, MA 02142.

IEEE Log Number 9212005.

gram. The major directions of the program have been described in [15]. This paper discusses the estimation and utilization of practical airport capacity as applicable within the scope of the ATMS and as relevant to the strategic management of air traffic.

Section II is devoted to representation and estimation of airport capacity. An empirical approach to estimating airport capacity is taken to obtain practically realizable values that reflect major restrictions to airport traffic throughput for the entire range of arrival/departure ratios.

Section III describes a method for optimization of airport capacity using the derived estimates. The optimization is achieved by dynamic allocation of the capacity over time between arrivals and departures. In general, the optimal solution provides time-varying capacity profiles which most effectively solve a predicted congestion problem by reflecting the dynamics of the traffic demand and the operational conditions at the airport. This approach better utilizes the available resources at the airport to increase the throughput of traffic.

Numerical examples which illustrate the benefits of the described approach are presented in Section IV.

II. REPRESENTATION AND ESTIMATION OF AIRPORT CAPACITY

A. Background

The intensive analytical studies on airport operational capacities began in the late 1950's. Since then, a large number of publications have addressed various aspects of the studies (see [3], [4], [9]–[11], [13], [14], [19], and [21]).

Airport capacity is defined as the maximum number of operations (arrivals and departures) that can be performed during a fixed time interval (e.g., 15 minutes or one hour) at a given airport under given conditions such as runway configuration, and weather conditions. It is calculated as the reciprocal of the mean permissible inter-operation time.

The existing analytical methods (see [10], [11], [16], and [19]) provide the estimation of the mean inter-operation times by taking into account the uncertainty in the time of aircraft appearance at particular points at different stages of arrival and departure, stochastic variability in speed, differences in runway occupancy times, as well as the uncertainty in aircraft fleet mix.

By making assumptions about the distribution functions of the random variables, one can estimate minimum inter-operation times, which provide a given probability of not violating safe separation distance requirements. The minimum inter-operation times are in turn used to calculate the airport capacities. The numerical results substantially depend on the *a priori* suppositions about probability distributions and their parameters. The reliability of the capacity estimates depends on the reliability of the *a priori* information (which is often not very good). A way to get more reliable, realistic estimates is to combine analytical and empirical methods. Empirical data, such as historical counts of arrivals and departures at the airport, makes it possible to correct the analytical models and their parameters.

It has been established that arrival and departure capacities are connected with each other through a convex, nonlin-

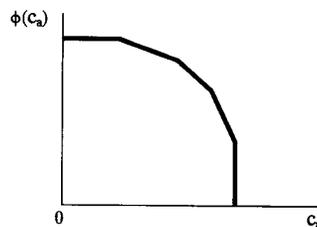


Fig. 1. Airport arrival/departure capacity curve (general view).

ear functional relationship ([16], [19]). The existence of the relationship reflects the fact that the arrival and departure capacities are interdependent. A specific relationship between the arrival capacity c_a and departure capacity $c_d = \phi(c_a)$ depends on various factors such as runway configuration, weather conditions, aircraft fleet mix, runway operating strategy, and characteristics of the air traffic control system.

Geometrically, the relationship can be shown on an arrival capacity/departure capacity plane by a capacity curve, illustrated in Fig. 1, which represents a set of capacity values that reflect the operational capabilities of the airport under certain conditions. To make the relationship specific for an airport requires a complex approach that includes a combination of mathematical modeling using empirical data, and validation of the results using the expertise of practicing traffic managers and controllers.

B. Estimation Method

Assuming the validity of a general convex shape of the capacity curve, an empirical method has been developed to estimate the curve by using real observed data on the number of arrivals and departures at the airport during a fixed time interval over a long period of time. Below, without loss of generality, the number of arrivals and departures per 15-minute intervals (i.e., 15-minute capacities) are considered.

The method is based on the assumption that during a period of time considered, the observed peak arrival and departure counts reflect the airport performance at or near capacity level. Therefore, the curves enveloping the peak data are considered as the airport capacity estimates.

The empirical method is applied to only the pacing airports, which are known to experience severe congestions and substantial delays during peak hours. The existence of significant delays can be considered as an indication that the airports operate close to or at their operational limits. Therefore for these airports, it is reasonable to assume that the historical peak data reflects the maximum operational capabilities and, hence, can be useful for capacity estimation.

The observed data can be organized according to the operational conditions at the airport to provide capacity curves for specific sets of conditions. To date, the observed data has been analyzed for runway configurations and weather. Each major airport has a set of runway configurations that are used with sufficient frequency that empirical data is available to estimate capacity curves for these runway configurations. Weather conditions are clustered into four operational weather

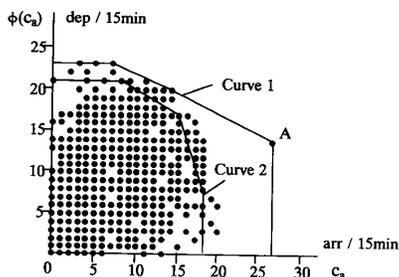


Fig. 2. Historical airport performance data and capacity curves.

categories that reflect conventional limitations on visibility and ceiling: VFR (Visual Flight Rules), MVFR (Marginal VFR), IFR (Instrument Flight Rules), and LIFR (Low IFR). Capacity curves can be estimated for these four different weather categories.

The essence of the method is presented below. Consider the arrival/departure plane shown in Fig. 2. The set of points corresponds to all observed arrivals and departures during 15-minute intervals over a long period of time (e.g., one month or more). The coordinates of each point show the number of arrivals and departures performed at the airport during the same 15-minute interval. The capacity curve is estimated by stretching a piecewise-linear convex curve over the set of points.

Basing capacity estimates on extreme values makes them sensitive to possible outliers in the observed data [12]. Outliers can be of two kinds. They can be caused by errors in the historical data collection process, or they can reflect real but rare events when an airport operates beyond its normal operational limits for a short period of time. In neither case should the outliers be included in the capacity estimation procedure.

The robustness (i.e., nonsensitivity to outliers) of the capacity estimates can be achieved by rejecting some extreme observations [12]. Rejection criteria are selected that reflect confidence levels for the resulting capacity estimates. This approach is illustrated in Fig. 2.

Curve 1 represents a nonrobust estimate that envelopes all observed data and includes absolute maximum values of observed numbers of arrivals and departures. Point A with 27 arrivals and 14 departures per 15 minutes is likely to be an outlier. Curve 1, which includes Point A, seems unrealistic.

Curve 2 in Fig. 2 represents a robust estimate derived from the algorithm that rejects some extreme observations; the rejected data are located outside the area bounded by the curve and the axes of coordinates.

The variety of rejection criteria determines the variety of estimation algorithms. The criteria can be based on various principles: the proximity of extreme observations to the nearest observations, the ranks of extreme values (the rejection criteria are based on order statistics), the frequency of occurrences of extreme observations. The latter is considered here. According to this criterion, the extreme observations that occurred less than a certain number of times during the period of time of interest are to be rejected. The criterion can provide the estimates that are practically insensitive to outliers. If, for

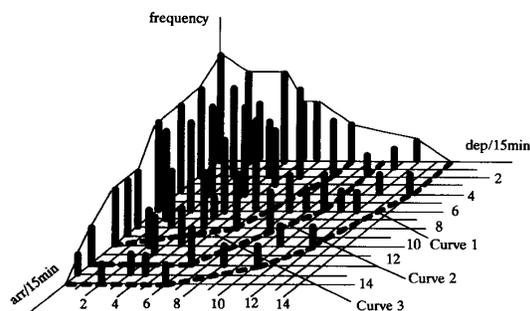


Fig. 3. Histogram of airport arrival/departure operations.

example, the probability for outliers of the same value to occur more than once is negligible small, then the capacity curve which courses through the extremes that occur more than one time is almost unlikely to include outliers.

In the example shown in Fig. 2, outlier A occurred only one time. Therefore, Curve 2 that courses through the data points, which occur at least twice, does not include the outlier.

Fig. 3 illustrates the statistical image of the capacity curve estimates based on the above criterion of rejecting the extreme observations.

The bars in Fig. 3 show frequencies of observed numbers of arrivals and departures during 15-minute intervals throughout the period of time considered. The frequency is determined as the number of occurrences of the same pair of values (arrivals and departures per 15 minutes) divided by the total number of observations. Capacity curve estimates are shown as sample two-dimensional percentiles that course through the extreme observations that occur with the frequency not less than an assigned level. This level reflects the amount of confidence in the capacity estimates and can be heuristically determined. The percentage of total observations enveloped by a curve determines the corresponding percentile represented by the capacity curve.

In Fig. 3, Curve 1 courses the extreme points that occur at least one time; the capacity curve represents the 100th percentile and is not robust. Curve 2 courses through the extreme points that occur not less than two times; the extreme points, which occur only once, have been rejected making the curve robust, insensitive to single outliers. Curve 3 is obtained by the algorithm that rejects the extreme observations that occur less than three times. Curve 3 is more robust than Curve 2.

Using the observed data for 15-minute intervals provides 15-minute capacity estimates that determine the upper limits for the number of arrival and departure operations that can be performed at the airport during a 15-minute interval. The same performance level may not be sustainable for several consecutive 15-minute intervals. The ability to sustain the extreme peak number of operations during a long period of time can significantly depend on the human factor. Empirical data shows that the peak arrivals and departures during 30-minute intervals are usually less than the doubled 15-minute peaks, and 60-minute peaks are less than the doubled 30-minute peaks.

This effect can also be caused by the characteristics of traffic demands when examining longer time intervals. As the time interval becomes longer, it is less likely that the available demand stresses the airport to its operational limits. This could explain some of the drop off in the empirically derived capacities.

Nevertheless, it is useful to estimate 30-minute, 45-minute, and 60-minute capacities in addition to the 15-minute capacities. The same estimation technique is used. The 15-minute observed values, represented as a time series for a long period of time, are recalculated to 30-minute, 45-minute, and 60-minute intervals by summing the 15-minute data within sliding time windows with the width of 2, 3, and 4 consecutive 15-minute intervals, respectively. The resulting capacity curves show an expected quantitative change in the airport capacity when the duration of peak demands increases from one 15-minute interval to two, three, and four consecutive intervals.

C. Some Preliminary Results and Remarks

The preliminary estimates of the capacity curves for major airports based on empirical data from 1989 to 1991 are represented in the Unisys interim memorandums [6]–[8].

As the empirical capacity estimates are the results of statistical procedures, several important questions need to be addressed. Among them are the amount of data needed to ensure the statistical significance of the characterization, the accuracy of the observed data, the stability of the estimates, and the sensitivity of the estimates. Proper consideration of these questions is beyond the scope of this paper. Nevertheless, several comments can be made at this time.

Preliminary analysis has been performed on capacity curves estimated for all pacing airports using tens of thousands of historical observations during the period from 1989 to 1992. This analysis shows good stability of the estimates when comparing curves estimated for each month during the overall period.

The estimated capacity curves were compared with EPS. Although the EPS values do not cover the whole range of arrival/departure ratios and hence, do not form a curve, the rate of proximity of the EPS to the curve can be very informative.

An example of the comparison is shown in Fig. 4, where the estimated curves for runway configuration #1 at San Francisco International Airport (SFO) are presented. The runway configuration includes four runways with two parallel runways (28L and 28R) devoted to arrivals and two parallel runways (01L and 01R) devoted to departures. Both sets of parallel runways cross at approximately midfield. The historical data used for the estimates includes the observations of actual numbers of arrivals and departures for 15-minute intervals during the eight months of August 1990 to March 1991, totaling 6688 pairs of observations. There are three curves shown for different percentages of rejected observations (only extreme observations were rejected): 99.5%, 95%, and 90% curves with 0.5%, 5%, and 10% rejected observations, respectively. The triangle in Fig. 4 shows EPS values for the same runway configuration—52 arrivals and 53 departures per hour. The triangle lies close to the 95% curve.

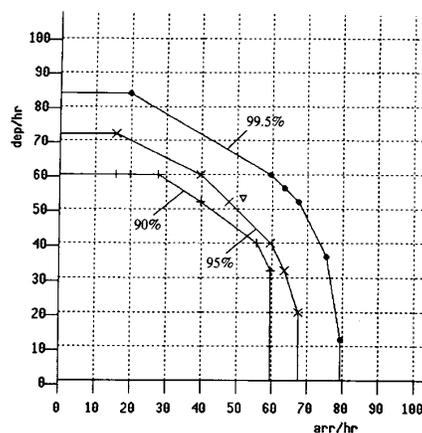


Fig. 4. Estimated capacity curves for SFO, runway configuration #1, VFR.

Similar comparisons performed for other pacing airports have shown that the EPS values are located below capacity curves that envelop 100% of observed performance data, and are typically close to the 90%–95% curves. The correlation with the EPS values supports the supposition that the estimated capacity curves represent operations at or near the practical capacities of the airports.

Further analysis of airport capacity will include the comparison of the empirical capacity curves with those obtained from MITRE's FAA Airfield Capacity Model.

It should be stressed that neither empirical nor analytical models can be expected to provide capacity curves that are completely acceptable for field use by practicing traffic managers. The estimates provided by any method must be subject to expert evaluation and correction by traffic managers and controllers using their experience and knowledge of the specific conditions at the airports. Only after such corrections can the capacity values be applied for solving real air traffic management problems.

III. OPTIMIZATION OF AIRPORT CAPACITY

A. Formulation of the Problem

Once the capacity curves have been estimated, traffic managers have detailed information about airport operational limits for the complete spectrum of arrival/departure ratios for given operational conditions. How should this information be used? Ideally, a manager would select capacity values from the given range to best satisfy the traffic demand. However, it is extremely difficult to find the best solution during a period of severe congestion as the demand profile may vary substantially during that time.

A method for optimization of airport capacity is presented here. The optimization of airport capacity is taken to mean the best allocation of airport capacities between arrivals and departures that optimally satisfy the predicted traffic demand over a period of time under given operational conditions at the airport.

The method is based on a mathematical model of interdependent arrival and departure processes at the airport. The airport arrival and departure capacities are also interdependent; their relationship is determined by a capacity curve. The model treats the airport capacities as the decision variables that are to be determined in accordance with an optimization criterion.

The choice of optimization criterion is an important step in formulating the problem. The effectiveness of arrival and departure operations at the airport can be measured by the total delay time of the flights being served (i.e., the total waiting time in the arrival and departure queues) or by the total number of flights in the queue over the time period of interest. These two measures both reflect the physical essence of the problem and are strongly correlated; larger queues cause longer delays. Which of the measures to use in the optimization criterion depends on factors such as the type of input data available and the simplicity of obtaining the optimal solutions.

In this paper, the total number of flights in the queues has been chosen for the optimization criteria. The major reason is that we consider strategic, not tactical, problems and hence, use the aggregated input data such as total demands for each 15-minute interval (not flight-by-flight data). The total demands can be easily used to calculate the length of the queues, but not the delay time for each individual flight in the queues. In addition, the use of total number in the queues provides less complex algorithms for obtaining optimal solutions. The optimal solution determines the arrival and departure capacity values for each 15-minute time slot to minimize total arrival and departure queues (or functions of the queues). The values can then be used in a flight-by-flight model that determines a schedule for each flight that minimizes the total time in the queues. The optimal capacity values provide the most favorable conditions for obtaining the lowest delays.

The following basic notations are used to formulate the problem:

T	=	a time interval of interest consisting of N time slots of length Δ (e.g., $\Delta = 15$ min); $T = N\Delta$
I	=	$\{1, 2, \dots, N\}$ = a set of time slots
a_i	=	demand for arrivals at the i th time slot
d_i	=	demand for departures at the i th time slot
X_i	=	arrival queue by the beginning of the i th time slot: $i = 1, 2, \dots, N+1$
Y_i	=	departure queue by the beginning of the i th time slot: $i = 1, 2, \dots, N+1$
Φ	=	$\{\phi^{(1)}(u), \phi^{(2)}(u), \dots, \phi^{(s)}(u)\}$ = a set of capacity curves that represent all runway configurations of the airport under all weather conditions
$\phi_i(u)$	=	an arrival/departure capacity curve, which determines the airport capacity at the i th time slot: $\phi_i(u) \in \Phi, i \in I$
u_i	=	airport arrival capacity at the i th time slot, $i \in I$
v_i	=	airport departure capacity at the i th time slot, $i \in I$.

In what follows, X_i and Y_i are state variables, and u_i and v_i are decision variables, $i \in I$. A decision vector $u^* = (u_1, u_2, \dots, u_N, v_1, v_2, \dots, v_N)$ is then introduced.

Consider the problem of managing arrival and departure traffic at an airport during a time interval T . The traffic demand is given by a sequence of arrival and departure demands,

a_i and d_i , respectively, for each time slot of the interval ($i = 1, 2, \dots, N$). According to weather forecasts and other operational conditions for the time interval, a set of runway configurations is assigned. A sequence of capacity curves $\phi_i(u)$ ($i = 1, 2, \dots, N$) is also given. The problem is to find the sequence of arrival and departure capacities (u_i and v_i) that best satisfy the traffic demand.

The general problem of optimization of airport capacity during the time interval T is formulated as follows.

$$\min_{u^*} \sum_{i=1}^N F_i(X_{i+1}, Y_{i+1}) \quad (1)$$

subject to

$$X_{i+1} = \max(0, X_i + a_i - u_i), \quad i \in I \quad (2)$$

$$Y_{i+1} = \max(0, Y_i + d_i - v_i), \quad i \in I \quad (3)$$

$$X_1 = X_0 \geq 0; Y_1 = Y_0 \geq 0, \text{ (given initial conditions)} \quad (4)$$

$$0 \leq v_i \leq \phi_i(u_i), \quad \phi_i(u) \in \Phi, \quad i \in I \quad (5)$$

$$0 \leq u_i \leq B_i, \quad i \in I \quad (6)$$

where $F_i(X, Y)$, $i \in I$ represent given, nondecreasing scalar loss functions that determine the optimality criterion, and B_i represents given maximum values of the arrival capacities that can be utilized during each time slot $i \in I$; X_i , Y_i , u_i , and v_i are integers.

The essence of the problem is well reflected in the following type of loss functions:

$$F_i(X, Y) = \gamma_i[\alpha_i X^k + (1 - \alpha_i)Y^k], \quad k > 0, i \in I. \quad (7)$$

The corresponding optimization problem is

$$\min_{u^*} \sum_{i=1}^N \gamma_i [\alpha_i X_{i+1}^k + (1 - \alpha_i)Y_{i+1}^k], \quad 1 \geq \alpha_i \geq 0 \quad (8)$$

subject to (2)–(6).

This is a problem of minimizing a weighted sum of the k th power of arrival and departure queues for all slots of the time interval T ; for instance, $k = 1$ corresponds to minimizing a sum of weighted queues, and $k = 2$ corresponds to minimizing a sum of weighted squares of queues. The power k can be used as a parameter in the optimization problem. The weight coefficients γ_i and α_i are related to each time slot. α_i determines the priority rate for arrivals at the i th slot, the corresponding priority rate for departures is $(1 - \alpha_i)$. γ_i

determines the relative cost of the i th slot. The coefficients γ_i can be normalized so that

$$\sum_{i=1}^N \gamma_i = 1, \quad \gamma_i \geq 0. \quad (9)$$

The coefficients γ_i can also reflect the confidence in predicting the traffic (arrival and departure demand) and the meteorological conditions. In general, distant time slots (i.e., those far into the future) have a less reliable forecast; hence, smaller values of γ_i can be assigned for those slots.

Equations (2) and (3) represent the airport as a multi-stage control system with initial conditions (4). Inequalities (5) and (6) describe the airport capacity constraints for each time slot.

Equations (2) and (3) describe the dynamics of arrival and departure queues at the airport during the time interval of interest. The number of flights delayed at the beginning of the next time slot depends on the number of delayed flights from the previous time slot and the difference between demand and capacity for the current time slot. If capacities u_i and/or v_i are greater than or equal to the number of aircraft waiting for service at the i th slot, then there is no queue left by the beginning of the next ($i + 1$)th slot (X_{i+1} and/or Y_{i+1} are equal to 0). Otherwise there is a queue (X_{i+1} and/or Y_{i+1} are greater than 0). Equations (2)–(4) guarantee nonnegativity of state variables.

The expressions (1)–(6) constitute a classical optimal control problem. Consider the optimization criterion with the linear loss function

$$\min_{u^*} \sum_{i=1}^N \gamma_i [\alpha_i X_{i+1} + (1 - \alpha_i) Y_{i+1}], \quad 1 \geq \alpha_i \geq 0 \quad (10)$$

that corresponds to (8) with $k = 1$. This minimizes a weighted sum of arrival and departure queues at all slots of the time interval T .

If there is interest only in the results of traffic management at the end of the time interval T , the loss function in (10) is applied to the N th time slot only, and the criterion becomes

$$\min_{u^*} (\alpha X_{N+1} + (1 - \alpha) Y_{N+1}), \quad 1 \geq \alpha \geq 0. \quad (11)$$

Here, a weighted sum of the arrival and departure queues at the end of the time interval considered is minimized. The weight coefficient α determines the rate of priority for the arrival process at an airport. With $\alpha = 1$ only the arrival queue is minimized, and consideration of the departure queue is dropped. $\alpha = 0$ corresponds to minimizing the departure queue only.

The expressions (2)–(6) with criterion (11) formulate an optimal terminal control problem.

B. Linear Programming Model

The optimization problem (10) or (11) subject to (2)–(6) can be reformulated as a linear programming (LP) problem by slightly modifying (2) and (3) and by using a specific property

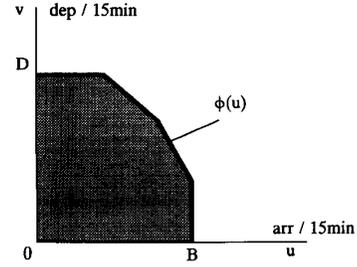


Fig. 5. The area restricted by the capacity curve: $0 \leq v \leq \phi(u)$.

of the nonlinear functions $\phi_i(u)$, namely their convexity and piece-wise linearity.

Let us write the magnitude of the queues and the constraints providing the nonnegativity of the state variables separately using equations (2) and (3). Then instead of equations (2) and (3), the following system of linear equations and inequalities can be written:

$$X_{i+1} = X_i + a_i - u_i, \quad i \in I \quad (12)$$

$$Y_{i+1} = Y_i + d_i - v_i, \quad i \in I \quad (13)$$

$$X_i \geq 0, \quad i = 1, 2, \dots, N + 1 \quad (14)$$

$$Y_i \geq 0, \quad i = 1, 2, \dots, N + 1. \quad (15)$$

Fig. 5 shows an example of the area (the shaded area) that corresponds to one of the constraints (5).

The shaded area can be represented by a system of linear inequalities. Hence, all the constraints (5) can be replaced by a system of linear inequalities. The number of inequalities for each constraint is determined by the number of vertices of the corresponding capacity curve.

Let B_i and D_i denote the maximum values of arrival and departure capacities, respectively, determined by the capacity curve $v = \phi_i(u)$, $i \in I$ (see B and D in Fig. 5). Let n_i denote the number of linear sloping sections of the capacity curve $v = \phi_i(u)$, $i \in I$ (excluding the sections parallel to axes v and u). Then constraints (5) and (6) can be replaced by

$$0 \leq v_i \leq D_i, \quad i \in I \quad (16)$$

$$v_i + g_{ji} u_i \leq b_{ji}, \quad j = 1, \dots, n_i; \quad i \in I \quad (17)$$

$$0 \leq u_i \leq B_i, \quad i \in I \quad (18)$$

where g_{ji} and b_{ji} are the constants that characterize the j th linear section of the i th capacity curve $\phi_i(u)$.

Now the optimization problem (10) subject to (2) - (6) can be easily reduced to a linear programming scheme. After a series of transformations in (12), (13), and (10), and considering (16) - (18), we obtain the following LP problem:

$$\max_{u^*} \sum_{i=1}^N \sum_{p=i}^N \gamma_p [\alpha_p u_i + (1 - \alpha_p) v_i], \quad 1 \geq \alpha_p \geq 0 \quad (19)$$

subject to

$$\sum_{p=1}^i u_p - U_i \leq 0, \quad i \in I \quad (20)$$

$$\sum_{p=1}^i v_p - V_i \leq 0, \quad i \in I \quad (21)$$

$$U_i = X_1 + \sum_{p=1}^i a_p, \quad i \in I \quad (22)$$

$$V_i = Y_1 + \sum_{p=1}^i d_p, \quad i \in I \quad (23)$$

$$X_1 = X^0 \geq 0, \quad Y_1 = Y^0 \geq 0 \quad (24)$$

$$v_i + g_{ji} u_i \leq b_{ji}, \quad j = 1, \dots, n_i; \quad i \in I \quad (25)$$

$$0 \leq v_i \leq D_i, \quad i \in I \quad (26)$$

$$0 \leq u_i \leq B_i, \quad i \in I. \quad (27)$$

The criterion (19) maximizes a weighted sum of arrival and departure capacities during the entire time interval considered. The weight coefficients depend on γ_i (the relative cost of slots) and α_i (priority rates for arrivals at each slot). Having satisfied criterion (19) we automatically satisfy criterion (10) to minimize a weighted sum of arrival and departure queues for all slots of the time interval T . The values of U_i and V_i in (20)–(23) are cumulative arrival and departure demands by the end of the i th time slot, respectively. Inequalities (25)–(27) restrict the area for decision variables below the capacity curve (including the curve).

In the case of equal relative values for all time slots and constant priorities for arrivals and departures during the whole time interval (the coefficients γ_i and α_i are constant: $\gamma_i = \gamma$ and $\alpha_i = \alpha, i \in I$), the objective function (19) is transformed to

TABLE I
PREDICTED DEMAND

TIME	DEMAND	
	arrival	departure
12:00-12:15	$a_1 = 13$	$d_1 = 35$
12:15-12:30	$a_2 = 32$	$d_2 = 2$
12:30-12:45	$a_3 = 24$	$d_3 = 28$
12:45-13:00	$a_4 = 10$	$d_4 = 20$
Total demand	79	85

$$\max_{u^*} \sum_{i=1}^N (N - i + 1) [\alpha u_i + (1 - \alpha) v_i], \quad 1 \geq \alpha \geq 0 \quad (28)$$

which corresponds to

$$\min_{u^*} \sum_{i=1}^N [\alpha X_{i+1} + (1 - \alpha) Y_{i+1}], \quad 1 \geq \alpha \geq 0. \quad (29)$$

The LP version of the optimal terminal control problem (11) subject to (2)–(6) is

$$\max_{u^*} \sum_{i=1}^N (\alpha u_i + (1 - \alpha) v_i), \quad 1 \geq \alpha \geq 0 \quad (30)$$

subject to (20)–(27).

Therefore, to minimize a weighted sum of arrival and departure queues by the end of the time interval of interest one should maximize a weighted sum of the cumulative arrival and departure capacities with the same weights as in (11).

IV. EXAMPLES

Let us consider an airport that, according to the forecast, is going to experience a severe congestion problem during one hour (e.g., from 12:00 to 13:00). The predicted arrival and departure demands for this hour exceed the available capacity and some of the flights have to be delayed. The problem is to find the optimal allocation of airport capacity between arrivals and departures during the hour to best satisfy the predicted demand.

Table I shows predicted demands for each 15-minute slot of the hour. Fig. 6 represents the airport arrival/departure capacity curve that corresponds to the operational conditions at the airport predicted for the hour (weather conditions and runway configuration).

The four dots in Fig. 6 correspond to the demands taken from Table I. The position of the dots beyond the area, restricted by the curve, shows the magnitude of the congestion problem.

To illustrate the scale of the curve, the coordinates of the vertices of the curve (15; 30), (21; 21), and (25; 12) are shown in Fig. 6. According to the curve, the maximum arrival capacity is equal to 25 flights per 15 minutes, and the maximum departure capacity is 30 flights per 15 minutes. Maximum total capacity (arrival plus departure) is 45 flights per 15 minutes. For equal arrival/departure mix operations, the

TABLE II
OPTIMAL ALLOCATION OF ARRIVAL AND DEPARTURE CAPACITIES

TIME	$\alpha=0.5$				$\alpha=0.7$			
	CAPACITY		QUEUE		CAPACITY		QUEUE	
	arrival	departure	arrival	departure	arrival	departure	arrival	departure
12:00-12:15	$u_1 = 13$	$v_1 = 30$	0	5	$u_1 = 13$	$v_1 = 30$	0	5
12:15-12:30	$u_2 = 25$	$v_2 = 7$	7	0	$u_2 = 25$	$v_2 = 7$	7	0
12:30-12:45	$u_3 = 17$	$v_3 = 27$	14	1	$u_3 = 21$	$v_3 = 21$	10	7
12:45-13:00	$u_4 = 21$	$v_4 = 21$	3	0	$u_4 = 22$	$v_4 = 22$	0	5
TOTAL	76	85	24	6	79	80	17	17

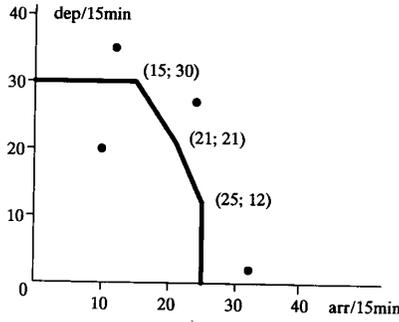


Fig. 6. Airport arrival/departure capacity curve.

airport capacity is 21 flights per 15 minutes for both arrivals and departures.

In this case, the time interval of interest T consists of 4 time slots ($N = 4$). The decision vector u^* contains 8 coordinates: 4 arrival (u_1, u_2, u_3, u_4), and 4 departure (v_1, v_2, v_3, v_4), capacities. Consider criterion (29) to minimize a weighted sum of arrival and departure queues during the entire time interval T . Suppose that there is no queue at the beginning of the time interval, i.e., $X_1 = Y_1 = 0$ in (4) or (24).

In this case, the corresponding LP problem (28) subject to (20) - (27) is formulated as

$$\max_{u^*} \sum_{i=1}^4 (4 - i + 1)(\alpha u_i + (1 - \alpha)v_i), \quad 1 \geq \alpha \geq 0 \quad (31)$$

subject to

$$u_1 \leq 13$$

$$u_1 + u_2 \leq 45$$

$$u_1 + u_2 + u_3 \leq 69$$

$$u_1 + u_2 + u_3 + u_4 \leq 79$$

$$v_1 \leq 35$$

$$v_1 + v_2 \leq 37$$

$$v_1 + v_2 + v_3 \leq 65$$

$$v_1 + v_2 + v_3 + v_4 \leq 85$$

$$0 \leq u_i \leq 25, \quad i = 1, 2, 3, 4$$

$$0 \leq v_i \leq 30, \quad i = 1, 2, 3, 4$$

$$v_i + 1.5u_i \leq 52.5, \quad i = 1, 2, 3, 4$$

$$v_i + 2.25u_i \leq 68.25, \quad i = 1, 2, 3, 4.$$

The optimal capacity values for two levels of priorities for arrivals ($\alpha = 0.5$ and 0.7) are presented in Table II. The table also shows arrival and departure queues at the end of each time slot.

$$u_2 = 25$$

The optimal capacity values vary from one slot to another in response to the dynamics of demand. Table II shows that, in response to increasing the arrival priority rate from 0.5 to 0.7, the arrival and departure capacities are reallocated in two time slots (from 12:30 to 13:00) to decrease the sum of arrival queues from 24 to 17 and increase the sum of departure queues from 6 to 17. The dynamics of the slot-by-slot queue evolution becomes more favorable to arrivals and less favorable to departures. At the end of the time interval T in the case of $\alpha = 0.5$, there is no departure queue and three flights left in the arrival queue. For $\alpha = 0.7$, there is no arrival queue and five flights left in the departure queue. In other words, in the case of $\alpha = 0.5$, the optimal strategy provides a complete solution of the departure problem, and in the case of $\alpha = 0.7$, the optimal strategy provides a complete solution of the arrival problem.

The number of flights delayed under different strategies of airport capacity allocation was calculated. For example, in the case of $\alpha = 0.5$, 14 arrivals and six departures were delayed; in the case of $\alpha = 0.7$, 10 arrivals and 12 departures were delayed. Increasing the arrival priority, the number of delayed arrivals decreases significantly. It can be expected that the

TABLE III
QUEUES FOR OPTIMAL CONSTANT AND VARYING CAPACITIES

TIME	$\alpha=0.5$				$\alpha=0.7$			
	QUEUE for optimal constant capacities		QUEUE for optimal varying capacities		QUEUE for optimal constant capacities		QUEUE for optimal varying capacities	
	arrival	departure	arrival	departure	arrival	departure	arrival	departure
12:00-12:15	0	14	0	5	0	16	0	5
12:15-12:30	11	0	7	0	10	0	7	0
12:30-12:45	14	7	14	1	12	9	10	7
12:45-13:00	3	6	3	0	0	10	0	5
Total	28	27	24	6	22	35	17	17

total time the flights spend in the arrival queue decreases also. A rough calculation of the number of 15 minute time slots the delayed flights were shifted shows that the total number decreases from 24 to 17 slots. At the same time the number of delayed departure flights increases from 6 to 12 flights, and the total time in queue increased from 6 to 17 slots.

Changing parameter α , a traffic manager would be able to generate other strategies for airport capacity allocation and choose any of the alternatives based on other reasons beyond the mathematical formalism.

To estimate the benefits that optimal dynamic (slot-by-slot) capacity allocation provides, the optimal constant capacities (which do not vary during the time interval) and the corresponding queues were determined. The optimization was performed for the same demand and the same criterion (minimum of weighted sum of arrival and departure queues) as above.

The optimization problem is formulated as follows:

$$\min_{u, v} \sum_{i=1}^N [\alpha X_{i+1} + (1 - \alpha) Y_{i+1}], \quad 1 \geq \alpha \geq 0 \quad (32)$$

subject to

$$X_{i+1} = \max(0, X_i + a_i - u), \quad i = 1, 2, \dots, N$$

$$Y_{i+1} = \max(0, Y_i + d_i - v), \quad i = 1, 2, \dots, N$$

$$X_1 = X^0 \geq 0; \quad Y_1 = Y^0 \geq 0$$

$$0 \leq u \leq 25$$

$$0 \leq v \leq 30$$

$$v + 1.5u \leq 52.5$$

$$v + 2.25u \leq 68.25$$

where u and v are constant arrival and departure capacities, respectively; the values of demands a_i and d_i are shown in Table I. In this example, $N = 4$.

The optimal constant capacities have been determined for different values of arrival priority α . Table III shows the arrival and departure queues at the end of each time slot as calculated for constant and optimally varying capacities, respectively, for $\alpha = 0.5$ and 0.7 .

The optimal constant capacities for $\alpha = 0.5$ are identical for arrivals and departures and equal to 21 flights per 15 minutes. By the end of a one-hour interval (by 13:00), the constant capacities produce a total queue of nine flights (three arrival and six departure flights). The varying capacities produce a total queue of three flights, which is substantially less.

For $\alpha = 0.7$, the optimal constant arrival and departure capacities are 22 and 19 flights per 15 minutes, respectively. By the end of the time interval considered, these capacities produce the total queue of 10 flights (0 arrival and 10 departure flights). The optimal varying capacities produce only five flights in queue (0 arrival and five departure flights), which is again significantly less.

Table III also demonstrates that the optimally varying capacities produce lower queues (in comparison with constant capacities) at each time slot within the interval.

The following calculation shows how effective the optimization procedure is in utilizing the airport operational resources. The total original demand for the one-hour interval is 164 flights: 79 arrivals and 85 departures (see Table I). In the case of $\alpha = 0.5$, the total variable optimal capacity for the hour is 161 flights: 76 for arrivals and 85 for departures (see Table II). These capacities are consistent with the demand and altogether provide three flights delayed ($164 - 161 = 3$) to the slots outside of the one-hour interval, three arrivals ($79 - 76 = 3$), and no departures ($85 - 85 = 0$). In the case of $\alpha = 0.7$, the total variable capacity for the hour is 159 flights: 79 for arrivals and 80 for departures. The total number of flights in the queue by the end of the time interval is 5 ($164 - 159 = 5$): 0 arrivals delayed ($79 - 79 = 0$), and 5 departures delayed ($85 - 80 = 5$).

The situation is quite different when the capacities are constant within the time interval and are not coordinated with the dynamics of demand. As was mentioned above, in the case of $\alpha = 0.5$, the optimal constant capacities for arrivals and departures are the same and are equal to 21 flights per 15 minutes. This corresponds to the total hourly capacity of 168 flights: 84 flights per hour for arrivals and 84 flights per hour for departures. In total, these capacities provide nine flights delayed in the slots outside of the one-hour interval: three arrival flights and six departure flights (see Table 3). Though the total constant hourly capacity (168) is greater than the

variable one (161), it provides a greater number of delayed flights at the end of the hour. Moreover, the comparison of total constant arrival capacity (84) with the total hourly demand for arrivals (79) shows that demand is less than the capacity, and we can expect no arrival flights delayed at the end of the hour. However, in reality there are three arrival flights delayed. These apparent paradoxes are caused by operating with the hourly capacities, which are not coordinated with the nonuniform distribution of demand within the hour. The same effect also takes place in the case of $\alpha = 0.7$.

These examples show that the optimal dynamic allocation of airport capacities between arrivals and departures provides a rational utilization of the airport capacities consistent with the dynamics of demand and can be very effective in solving congestion problems at the airport.

Note that in these examples, the demand profile for arrivals and departures has been selected to show significant slot-by-slot variation (peak arrival demands alternating with peak departure demands). In reaction to the variability, the optimization procedure generates capacities that vary from one slot to another. This made it possible to illustrate the benefits that can be obtained through the proper dynamic tradeoff between arrival and departure capacities at the airport. In the case of a small variability in demand during a period of time, the optimization procedure might select constant capacities as the best allocation of airport operational resources for the time interval.

V. CONCLUSIONS

This paper discusses important aspects of airport capacity studies concerning representation, estimation, and optimization within the scope of air traffic management.

Representation of airport capacity through a set of capacity curves that cover the airport operational limits over the entire range of arrival/departure ratios under various conditions has incontestable advantages over representation by fixed constants (one to three constants separately for arrival and departure capacities). However, the benefits of this representation can only be realized under two conditions: the capacity curves must be realistic, and the curves must be properly used to solve major problems of traffic management during congestion.

A method to obtain the realistic estimates of capacity curves has been presented. Using analytical results on the character of the functional relationship between arrival and departure capacities, and using historical data on the actual number of arrivals and departures at an airport during a long period of time, the functional relationship can be made specific for each runway configuration and weather condition. The resulting set of realistic capacity curves represents detailed information on the operational limits of the airport.

A method for optimal allocation of airport capacities between arrivals and departures to best satisfy the traffic demand has been presented. The mathematical model considers arrivals and departures as interdependent processes, treats the airport capacities as decision variables, and selects the optimal capacity values from the area restricted by the capacity curves.

The model can be used as an effective decision support tool for air traffic managers. The output of the optimization

procedure presents an airport capacity profile that suggests to traffic managers how many arrivals and departures would best be performed in each time slot. The capacity optimization model allows a traffic manager to generate effective strategies for managing arrival and departure flows. Alternative capacity profiles, and hence, alternative management strategies, can be obtained by changing the parameters of the model. The manager may then evaluate the alternatives and choose the best solution.

The presented approach can be extended to a network with multi-airport connections. Optimization of capacities for the set of airports could further improve NAS utilization and increase NAS throughput.

VI. ACKNOWLEDGMENTS

The author thanks the guest editor Dr. Stephen Kahne and the referees for their valuable comments and suggestions that improved the text. The improvement would not have been possible without the assistance of Kenneth Howard to whom the author expresses his deep gratitude.

REFERENCES

- [1] G. Andretta, and G. Romanin-Jacur, "Aircraft flow management under congestion," *Transportation Science*, vol. 21, no. 4, pp. 249-253, 1987.
- [2] L. Bianco, and M. Bielli, "Air Traffic Management: Optimization models and algorithms," *J. Advanced Transp.*, vol. 26, no. 2, pp. 131-167, 1992.
- [3] A. Blumstein, "An analytical investigation of airport capacity," Rep. TA-1358-8-1, Cornell Aeronautical Lab., 1960.
- [4] H.S.L. Fan, "Effect of local operational constraints on runway capacity—A case study," *J. Advanced Transp.*, vol. 26, no. 2, pp. 169-184, 1992.
- [5] I. Frolow, and J. H. Sinnott, "National airspace system demand and capacity modeling," *Proc. IEEE*, vol. 77, no. 11, pp. 1618-1624, 1989.
- [6] E. P. Gilbo, "Airport capacity estimation based on observed data," ATMS/ETMS Project Memorandum, Unisys Corp., Cambridge, MA, Mar. 1990.
- [7] E. P. Gilbo, "Arrival-departure capacity estimates for major airports," ATMS/ETMS Project Memorandum, Unisys Corp., Cambridge, MA, Nov. 1990.
- [8] E. P. Gilbo, "Airport capacity estimates for various weather conditions," ATMS/ETMS Project Memorandum, Unisys Corp., Cambridge, MA, July 1991.
- [9] S. G. Hamzawi, "Lack of Airport Capacity: Exploration of Alternative Solutions," *Transp. Research*, vol. 26A, no. 1, pp. 47-58, 1992.
- [10] R. M. Harris, "Models for Runway Capacity Analysis," The MITRE Corp. Techn. Rep. MTR-4102, Rev. 2, 1972.
- [11] S.L.M. Hockaday, and A.K. Kanafani, "Development in Airport Capacity Analysis," *Transp. Research*, vol. 8, pp. 171-180, 1974.
- [12] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [13] M. Janic, and V. Tomic, "Terminal airspace capacity model," *Transp. Research*, vol. 16A, pp. 253-260, 1982.
- [14] L. J. McCabe, "Airport capacity prediction (ACP) system development," Project Memorandum, DOT-TSC-FA058-PM-90-4, TSC, Cambridge, MA, 1989.
- [15] M.F. Medeiros, "Advanced traffic management system automation," *Proc. IEEE*, vol. 77, no. 11, pp. 1643-1652, 1989.
- [16] G. F. Newell, "Airport capacity and delays," *Transportation Science*, vol. 3, no. 3, pp. 201-241, 1979.
- [17] A. R. Odoni, "The Flow Management Problem in Air Traffic Control," in *Flow Control of Congested Networks*, A. R. Odoni, L. Bianco, and G. Szego, Eds. New York: Springer-Verlag, 1987, pp. 269-288.
- [18] O. Richetta and A. R. Odoni, "Solving optimally the static ground—holding policy problem in air traffic control," *Transportation Science*, vol. 27, no. 3, pp. 228-238, 1993.
- [19] W. J. Swedish, "Upgraded FAA airfield capacity model," The MITRE Corp. Tech. Rep. MTR-81W16, vols. 1 and 2, 1981.
- [20] M. Terrab, and A. R. Odoni, "Strategic flow management for air traffic control," *Oper. Res.*, vol. 41, no. 1, pp. 138-152, 1993.
- [21] V. Tomic, and R. Horonjeff, "Effect of multiple path approach procedures on runway landing capacity," *Transp. Res.*, vol. 10, pp. 319-329, 1976.



Eugene P. Gilbo received the M.S. degree in mechanical engineering, and the Ph.D. and Doctor of Sciences degrees in electrical engineering and applied mathematics from Leningrad Polytechnic Institute in 1960, 1965, and 1975, respectively.

His research and teaching career in the USSR was connected with the Department of Mechanics and Control Processes of Leningrad Polytechnic Institute. Since January 1989, he has been with Unisys Corporation, Cambridge, MA. His major research interests include system optimization, robust statistics, and nonlinear statistical signal processing and control. His current research concentrates on the application of optimization and modern statistical techniques to air traffic management and control with the emphasis on airport operations. Dr. Gilbo is the author or coauthor of more than 70 scientific publications including a monograph (with I. B. Chelpanov) *Signal Processing Based on Order Statistics*, Soviet Radio Publishing House, Moscow, 1975 (in Russian).

Dr. Gilbo is a member of the Operations Research Society of America.