S.M. Pollock et al., Eds., *Handbooks in OR & MS, Vol. 6*

Chapter 5

Models in Urban and Air Transportation

Amedeo R. Odoni
Massachusetts Institute of Technology, Room 33-404, Cambridge, MA 02139, U.S.A.

Jean-Marc Rousseau
GIRO Enterprises Inc., 75 Port-Royal Street East, Suite 500, Montreal, Que. H3L 3T1, Canada

Nigel H.M. Wilson
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Introduction

Transportation undoubtedly ranks as one of the areas in which operations research is most intensively applied. The purpose of this chapter is to review briefly operations research models in two major sectors of transportation, urban transportation and aviation. The review is not exhaustive: for urban transportation, it is restricted to planning and operational problems in urban transit authorities while, in the case of aviation, only work related to the infrastructure of the air transportation system (airports and air traffic control) has been covered. These limitations in scope stemmed from the desire to strike a compromise between, on the one hand, providing a review of manageable length and, on the other, presenting our material at a reasonable level of detail. Nonetheless, it is clearly impossible, even under these constraints, to offer a truly comprehensive survey of the extensive literature that has appeared on our two subject areas in the course of the past 40 or so years. Instead, we have tried to offer (the first author on the subject of aviation infrastructure and the second and third on urban transit) an overall perspective on the state-of-the-art and to suggest areas which may require further research.

PART I. AVIATION INFRASTRUCTURE

1. Background

Airports and the air traffic control (ATC) system are the two principal types of infrastructure for the air transportation system. Airport services and facilities are further subdivided into 'airside' and 'landside'. Runways, taxiways, apron areas and hangars are the principal airside facilities and they are collectively referred to as the *airfield*. Landside facilities consist primarily of passenger and cargo terminal buildings, access roads on the airport proper and such supporting facilities as automobile parking areas, power generation stations, etc. Most existing operations research models deal with issues related either to the runway and taxiway systems of airports or to passenger terminals or to ATC. We shall review each one of these three areas of work in Part I.

2. Airport airside analysis

Airfields and, specifically, runway complexes have been for the past thirty years (and continue to be today) the principal bottlenecks of the entire ATC and airport system. The reason, quite simply, is that runways constitute the 'interface' between the three-dimensional airspace and a 'single-file-flow' traffic regime. Moreover, all too often, there are too few of them: runways are expensive to build, 'consume' a great amount of land area and, most important, have significant environmental and other external impacts that necessitate long and complicated review-and-approval processes with uncertain outcomes.

2.1. A model for estimating the capacity of a single runway

It is no accident then that the first important analytical model of airport operations, due to Blumstein [1959, 1960], dealt with the issue of estimating the capacity of a single runway. We shall present this model in considerable detail because it remains valid today, illustrates the effects of the principal factors that affect airport capacity and provides a fundamental building block for many of the computer-based models which are now widely used to estimate capacity and delays at runway complexes.

Consider then a single runway, as shown in Figure 1, which is used for landings only. To land on such a runway, aircraft must descend in single file along the final approach path (Figure 1) where they decelerate and exit onto the taxiway system. Thus, aircraft paths merge in the vicinity of the 'gate' to the final approach – typically 5–8 miles away from the runway threshold. Throughout the final approach, aircraft must, of course, maintain a safe longitudinal distance from each other; moreover, ATC requires 'single occupancy' of runways by landing aircraft, i.e., each aircraft must be safely out of the runway before the next landing can 'touch down'. These safety rules impose limits on the maximum 'acceptance rate' of the runway, i.e., on its 'capacity'.

Let us define the following quantities for an aircraft of type i:

n = the length of the common final approach path.

v_i = ground speed (i.e., 'net' of wind speed) on final approach assuming, as a reasonable approximation, that each aircraft maintains a constant speed throughout the approach.

o_i = runway occupancy time, i.e., the time that elapses from the instant when the aircraft touches down on the runway to the instant when it leaves the runway at one of the runway exits.

Consider then the case in which an aircraft of type i is landing, followed immediately by another aircraft of type j. Denote by s_{ij} the minimum separation required by ATC between the two aircraft, while they are both airborne. Let T_{ij} denote the minimum acceptable (in the sense of not violating any ATC separation requirements) time interval between the successive arrivals at the runway of aircraft i and j. We have [Blumstein, 1959]:

$$T_{ij} = \max\left[\frac{n + s_{ij}}{v_j} - \frac{n}{v_i},\ o_i\right] \quad \text{for } v_i > v_j, \tag{1a}$$

$$T_{ij} = \max\left[\frac{s_{ij}}{v_j},\ o_i\right] \qquad \text{for } v_i \leq v_j. \tag{1b}$$

The situation in which $v_i > v_j$ is known as the 'opening case' because the airborne distance between the two aircraft keeps increasing as they fly along the final approach path on the way to the runway. In this case, the two aircraft are closest to each other at the instant when the first of the two, aircraft i, is at the gate of the final approach path (Figure 1). If at that instant the final approach controller has managed to achieve the minimum allowable separation s_{ij} between the two aircraft, then the time interval between the successive landings of the two airplanes will be given by (1a). By contrast, in the 'closing case' ($v_i \leq v_j$) the two aircraft are closest to each other at the instant when the first aircraft is at the runway threshold; T_{ij} is then given by expression (1b).

Suppose now that the 'long-run' probability of a 'type i aircraft followed by a type j aircraft' is p_{ij}. Then, we have:

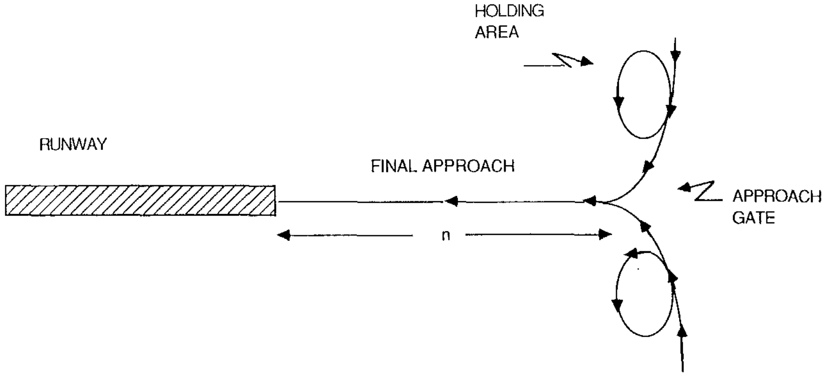


Fig. 1. Idealized representation of final approach and runway.

$$\mathrm{E}\left[T_{ij}\right] = \sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij} \times T_{ij}. \tag{2}$$

Example. A numerical example will be helpful at this point. The FAA subdivides aircraft into three classes with respect to the separations s_{ij} required on final approach: 'heavy' (H), defined as aircraft with a maximum take-off weight (MTOW) of 350 000 lbs. or more; 'large/medium' (L/M) with MTOW between 12 500 and 350 000 lbs.; and 'small' (S) with MTOW of 12 500 or less. Wide-body commercial jets generally belong to the H category; the L/M class includes practically all types of narrow-body commercial jets (L) as well as many of the non-jet airplanes typically used by regional or 'commuter' air carriers (M); finally, most small general aviation aircraft belong to the S class. Because the aircraft in class L/M have widely varying characteristics, it is customary in airport capacity analyses to subdivide this class into two more homogeneous subclasses, L and M. Let us denote the classes H, L, M and S with the indices 1 through 4, respectively.

Assume now that, at a major airport, a particular runway which is used for long periods of time for arrivals only (another runway(s) is presumably used for departures) serves an aircraft population with the following characteristics:

i (a/c type)	p_i (probability)	v_i (knots)	o_i (s)
1(H)	0.2	150	70
2(L)	0.35	130	60
3(M)	0.35	110	55
4(S)	0.1	90	50

1 knot = 1 nautical mile per hour ≈ 1.15 statute miles per hour ≈ 1.8 km per hour.

The probabilities, p_i, indicate the 'traffic mix' at this runway (e.g., 20% of the aircraft are of type H).

In instrument meteorological conditions (IMC) when aircraft are 'spaced' strictly according to the FAA's Instrument Flight Rules (IFR) the separation requirements, s_{ij}, used by ATC at several major airports in the United States are as follows:

$$\begin{array}{ll} & \text{Trailing aircaft} \\ \text{Leading aircraft} & \begin{array}{ccc} \text{H} & \text{L/M} & \text{S} \end{array} \end{array}$$

$$\left[s_{ij}\right] = \begin{array}{c} \text{H} \\ \text{L/M} \\ \text{S} \end{array} \begin{bmatrix} 4 & 5 & 6 \\ 2.5 & 2.5 & 4 \\ 2.5 & 2.5 & 2.5 \end{bmatrix}$$

[Separations specified in nautical miles].

Assuming now that the length, n, of the final approach path is equal to 5 n. miles and applying expressions (1a) and (1b) we find the following matrix, T, of minimum time separations, T_{ij} (in seconds) at the runway:

$$
\begin{array}{c c}
 & \text{Trailing aircaft} \\
\text{Leading aircraft} & \begin{array}{cccc} 1(H) & 2(L) & 3(M) & 4(S) \end{array}
\end{array}
$$

$$T = \begin{array}{c} 1(H) \\ 2(L) \\ 3(M) \\ 4(S) \end{array} \begin{bmatrix} 96 & 157 & 207 & 240 \\ 60 & 69 & 107 & 160 \\ 60 & 69 & 82 & 160 \\ 60 & 69 & 82 & 100 \end{bmatrix}. \qquad (3)$$

Air traffic controllers throughout the world currently use first-come, first-served (FCFS) sequencing of aircraft wishing to land at an airport. [In actual practice, at some busy airports, there are some occasional deviations from strict FCFS, for reasons to be discussed below.] For our example this makes it reasonable to assume that, given that an aircraft of type i has just landed at the runway, the probability that the next aircraft to land is of type j will be simply equal to the proportion p_j of type j aircraft in the aircraft mix. This means that

$$p_{ij} = p_i p_j. \qquad (4)$$

Thus, we obtain the following matrix, P, of aircraft-pair probabilities p_{ij}:

$$
\begin{array}{c c}
 & \text{Trailing aircaft} \\
\text{Leading aircraft} & \begin{array}{cccc} 1(H) & 2(L) & 3(M) & 4(S) \end{array}
\end{array}
$$

$$P = \begin{array}{c} 1(H) \\ 2(L) \\ 3(M) \\ 4(S) \end{array} \begin{bmatrix} 0.04 & 0.07 & 0.07 & 0.02 \\ 0.07 & 0.1225 & 0.1225 & 0.035 \\ 0.07 & 0.1225 & 0.1225 & 0.035 \\ 0.02 & 0.035 & 0.035 & 0.01 \end{bmatrix}.$$

Multiplying the corresponding elements of the matrices T and P to apply (2) we find $E[T_{ij}] = 101.3$ s, i.e., if the ATC system could always achieve the minimum allowable separations between landing aircraft, the runway of this example could serve one arrival every 101.3 s, on average, or up to 35.5 arrivals per hour.

In practice, it is, of course, very difficult to achieve as perfect a level of precision in spacing successive landing aircraft on final approach as implied by the matrix T above. Since spacing in IMC is carried out 'manually' through instructions from air traffic controllers to pilots, it is natural to expect some deviations from the separations indicated by the elements T_{ij} of T. In fact, in view of the expected conservatism of both pilots and controllers, one would expect the separations between given pairs of aircraft types to be, on the average, larger than the corresponding values of T_{ij}. This is indeed the case in practice: for example, in the

United States average spacing under IMC seems to exceed the minimum allowable by ATC rules by about 10–15 s. This effect can be captured by Blumstein's model by modifying the matrix T through the addition of appropriate 'buffers' to each element T_{ij}, designed to account for the additional spacing which is added, intentionally or unintentionally, in practice. For instance, under a particularly simple but reasonable approximation, one could just add the same constant buffer, b, to all T_{ij}, obtaining a new matrix T' of average (not minimum) separations expected between pairs of aircraft, with

$$T'_{ij} = T_{ij} + b. \tag{5}$$

If, to continue our numerical example, we set $b = 10$ s in (5), we shall obviously obtain $\mathrm{E}[T'_{ij}] = 111.3$ s, as the expected separation between successive landing aircraft. This leads to an expected service rate of $C = 32.3$ aircraft per hour, a number which is typical of the service rates that would be observed in today's ATC system in an airport in the United States with a traffic mix similar to that described above. The quantity C is referred to as the 'saturation capacity' or 'maximum throughput' by airport and ATC specialists. (Note that, technically, maximum throughput is truly achieved only when $b = 0$, in which case $\mathrm{E}[T'_{ij}] = 101.3$ s and $C = 35.5$ aircraft per hour for this example.)

Despite its simplicity, Blumstein's model has proved extremely valuable as a planning tool in ATC and airport studies, especially in the context of assessing the effects of various procedural or ATC equipment changes on airport capacity. For example, it was not until recently that the FAA allowed controllers at certain major airports to use 2.5 n. miles separations for the L/M–H, L/M–L/M, S–H, S–L/M and S–S aircraft pairs. Until then the separations required for these aircraft pairs was 3 n. miles – and this is still the case at most airports. This change from 3 to 2.5 n. miles provides an increase of approximately 2.7 arrivals per hour (from $C = 29.6$ to $C = 32.3$) or 9.1% in the runway (saturation) capacity. Similarly, ATC controllers attempt to achieve more uniform final approach speeds at congested commercial airports, typically by recommending that pilots land the smaller and slower aircraft at speeds more similar to those of some of the commercial jets. For instance, if in the numerical example above, ATC could somehow achieve $v_3 = 130$ knots and $v_4 = 110$ knots by assigning higher final approach speeds to 'medium' and 'small' aircraft, we would obtain $C = 36.8$ arrivals per hour, a 13.9% increase over $C = 32.3$.

Other promising possibilities for increasing runway capacity also emerge from the study of the Blumstein model and of realistic examples of its application, such as the one above. For instance, inspection of the matrix T in (3) indicates that certain aircraft sequences may be more desirable than others (e.g., the sequence 1–4, or 'H–S', requires at least 4 min of separation between successive landings, while the sequence 4–1, or 'S–H', requires only a 1-min separation. This suggests the possibility of computer-aided sequencing of aircraft waiting to land at an airport, an idea that has been investigated in detail by several researchers [Dear, 1976; Psaraftis, 1980; Venkatakrishnan, Barnett & Odoni, 1993]. The 'runway sequencing problem' is typically formulated as a constrained optimization problem with the objective of finding sequences which maximize the service rate of runways (or some other aggregate measure of performance) while maintaining a certain level of 'fairness', so that some types of aircraft are not unduly penalized by being always relegated to the end of the queue of arrivals waiting to land. This approach is now reaching the stage of initial implementation through innovative terminal area ATC systems currently being installed in Germany, in France and in the United States. Note that, when sequences are in use which are not based on a FCFS discipline, expression (4) is no longer valid and must be replaced by an expression (or an algorithm for computing the probabilities p_{ij}) which reflects the sequencing scheme actually in use.

A number of improvements and extensions of the Blumstein model of a single runway used for arrivals only were proposed soon by early researchers. It is obvious, for example, that some of the parameters which are treated as constants in the Blumstein model (e.g., the approach speeds, v_i, and the runway occupancy times, o_i, of each type of aircraft) can be treated more realistically as random variables with associated probability distributions [Odoni, 1972]. The distances between successive aircraft on final approach can also be treated as random variables whose probability distribution depends on the required ATC separations and on the characteristics and performance of the terminal area ATC system, including the controllers and pilots [Harris, 1972].

Entirely analogous concepts can also be applied for estimating the capacity of runways used only for takeoffs, as well as of runways used for both landings and takeoffs. In the latter case, it is important to identify the strategy used by ATC controllers to sequence landings and takeoffs on the runway. Under the strategy most commonly used, controllers during peak demand periods may serve a long 'string' of successive arrivals (e.g., 5 to 10 arrivals in a row), then a long string of successive departures, then another string of arrivals, and so on. In this case the runway capacity can be approximated as a simple weighted average of C_a, the runway capacity when the runway is used only for arrivals, and of C_d, the runway capacity when the runway is used only for departures, the weights being equal to the fractions of time when the runway is used for serving arrivals and departures, respectively [Odoni, 1972]. However, at busy airports in the United States, controllers often use a strategy whereby arrivals alternate with departures on the runway, i.e., the separations on final approach between successive arriving aircraft are 'stretched', so that a departure can take place during the time interval between the two arrivals. This is a procedure that requires considerable skill but, if performed accurately, can increase considerably the total operations (landings and takeoffs) capacity of the runway. A model of this operating strategy was developed by Hockaday & Kanafani [1972] and was subsequently generalized by several researchers [see, e.g., Swedish, 1981].

2.2. More complex runway configurations

Many major airports operate with more than a single runway: some use two, three or four runways simultaneously, while the two busiest airports in the world, Chicago's O'Hare (ORD) and Dallas/Ft. Worth (DFW) often operate with six simultaneously active runways. In fact, the number and identity of runways in use at any given time, as well as the allocation of types of aircraft and mix of operations to them, may change several times a day at many of these large airports. The selection of the specific runways to be operated at any one time depends on many factors: demand (e.g., during periods of low demand an airport may accommodate all its traffic on a single runway, even though more than one runways may actually be available); weather conditions, including visibility, precipitation and wind speed and direction; mix of traffic (e.g., during peak periods for flight arrivals, one or more runways may be dedicated to serving exclusively arrivals – and conversely for peak departure periods); and noise regulations which, for example, may prohibit or discourage the use of certain runways during the night or during certain parts of the year. For an airport with several runways, there can be a large number of combinations of weather (ranging from 'VFR-1' – good weather conditions – to 'IFR-3' – very limited visibility all the way down

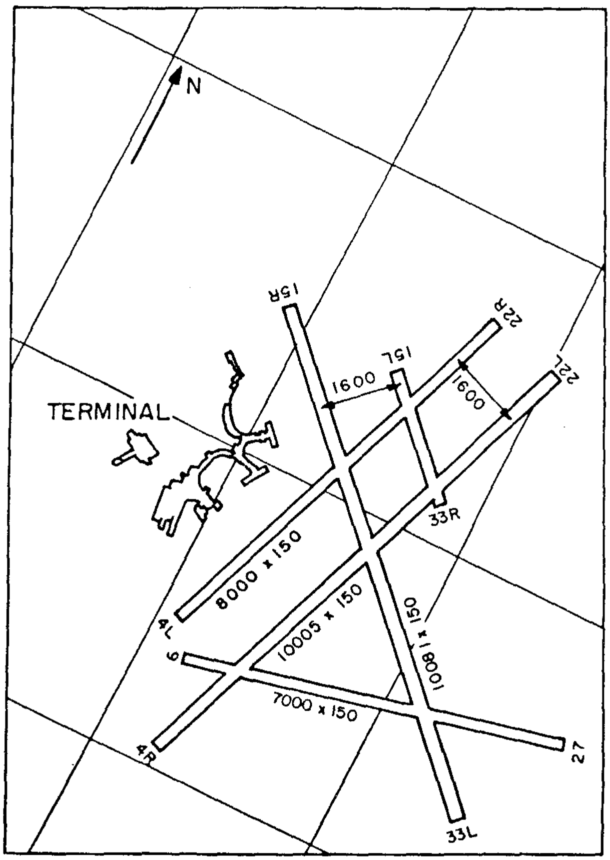


Fig. 2. Runway layout plan, Boston Logan International Airport.

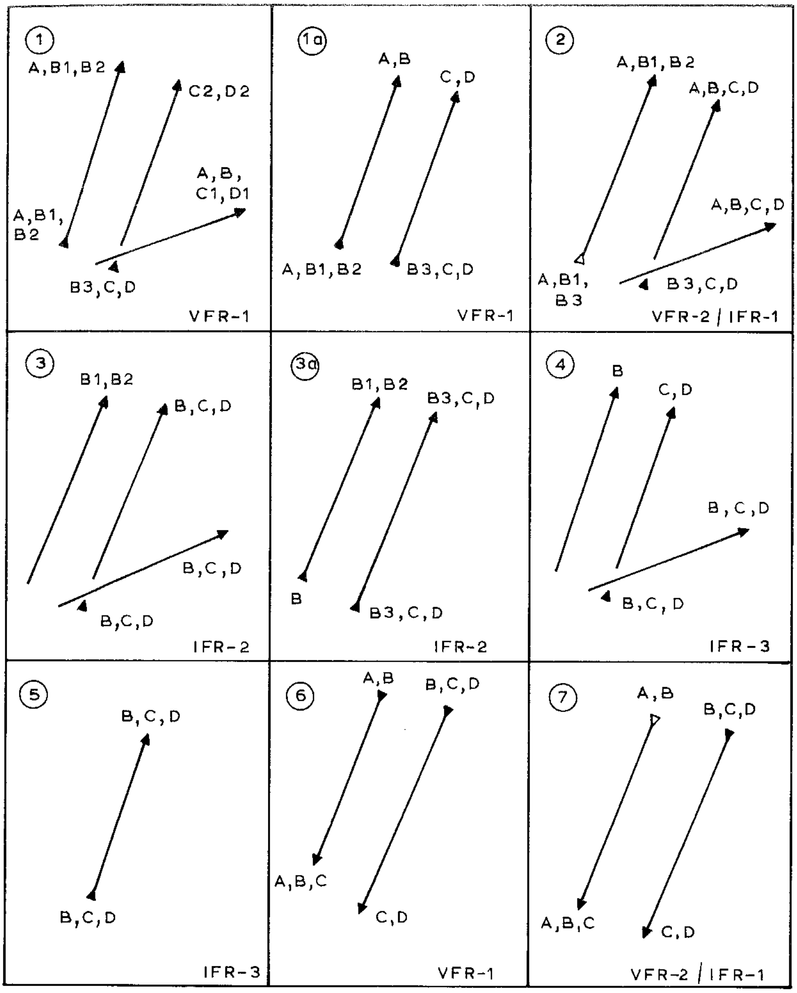


Fig. 3. Examples of runway configurations at Boston Logan under various weather conditions (VFR-1, VFR-2/IFR-1, etc.). Inbound arrows indicate direction of landings, outbound of take-offs. Aircraft types (A, B, B1, B2, ...) assigned to each runway are also indicated.

to the runway surface), active runways and allocation of traffic and operations to the runways. Each of these combinations is called a 'configuration'. For example, Boston's Logan International Airport (BOS) with five runways (Figure 2) can operate under 39 different configurations (Figure 3)!

For computing airport capacity and improving airport design and operations in such more complex cases, the use of OR approaches is a necessity. The specific tools that are used in practice can be analytical models or simulations. To start with, it is rather straightforward to extend the type of analytical model described in the previous section to airport configurations involving two simultaneously

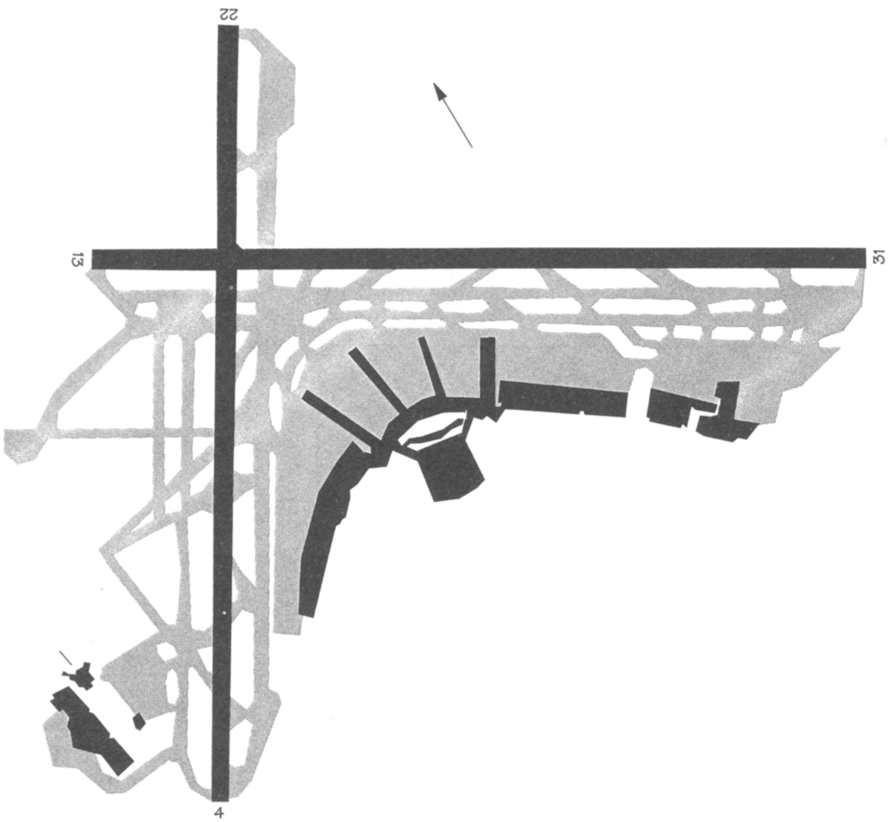


Fig. 4. LaGuardia Airport, New York.

active runways. Given a set of priority rules for sequencing operations on the two active runways, one can develop the usual time separation matrix for successive operations and compute quite accurately the available capacity. For example, New York's LaGuardia Airport (LGA) is usually operated with two intersecting runways, one for arrivals and the other for departures (Figure 4). Air traffic controllers will typically alternate arrivals and departures in this case (for instance, an arrival on runway 31, then a departure on 4, then an arrival on 31, etc.). Given the location of the runway intersection, it is easy to compute LGA's capacity using a Blumstein-like methodology. Note also that, depending on wind directions, the assignment of landings and takeoffs to runways at LGA may change, giving rise to additional configurations. Because of the change of the location of the runway intersection relative to the points where takeoffs are initiated or where landing aircraft touch down, the capacities of these various configurations at LGA (all involving two active runways) may not be equal. Similarly, in the case of two parallel runways the degree of interaction between the runways and the capacity of the airport depends on the distance between the runways.

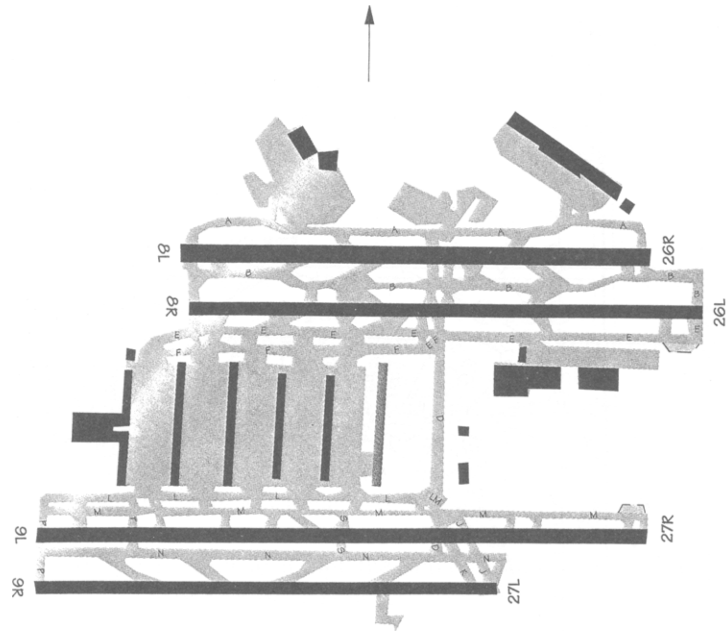


Fig. 5. Atlanta Hartsfield International Airport.

The application of analytical models to configurations involving three or more simultaneously active runways is more problematic, because of the large number of possible interactions among operations taking place at different runways and the complex operation-sequencing strategies that can be used. The typical analytical modeling approach taken in such cases is to approximate the representation of the airport's operation by 'decomposing' the airport into more easily analyzable components. For example, Atlanta's airport, one of the world's busiest, usually operates with two close pairs of parallel runways (Figure 5). One runway in each pair is used for arrivals and the other for departures. Since the two pairs of runways are at a sufficient distance from each other to be operated more or less independently for ATC purposes, one can approximate the capacity of the airfield by adding together the capacities of each individual close pair of runways. (Note that the capacities of the two individual pairs may not necessarily be equal – for instance, if the mix of traffic served by the 'North' pair happens to be different from that served by the 'South' pair.) The accuracy of such a decomposition approach depends, of course, on the geometry of the airfield, i.e., on how 'decomposable' the runways' layout is in the first place.

2.3. Simulation models

A second approach is the obvious one of simulating the airfield. Beginning in the early 1970's, a large number of general-purpose simulation packages have

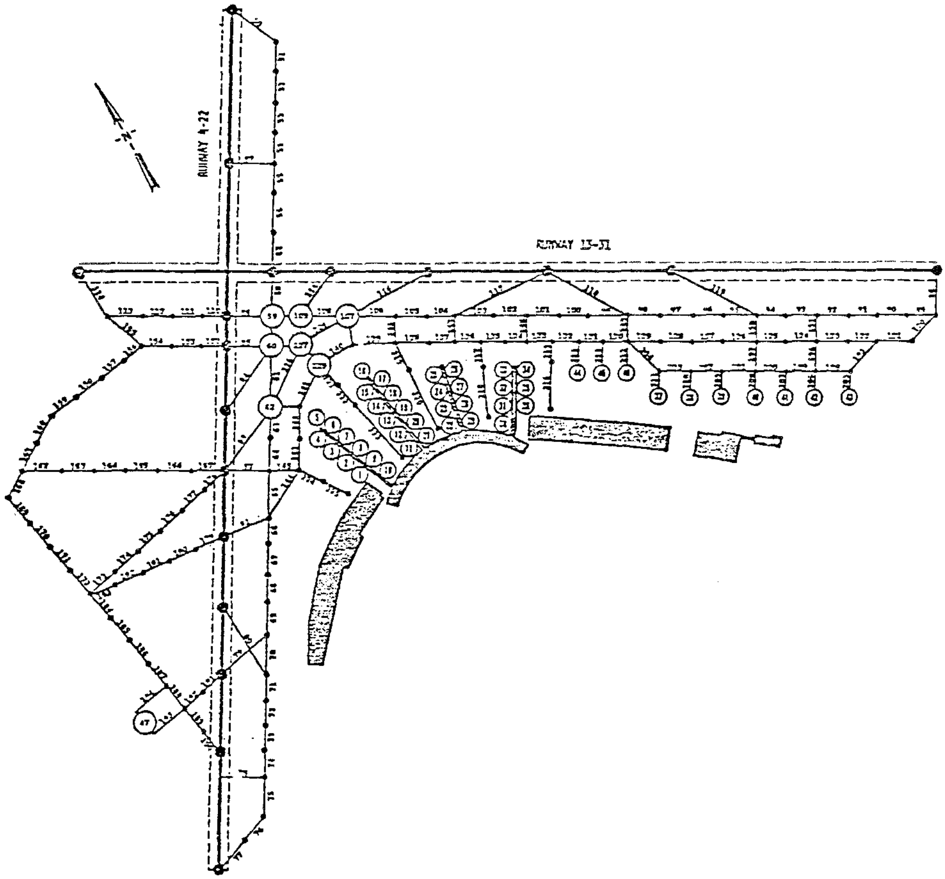


Fig. 6. Network representation of LaGuardia Airport. Line segments indicate network 'links', each of which can hold one aircraft at a time. Small circles indicate apron gates, larger ones key intersections on the airfield.

been developed for application to the analysis of airport airside operations, often covering not only runways but also aircraft movements on taxiways and aprons [for a review, see Odoni, 1991]. Some of these simulation packages are publicly available, while others are proprietary. Practically all of them represent the airfield as a network of nodes and links (see Figure 6 for an example). Aircraft move on this network along prescribed paths that consist of strings of nodes and links. Typically each link can be occupied by a single aircraft at a time. Thus a delay occurs whenever an aircraft attempts to use a link which is already occupied by another aircraft. Whenever two or more aircraft attempt to occupy a free link at the same time (e.g., two aircraft approach a taxiway intersection from different directions) the logic of the model resolves the conflict according to ATC priorities and assigns the free link to one of the candidate aircraft.

The network representation has both advantages and disadvantages. On the positive side, the network structure is intuitively appealing, can be used to develop a highly detailed representation of the airfield and provides a convenient base for collecting and reporting occupancy and delay statistics. On the negative side, the network structure can impose high set-up costs, reduce flexibility and slow down program execution. For example, the process of identifying the paths to be traversed by aircraft on the airport's surface is far from trivial and has major implications for the functionality of the simulation. In this respect, until recently, most available simulation packages required among their inputs a listing of assigned paths connecting each of the apron 'gates' (aircraft parking positions) with each runway end. For a large airport with 5 runways (i.e., 10 runway ends) and 100 gates, this meant listing 1000 different paths, each consisting typically of 30 or more links – or a total of 30 000 or more input numbers. Interestingly, the more sophisticated solutions to this particular problem that have been adopted by some modelers in the last few years, often have serious drawbacks as well. To avoid listing all paths, variations of shortest path algorithms have now been incorporated into some models: for each configuration of the airport they compute the paths that aircraft would travel between a gate and each active runway's end. (The 'variations' are needed to adopt the standard shortest path algorithms to the peculiarities of airfield traffic.) However, through this approach, the model still 'knows' of only a single best path between a gate and a runway. As a result, flexibility and realism suffer: the model does not react dynamically to operating conditions, as the ATC tower would in practice. For instance, if a major congestion problem develops at some location on the taxiway system, the model would continue to send aircraft to that location (since the aircraft paths are fixed) instead of routing them around the bottleneck to relieve congestion.

The above is but one example of the many – relatively minor but irritating – technical deficiencies with which users of existing airfield simulation models may still have to contend. A new generation of such models, currently under development at a number of organizations, will probably overcome such problems. Nevertheless, even the existing models are now meeting with growing acceptance and use for a variety of reasons, including the appeal of the traffic display capabilities of some of them. Two such models worth specific mention are the Airport Machine and SIMMOD, because they are extensively used by the FAA and several airport and ATC organizations overseas. The former is a simulation model that covers the airfield only (runways, taxiways, aprons) while SIMMOD is both an airspace and an airfield modeling tool and can, in fact, be used to simulate operations in a regional ATC system that may include several major airports. (It was used recently, for instance, to assess the impacts of a third major airport in the Chicago area.) SIMMOD enjoys strong and continued support from the FAA and has several companies involved in its application and further development, including a significantly improved forthcoming version. It is thus acquiring the status of the 'semi-official' model of the FAA and may become, at least for a while, the standard tool in highly detailed airfield and airspace analyses.

2.4. Delay analysis

A natural consequence of airport congestion is the widespread incidence of significant airport delays. Airport delays are generally considered as one of the most vexing (and apparently long-term) problems confronting air transportation in much of the world. Estimating airport delays, given actual or anticipated demand and capacity data, is thus a very important aspect of airport planning and design.

The estimation of airport delays poses an interesting challenge to operations researchers: it provides a prominent example of a practical congestion problem for which the results developed in the voluminous literature on classical steady-state queueing theory are largely irrelevant – at least, when it comes to the really interesting cases. The reason is that practically every closed-form expression developed in that body of work assumes (i) arrival and service rates at the queueing system which are constant over time, and (ii) arrival rates which are strictly less than service rates ($\rho < 1$). Both of these assumptions are generally false, often for extended periods of time, at busy airports. Figure 7 shows a typical example of a 24-hour demand profile for a major airport which illustrates this point. A number of authors [see especially Green, Kolesar & Svoronos, 1991; Odoni & Roth, 1983] have demonstrated in detail the poor performance of even sophisticated approximation schemes which try to estimate airport delays by judiciously applying steady-state expressions to the type of situation shown in Figure 7.

In the absence of reliable, closed-form analytical expressions, two approaches have been developed by now for computing airport delays: numerical solution of analytical models and simulation. The former of these approaches was pioneered by Koopman [1972]. His main argument was the following:

(i) The non-homogeneous Poisson process provides a reasonable approxima-

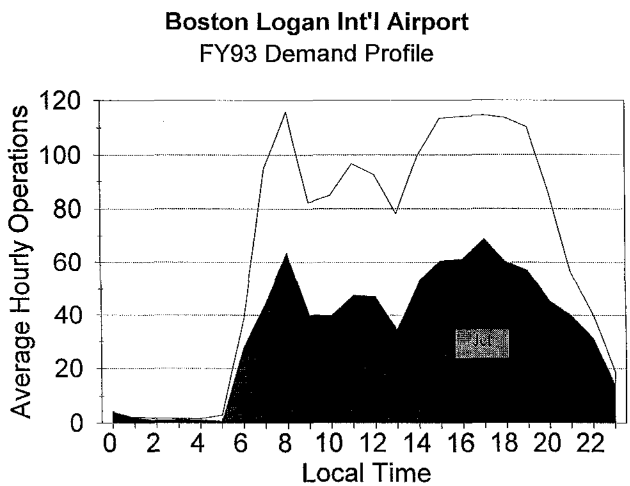


Fig. 7. A typical day's demand profile for Boston Logan.

tion of demand (arrivals and departures) for access to the runway systems of major airports.

(ii) The random variable which describes the duration of service times for arrivals and/or departures at an airport has a probability distribution which is 'less random' than the negative exponential and 'more random' than 'deterministic' (constant).

(iii) Hence, the queueing characteristics of an airport with k runways ('servers') must 'fall somewhere between' the characteristics of the $\mathrm{M}(t)/\mathrm{M}(t)/k$ and the $\mathrm{M}(t)/\mathrm{D}(t)/k$ queueing systems – with the two models providing 'worst case' and 'best case' results, respectively. (The t makes explicit the fact that these models allow for dynamic changes in the demand rates and service rates over time.)

Based on these premises, Koopman solved the Chapman-Kolmogorov equations describing the $\mathrm{M}(t)/\mathrm{M}(t)/k$ and the $\mathrm{M}(t)/\mathrm{D}(t)/k$ queueing systems, using airport demand profiles, such as the one shown in Figure 7, to depict demand rates and typical airport capacities, such as those discussed in Section 2.1, for the service rates. (These service rates can obviously be made to vary over time depending on the particular demand/capacity scenarios being analyzed.) For typical data specific to JFK International and to LaGuardia Airports in New York, Koopman plotted the expected queue length (expected number of aircraft waiting to use the airfield) as a function of time over a 24-hour period and showed that the results obtained from the two queueing systems were very close. He thus argued that queueing characteristics at airports are relatively insensitive to the precise form of the service time distribution.

Extending the work of Koopman, the $\mathrm{M}(t)/\mathrm{E}_k(t)/k$ queueing system was proposed by Kivestu [1976] as a model that could be used directly to compute approximate queueing statistics for airports – rather than solving separately an $\mathrm{M}(t)/\mathrm{M}(t)/k$ and an $\mathrm{M}(t)/\mathrm{D}(t)/k$ model and then 'interpolating' their results. The Erlang (E_k) family of random variables is indeed very flexible and can be used to represent: negative exponential service times (M) by setting $k = 1$; constant service times (D) by setting k to infinity; and 'inbetween' situations by selecting an appropriate value of k. Kivestu also developed an approximation scheme that solves numerically the Chapman-Kolmogorov equations for the $\mathrm{M}(t)/\mathrm{E}_k(t)/1$ system quickly and with high accuracy.

Recent research [Abundo, 1990; Horangic, 1990] has indicated that Kivestu's suggested approach is very practical computationally and produces realistic estimates of airport delays. In fact, a software package, DELAYS, that implements this approach on a range of personal computers is beginning to receive extensive use by a number of organizations, including the FAA.

The second approach, i.e., the use of simulation models to estimate airport delays, while straightforward conceptually, is less attractive than analytical methodologies for the usual reasons. To obtain results which are statistically meaningful, a large number of simulation experiments ('runs') is typically necessary; moreover, determining the correct number of required runs as well as the statistical significance of the results obtained from a simulation model is not an easy task for systems, such as airports, that operate under (and are simulated for) dynamic conditions that may vary greatly over the course of a typical day. An additional complication in the case of airports is that the user of the simulation model is usually interested in estimating distributions, not just averages. For example, a quantity of great interest to ATC and airport planners is the fraction of flights or operations delayed more than a certain amount of time, e.g., 15 or 20 minutes. This increases further computational requirements, as well as the need for sophisticated statistical analysis.

3. Airport passenger terminal design

While the airfield is the element that truly determines an airport's eventual capacity, as well as its land requirements, the passenger terminal plays a crucial role in defining the overall 'image' that most air travelers really associate with any individual airport. The passenger terminal is also a very expensive part of the airport, both absolutely and per gate of aircraft: typical recent fixed costs per gate of major international terminals are in the $15–25 million range, so that a large building with 30 gates could easily cost $600 million to build and equip.

Partly in response to these practical considerations and also because the modeling of passenger terminals offer some interesting challenges, a considerable amount of effort has been dedicated by operations researchers over the last 15 years to developing improved quantitative tools for supporting the design and operation of these large buildings. Much of this work is reviewed, in far more detail than is possible here, in two recent publications, a special issue of *Transportation Research* [1992, see especially Tosic, 1992] and a volume published by the Transportation Research Board [1987]. However, despite significant progress, it is fair to say that the 'state of the art' in this area still leaves much to be desired.

The 'elements' of a passenger terminal are generally classified into three categories: the terms *processing facilities* (e.g., ticket counters, check-in counters, security and passport controls, baggage carousels, customs counters, etc.), *holding areas* (e.g., lobbies, atria, gate lounges, etc.) and *passageways* (e.g., corridors, escalators, moving sidewalks, etc.) are often used to refer to servers or spaces where passengers and other airport users, respectively, (i) are processed or are offered a service, (ii) congregate while waiting for an event (such as initiation of boarding) or spend time voluntarily, and (iii) travel on foot. Holding areas also include ancillary facilities and concessions. In the case of processing facilities questions of interest include: the determination of the number of servers needed; the amount of linear frontage necessary (e.g., for check-in counters or for baggage claim devices); and the area that should be set aside for the servers and for the queues that form in front of them. For holding areas, the key issues concern the amount of space that should be reserved and the physical configuration of that space. This physical configuration is important, for example, in ensuring that occupants of the holding area (e.g., of a large lobby for departing passengers) have a good sense of orientation or are distributed quite uniformly within the holding area and are not all crowded in one or more parts of it. Finally, in the case of passageways, one is interested in the width or, more generally, the dimensions and physical configuration that would achieve certain flow requirements.

3.1. Standards for 'level of service'

None of these questions could be addressed in the absence of specified standards for desirable 'levels of service' (LOS). For example, in order to determine the required number of check-in counters, one must know what is considered an acceptable amount of waiting time for checking in; similarly, in determining how much space should be set aside for a departure gate lounge, one should have some idea of acceptable levels of 'crowding' (number of lounge occupants per unit area). As the reader would surmise, specifying such general-purpose LOS standards is far from simple. However, since 1981 a consensus of airlines and airport operators has been reached on a set of verbal definitions of LOS ranging from 'A' ('excellent level of service; conditions of free passenger flow; no delays; excellent level of comfort') to 'E' ('inadequate level of service; condition of unstable flow; unacceptable delays; inadequate level of comfort') – with 'F' labelled as 'entirely unacceptable LOS'. To be useful in guiding passenger terminal design and operations, it is, of course, necessary to 'translate' these verbal descriptions of LOS into quantitative sets of standards. Obviously, one such set should specify waiting time standards for processing facilities. Another set should specify 'space per occupant' standards for holding areas and for areas used for queueing in front of processing facilities. Yet a third set should refer to passenger and other flows in passageways.

To date, two such sets of standards have been proposed, one for space and the other for passageways, and seem to be gaining wide international acceptance. Table 1 lists the space LOS standards [AACI/IATA, 1981]. LOS standards for passageways [Benz, 1986; Fruin, 1971] are stated in terms of passengers per meter width (of the passageway) per minute with typical recommended values around 40 (LOS = C) or 54 (LOS = D).

Unfortunately, no similar progress has been made in the case of waiting time standards. Various airlines and airport operators have, however, developed their own standards in this regard. Ashford [1988] presents a good survey of the standards used by several major European Airport Authorities. As a typical

Table 1
LOS standards for holding areas

LOS	Wait circulate	Check-in bag claim	Holdroom inspection
A	2.7	1.6	1.4
B	2.3	1.4	1.2
C	1.9	1.2	1.0
D	1.5	1.0	0.8
E	1.0	0.8	0.6

example, the British Airports Authority specifies that: 95% of passengers should be processed at baggage check-in in less than 3 minutes overall and 80% of passengers in less than 5 minutes during peak traffic periods; or, that 95% of passengers on 'regular' flights should be screened for security in less than 3 minutes apiece, while for 'high security' flights 80% of passengers should be screened in less than 8 minutes. This illustrates the fact that many of the existing waiting time standards are distributive in nature ('X% wait less than Y minutes') and that different standards may be set for different processing facilities, to take into consideration the nature of the service rendered or factors related to the 'psychology of queueing'.

3.2. *Models*

As in the case of the airfield, both analytical models and simulations are available for passenger terminals. For questions involving individual elements of the passenger terminal, it is often quite easy to develop approximate analytical models that help resolve such issues as the capacity or the number of servers required at a processing facility, or the amount of space needed to reach a desirable LOS standard. These models are sometimes as simple as a 'back-of-the-envelope' formula. The following is a typical example.

Example. To estimate the amount of space to be provided for a departure lobby shared by several flights at a time, a model recommended by IATA [1981] states:

$$S = K_{\mathrm{LOS}}(\lambda D + 2\sqrt{\lambda D}) \tag{6}$$

where

S = the amount of space needed (in m^2),
K_{LOS} = the appropriate LOS space coefficient (see Table 1),
λ = the number of passengers per hour flowing through the lobby during the terminal's period of peak demand,
D = the average 'dwell' time (in hours) of passengers in the lobby during this period of peak demand.

If, for instance, the rate of flow of passengers through the lobby during the peak period is $\lambda = 1200$ per hour, the average dwell time in the lobby is 20 minutes ($D = 1/3$) and the desired LOS is C ($K_{\mathrm{LOS}} = 1.9$ from Table 1) then the space required is about 844 m^2 (or about 9000 square feet).

This example illustrates a number of general observations about this area of work. First, models, whether analytical or simulations, are very 'data intensive'. Any type of planning for passenger terminals requires quite extensive data (or, as an alternative, numerous assumptions) about passenger flows and passenger characteristics and behavior – in our example, how many passengers will be passing through the lobby and how long they will be staying there. Obviously, such data or assumptions are subject to massive uncertainty, especially when one is designing a terminal for a 'target date' which may easily be 10 years in the future.

It is important, therefore, to have models which are flexible and easy-to-use so that planners can perform extensive sensitivity analyses and come up with 'robust' operational designs.

Second, most models include some probabilistic considerations. In Equation (6), for instance, the underlying assumption is that the flow of passengers through the lobby can be approximated as a Poisson process and the term $2\sqrt{\lambda D}$ provides extra space to accommodate a '2-standard-deviations' fluctuation from the average rate of flow.

Finally, the application of these models relies heavily on the user's judgement and experience. An obvious example is the choice of the LOS factor, K_{LOS}, in (6) and the trade-offs it implies between passenger comfort during peak traffic periods, on the one hand, and utilization of the terminal during off-peak periods, on the other.

A few specific families of analytical models that have been applied to airport terminals and deserve mentioning are: (i) models based on classical steady-state queueing theory; (ii) graphical analyses using 'cumulative diagrams'; and (iii) 'macroscopic' models based on geometrical probability.

Models in (i) include some of the earliest applications of queueing theory, described in an outstanding early book by Lee [1966]. Unfortunately, the practical value of these models is quite limited, since most processing facilities in passenger terminals are essentially never in steady-state: they are almost always undergoing some kind of dynamic change. The only significant exceptions would seem to involve facilities involving several parallel servers shared by many flights (or by several airlines, as in the case of common check-in areas encountered at many European airports) where roughly steady rates of demand at high server utilization levels are achieved for significant periods of time.

To take into account dynamic phenomena, deterministic queueing models have been used extensively [Tosic, 1992]. The technique most often used is that of cumulative diagrams, i.e., graphical displays of the total number of arrivals and service completions over time at a processing facility [Newell, 1971]. This approach assumes that the patterns of demand over time, at least on average, are known. Thus, it is best suited for short-term planning, e.g., in reconfiguring a particular space within an existing structure.

A more recent family of models [Bandara & Wirasinghe, 1991, 1992] has attempted to develop approaches for optimizing the dimensions of passenger terminals and for selecting the most appropriate geometrical configuration among various alternatives (pier terminals, satellite-type terminals, etc.). The primary criterion used is a weighted objective function of average walking distance and average travel time on mechanical devices (e.g., people movers) for passengers, including those transferring between flights. To compute the relevant objective functions for any given configuration and allocation of flights to gates some use is made of geometrical probability concepts. A body of work with a related theme uses mathematical programming models (integer programming, the assignment problem, etc.) to allocate a day's schedule of arriving and departing flights to aircraft gates in such a way as to minimize expected passenger walking distances or, more generally, travel times in the terminal [Babic, Teodorovic & Tosic, 1984; Mangoubi & Mathaisel, 1985]. Obviously, these models must take into account the amount of time a gate will be occupied by an aircraft, as well as the fact that some gates may be able to accommodate only certain specific types of aircraft.

4. Air traffic control studies

Operations research models have also been used widely to analyze a number of problems arising in air traffic control (ATC). Probably the two principal areas of application are in estimating sector workload and in determining flow management strategies for air traffic. We shall discuss briefly below the first of these two topics.

One of the principal indicators of the workload of an ATC system is the frequency with which the system is required to resolve 'conflicts' between pairs of aircraft. Estimating this frequency is an important step in the process of either designing (i.e., specifying the dimensions and overall configuration) ATC 'sectors' (= local control regions) or of determining the personnel or equipment requirements of such sectors.

In attempting to estimate these conflicts a volume of *protected airspace* is defined around each flying aircraft. Typically this volume has the shape of a cylinder with radius r and height h. A conflict is then defined as the overlap of any two of these cylinders in airspace. (One can visualize the cylinders as flying ice-hockey 'pucks' – the flat sides of the pucks always parallel to the horizontal plane – and a conflict as the collision of two pucks.) Depending on the dimensions of the protected volume, a conflict then takes different interpretations. For example., when $2r$ and h are set, respectively, equal to the minimum horizontal and vertical ATC separation requirements between two aircraft (usually 3–5 n. miles and 1000 feet, respectively, for the airspace above the continental United States) a conflict is equivalent to a violation of ATC separation standards. In this case, each aircraft can be approximately represented by a point mass at the midpoint of its associated cylinder's axis. If, on the other hand, $2r$ and h correspond respectively to the wingspan and to the height of the hull of aircraft, a conflict is equivalent to a mid-air collision. In this latter case, representation of the aircraft as a cylinder is clearly an idealization that may necessitate some adjustments of the estimated probabilities and frequencies of conflicts.

Several authors have analyzed various types of conflicts arising in connection with the general situation described above. One such type is *overtaking conflicts*, which refer to cases involving aircraft flying on the same 'airway' (i.e., a straight-line path between two navaids) and in the same direction, but not at the same speed. Given the probability distribution of aircraft speeds, the traffic density on the airway and the length of the airway, Siddiqee [1974] and Dunlay [1975] have derived a simple expression for the expected number of overtaking conflicts per unit of time on the airway, assuming that aircraft enter the airway independently in a Poisson manner and that each maintains its preferred constant speed. This

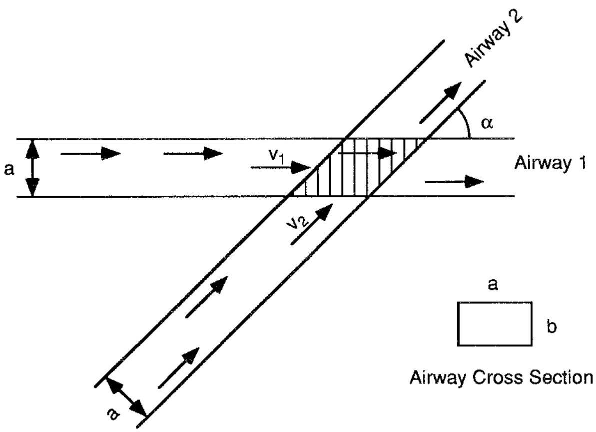


Fig. 8. Geometric model of the intersection of two airways.

expected number indicates how often the air traffic controller responsible for the airway will have to oversee the safe overtaking of one aircraft by another. It is interesting to note that the airspace model was developed independently of an essentially identical model in traffic flow theory which estimates the frequency with which a car, moving at a given constant speed on a lightly traveled stretch of highway, passes or is passed by other cars, given the overall probability distribution of car speeds.

A second fundamental model is the one that predicts the number of conflicts at the *intersection of two airways* (Figure 8). In its simplest form [Dunlay, 1975] this model assumes that the flows of traffic on each of the airways constitute independent homogeneous Poisson processes with intensities of f_1 and f_2 aircraft per unit of time. It is also assumed that the width and height of both airways (denoted by a and b in Figure 8) are equal to zero and that all aircraft on airways 1 and 2 travel at constant speeds V_1 and V_2, respectively. The model was subsequently extended to the more general geometry of Figure 8 by Endoh [1982] and to intersections of an arbitrary number of airways (not just two) and more generally distributed traffic flows (not just Poisson) by MacDonald [1987].

The third fundamental model in this area is the so-called *gas model*. As its name suggests, this uses the classic physical model of gas molecules in a heated chamber to estimate the number of conflicts between aircraft occupying some part of airspace. In its original form [Marks, 1963], the model assumes that aircraft are uniformly and independently distributed within an area, when movement is restricted to the horizontal plane (think of pucks on an ice-hockey rink) or a volume, when changes in aircraft altitude are permitted [see also Flanagan & Willis, 1969]. It is further assumed that aircraft travel in straight lines in directions which are independently and uniformly distributed between 0 and 2π and with speeds that are independent of the direction of travel and are drawn, independently for each aircraft, from a probability distribution $p_V(.)$.

Clearly this simple representation may be better suited to an uncontrolled part of airspace occupied by pleasure fliers who may indeed be flying in random directions. Nevertheless, estimates of frequencies of potential conflicts obtained through the gas model [Alexander, 1970; Graham & Orr, 1970] played an important role in the work of the Department of Transportation's ATC Advisory Committee which in 1969 charted much of the future of the ATC system in the United States. These estimates were used extensively in determining some of the computer specifications for the system. It was argued that estimates of the expected number of conflicts obtained through the model (i.e., assuming random directions of travel) are conservative upper bounds for the actual situation in controlled airspace where directions of travel are far from random. Indeed it was subsequently shown [Endoh, 1982; Endoh & Odoni, 1983] that, for any ensemble of aircraft flying in straight lines, the expected number of conflicts is maximized when the directions of aircraft travel are uniformly distributed between 0 and 2π, as assumed in the gas model. This may be a counter-intuitive result since it might seem that the expected number of conflicts would be maximum if half the aircraft were flying in one direction and the other half in exactly the opposite direction.

Another more obvious result can be proven easily: the distribution of aircraft with respect to *altitude* which *minimizes* the expected number of conflicts, everything else being equal, is the uniform distribution [Endoh, 1982; Ratcliffe & Ford, 1982]. Thus, the universal practice in ATC that concentrates aircraft around specific altitudes, typically at 500- or 1000- or 2000-feet intervals, increases the *potential* number of conflicts that must be resolved.

The assumption of uniformly distributed travel directions in the 2-dimensional and 3-dimensional gas models is unnecessarily restrictive. Endoh [1982] and Endoh & Odoni [1983] present a *generalized gas model* in which the probability distribution for the direction, θ, of travel can be arbitrary. For example, if for a given region we have

$$f_\theta(\theta) = K \times \delta(\theta - \gamma) + \frac{1-K}{2\pi}, \quad \text{for } 0 \le \theta \le 2\pi \tag{7}$$

where $\delta(.)$ indicates the impulse function, then a fraction K of aircraft are flying in the same direction, γ, while the directions of the remaining aircraft are uniformly distributed between 0 and 2π. The generalized gas model can be used to derive expressions for the expected number of conflicts for several interesting traffic scenarios. In addition, it is shown in Endoh & Odoni [1983] that the overtaking conflicts model and the airways intersection model are but special cases of the generalized gas model.

All the models described above have one consistent aspect, namely they show that the expected number of conflicts per unit of time is proportional to the square of 'traffic density', with the constant of proportionality depending on the 'geometry' and other characteristics of the particular situation being modeled. The term traffic density may mean the number of aircraft per unit volume or per unit area of airspace in the case of 3-dimensional and 2-dimensional gas models, respectively, or the number of aircraft per unit length in the case of overtaking and intersection conflicts where traffic is restricted to move along straight line paths (airways).

PART II. URBAN TRANSIT

1. Background

The purpose of this section is to review how operations research techniques have been used to solve planning and operational problems in public transport authorities. This review is not exhaustive, but emphasizes operations research techniques that have been applied or that, in our opinion, have potential applications. A further limitation is that it examines public transport problems in isolation from two larger problems in which they are embedded. First, public transport is just one element of the larger urban transportation problem. Street congestion, availability of reserved lanes (road-space reserved for public transport), parking policy and pricing, and limits on private car utilization all have a significant impact on public transport planning and operations. Second, urban development plans may directly influence decisions concerning the location of major public transport infrastructure (e.g., subways, light rail) so as to encourage desired urban development. These important interrelations are not considered here.

2. The elements of transit problems

Figure 9 presents the various elements necessary for the planning and operation of public transport service from an operations research perspective. As can be seen, there are many interrelations between the elements of the problem. However, mathematically it is not feasible to take into account all these elements at the same time. It is part of the art of operations research to split large interconnected problems into pieces that are significant yet tractable. These pieces often correspond to the internal organization of public transport authorities. Generally strategic planning activities, such as designing new routes, or constructing new infrastructure, are separated from the timetabling and vehicle and crew scheduling elements, which are in turn separated from the day-to-day management of operations. We will describe these problems in more detail in the following sections together with the operations research techniques used in each context.

We classify public transport problems into those at the strategic, tactical and operational levels. Strategic planning is concerned with long-term development of the system including fixed investments such as subway and light rail facilities, and also in major modifications to the bus route network. Since it is difficult for the public to adapt to major revisions in the bus network, most changes of this type are the result of comprehensive studies and must stay in place for some time to achieve their full potential impact. As such, there is a stability to the bus network structure, making it appropriately part of the strategic planning problem.

The tactical planning problem includes minor route revisions, the structuring of routes in high ridership corridors and the assignment of service frequencies to routes. These aspects of the overall problem often do not receive much attention, with some of the attendant strategies being implicitly part of the planning

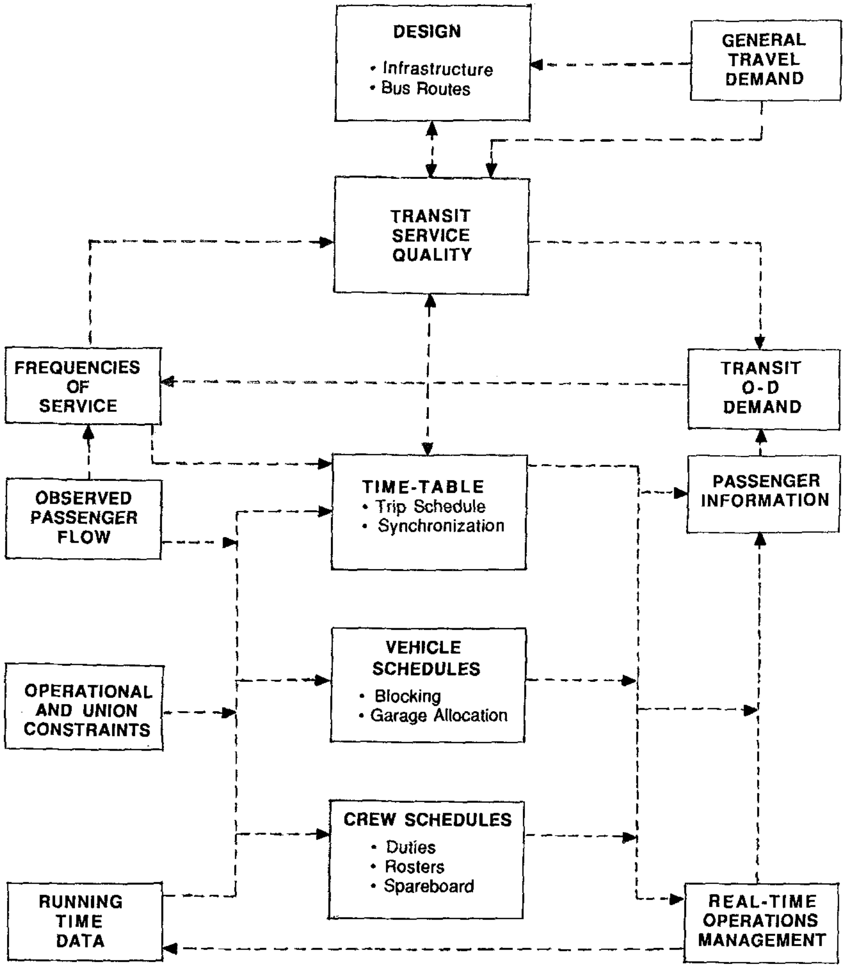


Fig. 9. Main interrelations between the various elements of the transit planning and operations problem.

department while others are included in the operational planning function. We will however discuss these problems independently because we believe that further research is warranted in this area.

Operational planning encompasses all scheduling-related activity, including the production of timetables, the scheduling of vehicles, the generation of daily duties for drivers, and the construction of rosters comprising the daily duties of individual drivers together with their days off and holidays.

Finally, we will briefly refer to the increasingly important topic of public transport operations control.

3. The strategic planning problem

The transit network design problem is one of a class of very difficult network design problems covering most scheduled carrier systems and including airlines and rail passenger systems. There are several reasons for this complexity. First is the existence of the route as an intermediate structure between the link and the full network. Second is the strong interdependence between the design of routes by the operator and the choice of travel itinerary carried out independently by users. We are in a user optimum context rather than a system optimum context because the network planning agency does not decide the actual routing (or routing strategies) of the traffic on its network (which is the case for telephone companies, pipelines, trucking, etc.). In addition, the objective function is ill-defined as in many design problems. The planner certainly wants to minimize the total travel time of the population, but he or she also wants to give 'equitable' access to the system to every taxpayer, design 'attractive routes' that will be 'easy' to use, maybe 'favor' the use of heavy infrastructure such as subways to minimize operating cost, while taking into account 'acquired rights' to certain services and dealing with political pressures for better services in individual political jurisdictions.

For all these reasons, it seems appropriate to favor an approach by which the design of new infrastructure or new bus routes is done interactively, with the planner choosing scenarios (route and network structures) while the computer performs the evaluation of these scenarios, providing the planner with information to help in the design of new scenarios, etc. Indeed, while many attempts have been made to design routes heuristically by computer [see, for example, Hasselström, 1981; Lampkin & Saalmans, 1967; Last & Leak, 1976; Mandl, 1980], all transit network models which have received widespread application are of an interactive type.

VIPS-II, developed jointly by VTS and Stockholm Transport, is the only one of these models which allows the automatic generation of transit routes, as described in Hasselström [1981]. This is essentially a typical location–allocation approach. It is first assumed that there is a direct transportation link between each pair of zones, and the frequency of service is calculated using a simple model of the type to be described in the following section. The algorithm allocates passengers to alternative paths between each origin–destination pair so as to minimize travel time. It sequentially drops from the network the least attractive routes, reallocating the passengers affected to the remaining routes and recalculating the frequency of services until a satisfactory solution is reached (see Figure 10). In addition to this automatic generation of routes, routes can be specified by the planner with the model then being used for network evaluation as in the approaches described below.

Several computer packages have been designed for transit network simulation and evaluation and so can be used in the interactive design of bus routes and transit infrastructure. Among these are Transcom [Chapleau, 1974], Transept [Last & Leak, 1976], NOPTS [Rapp, Mattenberger, Piguet & Robert-Grandpierre, 1976], VIPS [Andreasson, 1976] and its successor VIPS-II, U load [UMTA, 1977], TRANSPLAN [Osleeb & Moellering, 1976], IGTDS [General Motors, 1980], Ma-

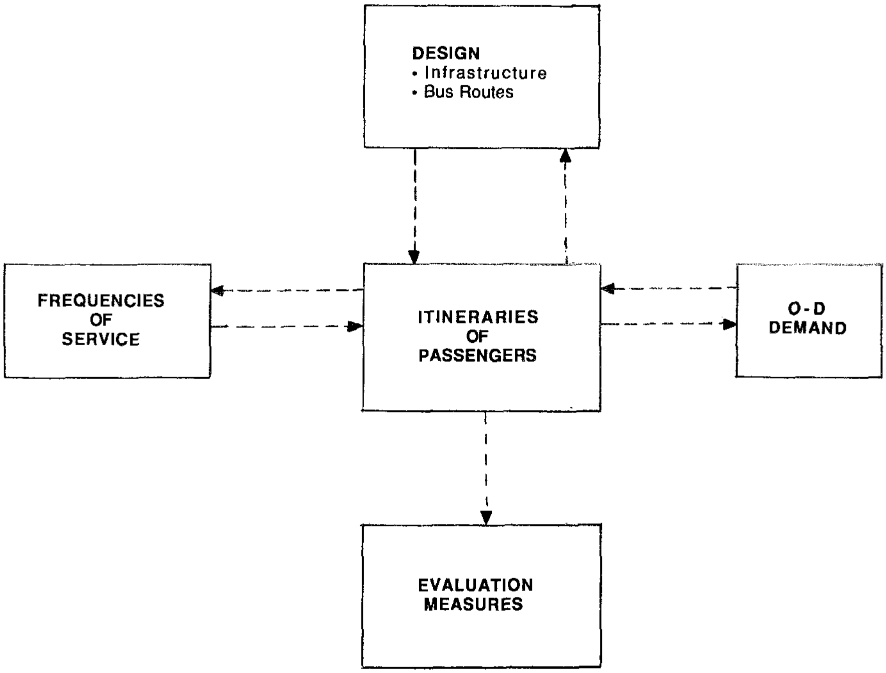


Fig. 10. The strategic planning problem.

dituc [Chapleau, Allard & Canova, 1982] and EMME/2 [Babin, Florian, James-Lefebvre & Spiess, 1982, 1984; Florian, 1985; Florian & Spiess, 1986]. Of these, VIPS-II and EMME/2 appear to be among the most sophisticated evaluation models. Both provide extensive graphical displays that assist the user in the evaluation of a transit network. In each, the planner can define scenarios interactively and the program then forecasts how the passengers will use the system, providing answers through a series of graphical and numerical outputs. Whereas VIPS-II is strictly a public transport planning tool, EMME/2 provides a multi-modal urban transportation planning environment letting the user define up to ten modes of transportation. Both systems have been used extensively around the world.

Any transit network evaluation tool supposes that the transit network is given, including the frequency of service on each route, and that total travel demand by transit between all zones is known (in some cases a modal split function is used to extract the transit demand from a total origin–destination travel matrix). From this information it calculates the itinerary of passengers and the transit flow on each line together with average trip duration between each O–D zone and other local and global measures. In this context, one feedback loop possible is the reevaluation of the matrix by using a mode choice function. The process may be repeated until a stable solution is obtained.

A key operations research problem in any public transport network evaluation model, as well as in any heuristic approach to network design, is to calculate the passenger assignment to routes in the network. The key behavioral principle is well-accepted; the transit user wants to travel at minimum 'inconvenience', this being a weighted combination of access time, waiting time, travel time, travel cost and transfer inconvenience. Any generalized cost functions can be calibrated and tested in each context assigning different weights to the various factors composing it.

However, even given this objective function, the solution of the passenger assignment problem is not as simple as it might appear at first sight. First, the O–D matrix is an aggregation at a zonal level; it will not always be true that persons living in the same zone and travelling to the same destination, will choose the same route. One reason for this is that their exact locations in the zone may dictate different choices, given the importance of access time. Second, several transit routes may serve a given street segment or a given intersection. The optimal choice of itinerary for a transit user may very well be dictated by the arrival of the first bus on any of these routes. In the same way, the transit user may well change his or her itinerary according to the presence of a given bus at a certain street intersection, etc.

The transit passenger assignment problem has been studied by many authors, either as a separate problem [see, for example, Andreasson, 1976; Chriqui, 1974; Chriqui & Robillard, 1975; Dial, 1967; Le Clerq, 1972; Marguier & Ceder, 1984; Rapp, Mattenberger, Piguet & Robert-Grandpierre, 1976; Spiess, 1983], or as a sub-problem of more complex models, such as transit network design [see, for example, Hasselström, 1981; Jansson & Ridderstolpe, 1992; Lampkin & Saalmans, 1967; Mandl, 1980; Schéele, 1977], or multimodal network equilibrium [Florian, 1977; Florian & Spiess, 1983].

Following Dial [1967], many of these algorithms are based on the assumptions of deterministic running times and exponentially distributed headways on all routes, in which case the market share for a particular route is simply its frequency share. Jansson & Ridderstolpe [1992] show that when headways are deterministic rather than exponential, the frequency share model does not always hold, with the result depending on the degree of schedule coordination. Chriqui & Robillard [1975] and Marguier & Ceder [1984] both examine route choice among routes on a common path. Chriqui & Robillard proposed a heuristic for selecting a set of acceptable paths among which the passenger would select the first bus to arrive so as to minimize the expected total travel time. They show that their heuristic is optimal when all routes have exponentially distributed headways and identical running times, but Marguier & Ceder showed that this heuristic is not optimal for all headway distributions. Jansson & Ridderstolpe present an alternative heuristic for this problem for the deterministic headway case.

Spiess [1983], Spiess & Florian [1989] and De Cea et al. [1988] present linear programming solutions to the transit path assignment problem as a relaxation of a mixed integer program under the assumption of exponentially distributed headways. There are two main differences between these models and other approaches. First and foremost, in these models the traveller does not simply choose a path, but rather selects a strategy. While in earlier methods the transit traveller's route choice is limited to one path, defined as a sequence of route segments, links or transfer points, the strategies considered in Spiess' approach allow the transit rider to select any subset of paths leading to the destination with the first vehicle to arrive, determining which of the alternative routes is actually taken on an individual trip. This concept of optimal strategies allows more realistic modelling of the traveller's behavior. Second, the resulting transit assignment problem, formulated as a mathematical programming problem, may be solved directly without recourse to heuristic approaches.

The principal difference between the approaches of Speiss and De Cea is in the representation of the problem in network form. Speiss' algorithm is in the form of a shortest-path problem, whereas de Cea's approach results in network search to define feasible paths, followed by Chriqui's heuristic to determine optimal paths. Speiss' algorithm solves the problem in polynomial time and so may be applied even to very large transit networks. This method is used in EMME/2. A method based on similar ideas has also been developed and implemented in Torino by Nguyen & Pallottino [1988].

In this general area of transit network design, additional development may come from the automatic generation of options or marginal improvements to systems that could be proposed to the planner. These suggestions could be generated through a set of rules gathered together in an expert system by the planners themselves. However, we do not envisage a totally automatic design tool for the transit network design problem in the near future, expert system or not.

4. Tactical planning

The tactical planning problem for transit has received relatively little serious attention. Figure 11 summarizes the various elements of this problem. The transit network is assumed known and the objective is to allocate the available resources (buses and drivers) in order to provide the best possible service at the least cost. Several problems can be studied at this level, including the transit assignment problem discussed at length in the previous section. Another problem that has received some attention [Fearnside & Draper, 1971; Last & Leak, 1976; Schéele, 1977], is that of choosing the optimal frequencies for bus routes, given a fixed fleet of vehicles, in order to minimize total travel time. The frequency of service influences the itinerary and duration of travel, which in turn may influence travel demand.

There are two fundamentally different ways of defining the frequency of service in a transit network. Most transit authorities set frequencies, or the number of vehicles required, so as to obtain a given number of passengers per vehicle at the peak loading point on each route, subject to a minimum frequency constraint. Typically both acceptable peak loads and minimum frequencies vary by time of day. This type of approach has been formulated (with extensions) by Ceder [1984]. An alternative approach treats the problem as one of resource allocation, with the resources, for example bus-hours, being allocated across routes and across time periods so as to maximize some objective subject to passenger loading constraints. Furth & Wilson [1981] showed that this problem could be

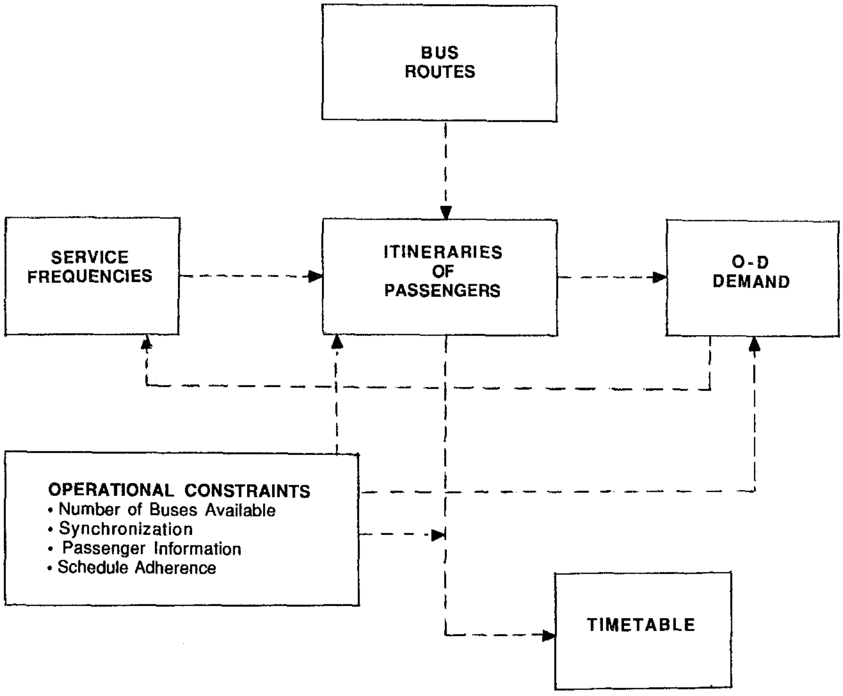


Fig. 11. Tactical planning level.

formulated either as a fixed demand minimization of total expected passenger waiting time, or as a variable demand consumer surplus maximization with very similar results. This work was extended by Koutsopoulos, Odoni & Wilson [1985] and later by Wilson & Banasiak [1984] to treat more complex objective functions and constraints. A similar approach was also taken by Hasselström [1981] for frequency determination in the VIPS model. Constantin in her thesis [Constantin, 1993] and in a subsequent paper [Constantin & Florian, 1993] use a non-linear bi-level programming approach to solve the problem of allocating the fleet of vehicles to bus routes in order to minimize passenger travel time given the passenger will travel according to the model of Spiess.

Another topic that falls within the tactical planning area is design of operating strategies in high ridership corridors which might be amenable to market segmentation techniques designed either to increase service quality or to increase productivity (or both). Turnquist [1979] formulated the design of zonal express service as a dynamic programming problem and this work was later extended by Furth [1986] to zonal design for bidirectional local service including light direction deadheading, and to branching as well as linear corridors. These models have received little application by transit planners largely because most real transit corridor planning situations have such limited design options that they can typically be solved by enumeration.

5. Operational planning models

In operational planning, we consider that the requirement for service on each route is known (frequency of service or number of buses), that the running times (to travel between various points on a route) are known for each period of the day, and all operating and crew contract constraints are fixed. We then have to define timetables, vehicle schedules and crew schedules. The determination of the number of spares drivers required and their schedules of work are also part of the operational plan.

5.1. Timetabling

Given a required frequency of service and the running times, the generation of trips is normally immediate. Combining trips into blocks is generally fairly trivial. After ensuring that a minimum layover is allowed to maintain the regularity of services [Wilson, 1986], a first-in first-out (FIFO) strategy is optimum on a route-by-route basis both for the number of vehicles necessary, and for the minimization of total duty times [Ceder, 1986; Saha, 1970]. A vehicle is freed as soon as it is no longer required or would have to wait an excessive amount of time for the next trip assignment. A vehicle is dispatched from the depot when needed for the trip schedule. However, it may be necessary in the timetabling problem to take into account crew scheduling constraints. For example, blocks of less than two hours may generate additional pay to the driver (because a piece of work for the driver must be at least two hours) and must thus be avoided as much as possible. It may also be useful to integrate a longer layover into a block that could be used later for a coffee break or lunch break for the driver. Some of these situations are described in Hamer & Séguin [1994].

There are few mathematical considerations in timetable preparation, but there is a need for a sophisticated management information system with graphic interfaces. Several bus scheduling software packages have been developed and are available commercially, most integrated with crew scheduling software [see Daduna & Wren, 1988; Desrochers & Rousseau, 1992; Paixao & Daduna, 1994; Rousseau, 1985; Wren, 1981].

However, when the service requirement is specified in terms of number of buses in service by time of the day, the problem of producing a smooth timetable is more complex, particularly at times when the number of buses required changes significantly [Ceder, 1986]. The adjustment of trip departure time can be difficult particularly when one is dealing with complex bus routes with several branches. One other topic which falls under the timetabling umbrella is the setting of timetables on connecting routes so that the service for transferring passengers is optimized [Bookbinder & Désilets, 1992]. While this topic is complex, particularly in the full network context, there may be scope for future application for route pairs involving large transfer volumes. Other references on the subject can be found in Desrochers & Rousseau [1992] and Paixao & Daduna [1994].

5.2. Vehicle scheduling problems

Vehicle scheduling refers to the class of problems in which vehicles must be assigned to timetabled trips or tasks, in such a way that each task is carried out by one vehicle, a given set of constraints is satisfied and cost is minimized. For simplicity here, the term 'trip' will be used to denote all vehicle tasks defined in the timetable (whether trips or blocks of trips). Very often, the cost structure reflects the fact that the number of vehicles must be minimized. We assume that for each trip, the departure time, trip duration, and origin and destination terminal are known. In addition to regular trips, we assume that deadheading trips are allowed between terminals. A survey paper and several papers on specific vehicle scheduling problems are published in Paixao & Daduna [1994].

5.2.1. The simple case

In this section we will discuss the problem as stated above with only one depot and no complicating constraints. The reader who requires more details on this problem could consult Ceder & Stern [1981], Bodin, Golden, Assad & Ball [1983], Carraresi & Gallo [1984b], and Scott [1986].

Desrosiers, Soumis & Desrochers [1982] presented a scheduling model which takes into account the use of a single vehicle to serve many chains of trips (see Figure 12). In this approach, each trip in the schedule is represented by an arc with a minimal and maximal flow of one, thus requiring a unit of flow to traverse it. There is no specific cost associated with the trips themselves because they all must be included in the solution. All feasible deadhead trips are also represented by arcs with a minimum flow of zero, a maximum flow of one and an associated cost c_{ij} corresponding to the cost of deadhead time and distance. Equivalently, a cost is associated with a deadhead trip to and from the garage $(c_{.i}, c_{i.})$. The originality of this formulation lies in the fact that the garage is not represented by a single node but by a sequence of time nodes. All buses in the garage can flow freely from one time node to the next time node without cost. A bound N_k on the maximum number of buses in the garage at time period k can be easily imposed on the solution. Finally, a return arc from the end of the day to the beginning of

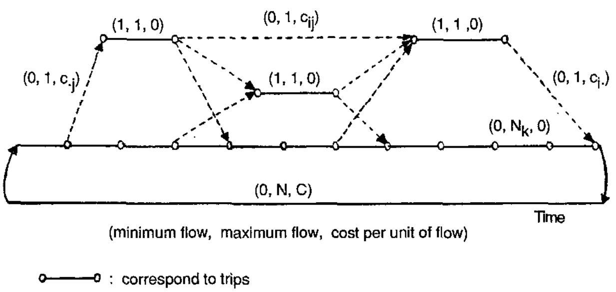


Fig. 12. A time space network for the transit bus scheduling problem.

the day ensures the conservation of flow and counts the number of buses required to operate the schedule. The fixed cost C of a vehicle and an upper bound N on the number of vehicles can be imposed on this arc.

A standard minimum cost flow algorithm can be used to solve this problem. This formulation accurately takes into account the number of buses required without generating too many arcs. However, Scott [1986] demonstrated that if deadhead cost is proportional to travel time and waiting time, which is characteristic of a broad class of vehicle scheduling problems, there need not be any conflict between the two objectives of minimizing fleet size and minimizing the cost of operating a schedule. In other words, there is never, in general, any need for the application of a fleet size penalty to the creation of independent chains when one intends to find a minimum-cost scheduling solution which uses a minimal fleet.

5.2.2. Complicating constraints

We will consider here three types of complicating constraints that can destroy the structure of the basic vehicle schedule observed earlier. First, we may want to impose a minimum length or a maximum length on the duration of a vehicle duty (set of trips assigned to a vehicle between its visits to the garage). The minimum length constraint arises when there is a minimum duration of each piece of work assigned to a driver, for example two hours. In this context, a vehicle duty of less than two hours duration will cost at least two hours of paid time. It is thus advisable to try to minimize the number of such vehicle duties. The maximum duration constraint is in general related to the necessity to refuel vehicles.

Both of these problems have to be dealt with heuristically. In practice the following heuristics have been found useful:

Step 0. Solve the problem without these maximum or minimum length constraints.

Step 1. Split the day into a certain number of time periods.

Step 2. For each time period, split the vehicle duty into two partial duties: the set of trips finishing before or during the time period, and the set of trips finishing after the time period. Solve an assignment problem between the two sets of partial duties taking into account the cost of having vehicle duty durations below the lower limit, and prohibiting duties from being longer than the upper limit.

Step 3. Cycle through the time period or various sets of time periods until no improvement is possible.

The second set of complicating constraints arises because one may want to take advantage of possible flexibility in the timetable concerning the departure time of the trip (for example, trip departure time may vary within ± 5 minutes of a desired time). Desrosiers, Soumis & Desrochers [1984] have described a combinatorial approach for this type of problem in which they solve a set-covering problem with column generation over all possible feasible vehicle duties respecting the time windows. These vehicle duties are not predefined but are generated as the algorithm progresses using shortest path or dynamic programming algorithms.

A third set of constraints occurs for companies that operate their fleets out of several garages such that each vehicle belongs to a single garage and must return to that garage at the end of its daily cycle. Each garage will typically have a capacity constraint and some garages may also have certain minimum usage levels, even if this is not cost-effective, for example in the case of a new garage located in an industrial rather than residential area. The company may also require that a certain number of vehicles be always out of the garage to facilitate maintenance activities. Finally for control reasons, the company may not allow vehicles assigned to a route to be split between several garages.

Various approaches have been proposed for this problem [Bodin, Golden, Assad & Ball, 1983; Carraresi & Gallo, 1984a; Mandl, 1980; Maze & Khasnabis, 1985; Rousseau, Lessard & Désilets, 1988]. Essentially, a choice must be made between clustering the trips and assigning them to the garage first, and then scheduling vehicles in each garage separately, or first scheduling the whole fleet as if there were only one garage and then assigning the resulting trips to each garage. If there are no constraints on separating a given route among several garages, Carraresi & Gallo [1984a] have proposed an approach based on a Lagrangean relaxation of the problem coupled with a subgradient optimization technique to find an optimal solution. On the other hand, Rousseau, Lessard & Désilets [1988] have proposed a simple clustering heuristic that can take into account all kinds of operational constraints. They sequentially assign all the trips of each route to the best garage and then search for a feasible solution by neighborhood exchange heuristics, moving routes from one garage to another in order to improve the cost or feasibility in relation to the various constraints. They then optimize the schedule for each garage. These methods can also be used to study the location of new garages and evaluate their impact on operational costs.

Although crew costs often represent more than 60% of operating costs, it does not appear useful to consider the crew scheduling constraints in preparing the vehicle schedule, except for the minimum length of a piece of work and some other similar constraints. For small problems with particular constraints it would also seem worthwhile to take these constraints into account [Scott, 1984]. Finally, in some contexts, it is necessary to schedule drivers and buses together because each driver is assigned to a given bus. This often happens in intercity contexts where deadheading drivers to a relief point is not practical. It also happens in some urban contexts where the authority wants to keep a given driver on a given bus in order that each driver takes better care of his/her bus.

5.3. Crew scheduling

Crew schedule planning in a transit company involves two different operations: the generation of the drivers' daily duties and the assignment of a set of daily duties (called a roster) to each driver. We will consider each of these problems in turn. The proceedings of the most recent workshops on computer scheduling of public transport constitute an excellent set of references on the subject [Daduna & Wren, 1988; Desrocher & Rousseau, 1992; Paixao & Daduna, 1994; Rousseau, 1985; Wren, 1981]. Wren & Rousseau [1994] also provides a survey of this problem.

5.3.1. Duty generation

The problem can be expressed simply as the creation of a set of duties of minimum cost, ensuring that each bus trip is covered by a driver (or crew), and that all union contract rules are respected. Several authors favor an approach based on the set-covering formulation of the problem. Theoretically, in this approach all feasible duties have to be generated according to union contract rules, and the least cost subset of duties that covers all bus trips is mathematically selected. The formulation is as follows:

$$\text{Min} \sum_j c_j x_j$$

$$\text{subject to} \quad \sum_j a_{ij} x_j = 1 \tag{1}$$

$$\sum_i b_{ij} x_j \geq d \tag{2}$$

$$x_j = 0 \text{ or } 1$$

where x_j is the duty variable with cost c_j, $a_{ij} = 1$, if duty j performs the task (trip) i. Constraint (1) insures that all tasks will be performed by a duty, while constraint (2) imposes other constraints on the problem such as a minimum number of duties of certain types, etc.

If the operating or union rules restrict a driver to working on only one bus route, and if there are not too many feasible duties due to strict working rules, this approach may be feasible. Otherwise, if for example full interlining is permitted (meaning that drivers are allowed to work on several routes during a duty), the number of feasible duties is extremely large and simply cannot be generated. Advocates of the set-covering approach then either have to split the problem into several sub-problems (evening problem, morning problem, etc. [see Ward, Durant & Hallman, 1981]) or heavily restrict the generation of duties according to heuristic rules (for example by eliminating relief points that are unlikely to be useful). This is the case for Busman, Ramcutter, and Crewsched, among other software packages [see Rousseau, 1985, for several papers based on this approach].

The main difficulty with this approach is that there does not exist a good set of rules for identifying what are likely to be good feasible duties. A typical duty may be very efficient under certain operating and union rules while being very inefficient under another set of rules. Desrochers [1986] and Desrochers & Soumis [1989] used the set-covering approach to the problem, but sequentially generated only the duties useful to improve the solution at each step of the algorithm. Thus only a small subset of the duties have to be generated. It seems, however, that the technique will apply only to small to medium-size problems, but could be quite useful in the context where the drivers have to work on only one bus route. Rousseau, Desrosiers & Dumas [1994] have reported on the results obtained with this approach.

The most common approach to the transit crew scheduling problem, in both manual and computer-based methods, proceeds as follows:

First, the vehicle duties, called blocks (the itinerary of a vehicle between its departure from the garage and its return to the garage) in the literature, are split into pieces of work (the period of time during which a driver works continuously). Second, the pieces of work are grouped to form driver duties, and finally, the solution is improved heuristically and/or manually. A matching or an assignment problem formulation for the construction of duties is generally used. Various strategies are possible for splitting the blocks into pieces of work. Most approaches use heuristics adapted to the type of problem considered. In RUCUS [Wilhelm, 1975], the first general package developed for transit crew scheduling, a set of parameters enables the user partially to determine the method by which the pieces will be defined. In Sage runcutting and RUCUS II [Luedtke, 1985], an interactive system for transit crew scheduling, the user has direct control on the ways the pieces are identified. Hoffstadt [1981] and Parker & Smith [1981] include heuristic approaches adapted to their respective problems related to German and British working conditions. The HASTUS system uses a relaxation of the problem to determine how to define the pieces. Ball, Bodin & Dial [1981] form the pieces and blocks by sequentially matching the trips.

While using this general decomposition approach, HASTUS uses a mathematical model to split the blocks into pieces of work. A full description of the algorithm can be found in Rousseau, Lessard & Blais [1985]. The authors chose to solve a relaxation of the crew scheduling problem first, trying to keep the main features of the problem while relaxing all the details. The idea is to retain the essentials of the bus schedule structure and its impact on the contract rules and costs while greatly simplifying the precise bus schedule.

First the user must define periods, normally about 30 minutes in length, although they can vary, depending on the problem, from 15 to 60 minutes. The schedules of all buses are approximated so that all blocks both start and end at the beginning (or end) of a period. Reliefs (the times, and places, at which change of drivers is possible) can also occur only at the beginning of a period. Moreover, the problem is further relaxed by requiring that the duties selected be sufficient to cover the total requirement of drivers per period instead of requiring that they exactly cover all the blocks individually.

All feasible duties made up of two or three pieces of work covering an integer number of periods and satisfying the union contract are generated. The generating rules eliminate, under instructions from the user, duties which are of no interest or which do not correspond to unwritten rules or common practice (for example, a duty with a long break during a peak hour). All the duties are costed according to the union contract in use. A linear program then searches for the set of duties that will provide the number of drivers required during each period and satisfy other additional constraints at minimum cost.

One type of additional constraint is related to the union contract. All the constraints related to the total number of duties of a certain type are considered. These include constraints on the minimum or maximum number of duties without break and constraints on the maximum number of duties with certain undesirable characteristics like long spreads (the time between driver sign on and sign off for the day), small breaks, etc.

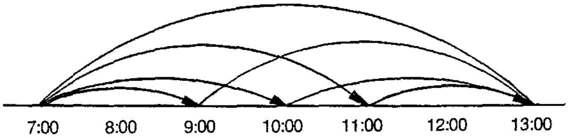
The relaxation is called HASTUS-macro and can also be used as a planning tool for the evaluation of the economic impact of changes in the union contract or in the service level. By changing the cost parameters or the service level, a new solution is obtained and can be compared with the current solution. The paper by Blais & Rousseau [1982] describes in detail the HASTUS-macro system and its use in that context. In practice, this relaxation has been found to be very accurate, producing cost estimates within 2% of the real operating costs for drivers.

The HASTUS-macro solution gives a guide to the number of duties of each type that should be constructed in a final solution. The next step uses this information to split the blocks into pieces of work that correspond as closely as possible to the ones generated in the HASTUS-macro relaxation.

Each block is sequentially split into pieces of work by solving a shortest path problem. It is interesting to note that the problem of splitting a block into pieces can be expressed as the problem of finding a path from the starting time of the block to its ending time using arcs that correspond to pieces of work. Figure 13 illustrates this concept with the feasible pieces of work (we are assuming here a minimum length of two hours) represented by arcs. We have assumed for simplicity that a relief opportunity exists at every hour in this example.

The cost associated with each piece of work in the flow formulation corresponds to a penalty indicating whether this given piece will increase or decrease the difference between the number of pieces of work of that type desired, and the number of pieces of work of that type already present in the current solution. We cycle through all blocks until no improvement can be achieved towards coming closer to the number of desired pieces of each type indicated by HASTUS-macro.

Once the blocks have been split into pieces of work, a matching algorithm is used to generate duties [Lessard, Minoux & Rousseau, 1989; Rousseau, Lessard & Blais, 1985] and several heuristics are used to improve the solution. The whole system is embedded in a software system [Rousseau & Blais, 1985] and is installed in more than seventy cities around the world (including Montréal, Québec, Calgary, Seattle, New York, Boston, Stockholm, Singapore, Newcastle and Barcelona). The savings of this system over a manual solution varies but is typically in the range of 15% to 20% of unproductive costs (unproductive time, overtime, special payments, etc.), which could represent 2–3% of total operating costs (for example, approximately 3 million dollars annual savings in Montréal). This application has been described in *Interfaces* [Blais, Lamont & Rousseau, 1990], where the mathematical model is also provided.



The feasible pieces are:

7 - 9	7 - 11	9 - 11	10 - 13
7 - 10	7 - 13	9 - 13	11 - 13

Fig. 13. The flow formulation for the partition of a block.

5.3.2. The rostering problem

A roster can be defined as the set of duties assigned to a particular driver. Making up rosters may be easy or difficult depending on the particular union contract. In most North American companies the assignment of duties is left to the drivers themselves and the decision is made on a seniority basis: this leaves no role at all for mathematical modelling! On the other hand, with many European companies, work load must be distributed evenly among the drivers. In fact, in these problems, few penalties are associated with unpleasant duties. Rather, these duties are evenly distributed among all drivers, who all receive approximately the same monthly pay. The rules for making monthly, weekly, or in some cases, several month rosters vary widely from one company to another. While there seem to be very few general packages available for rostering problems in transit, there is a consensus on the most efficient methods to be used to build such rosters. A more detailed description of the approach used can be found in Rousseau [1984], Belletti, Davini, Carraresi & Gallo [1985] and Hagberg [1985].

The basic method is a heuristic based on a construction phase and on a marginal improvement phase. The construction phase can be described as follows: given a set of partial rosters P_i constructed from day 1 to day i (initially the set of partial rosters is composed of all duties for day 1), construct P_{i+1} by defining an assignment problem between the partial rosters in P_i and the duties that have to be performed on day $i + 1$.

The assignment problem can be solved with standard well known algorithms. The objective function is a penalty function measuring how the resulting partial rosters will differ on various criteria from the ideal partial rosters. For example, if the main criterion is to equalize as much as possible the total number of driving hours among the rosters, the penalty incurred by assigning duty k to partial roster j might be the square of the difference between the number of driving hours in the resulting roster and the desired level for such a number after $i + 1$ days. A general penalty function is described in Nicoletti [1973, 1975] for the airline rostering problem.

Many kinds of exchange algorithms well known in routing and scheduling can be used to obtain marginal improvements. An interesting approach, however, has been applied by Carraresi & Gallo [1984a] for the transit crew rostering problem in which the objective is to minimize the maximum total driving time in the period considered. This problem is described as the multi-level bottleneck problem and it is shown that the problem is NP-complete. An asymptotically optimal algorithm is presented for this case. In general, the objective function is different, but their type of algorithm could easily be adapted to the general situation when starting from a feasible solution.

Sequentially consider each day i in the rostering period.

Option 1. Remove from each roster the duties of day i. Define and solve an assignment problem between the duties just removed and the partial rosters with the penalty function in use, allowing only feasible assignments.

Option 2. Define and solve an assignment problem with on one side the partial rosters composed of all the duties starting before day i, and on the other side the partial rosters composed of all the duties starting on or after day i. Of course, only feasible assignments are allowed and the same cost function as in the construction heuristic is used.

If no improvement in the objective function is obtained after all days in the rostering period have been considered, the process stops. If an improvement has been achieved, another iteration is performed on the period. Based on the results of Carraresi & Gallo on their specific problem, we can expect this algorithm to perform very well on all rostering types of problems even if the theoretical properties demonstrated by the authors may not hold in all situations. These methods obviously need to be adapted to particular situations, but they have been used successfully in several contexts.

5.4. Spare driver planning

A final aspect of operational planning which is now beginning to receive more serious attention is the determination and assignment of spare drivers (often referred to as the extraboard) who are required to be available both to fill in for absent regular drivers and to provide unscheduled extra service which is required from time to time. This problem can be viewed at both a strategic level and a tactical level, and both problems have received some attention in the past decade [Koutsopoulos & Wilson, 1987]. At the strategic level the principal problem is to find the optimal extraboard size which minimizes total operator cost while providing a given level of service reliability. This problem has been formulated as a non-linear program with the non-linearity arising from the nature of the overtime term in the total cost function [Hickman, Koutsopoulos & Wilson, 1988] and has also been solved heuristically [Shiftan & Wilson, 1993a,b].

The tactical problem is, given a certain size of extraboard on a particular day, how to schedule these spare drivers to provide the best coverage of the (uncertain) absences and extra work. This problem is more closely related to the driver scheduling problem discussed in the previous section, but has a much stronger flavor of scheduling under stochastic demands. It also has been approached both heuristically and optimally [Kaysi & Wilson, 1990; Koutsopoulos, 1990]. It is an area where future work is likely to have a real impact on practice.

6. The dispatching problem

In some respects dispatching problems are similar to operational planning problems. Vehicle schedules including additional or chartered trips must be planned, and spare drivers must be allocated to perform pieces of work left open by regular drivers on vacation or reporting sick. The same type of techniques as described earlier can be used in this context. Here again the technology is opening a whole new field related to the optimum control of vehicle schedules. With an automatic vehicle location and control system (AVLC), the dispatcher now has several strategies for dealing with a vehicle that is delayed: take measures to help the vehicle catch up with the schedule, or delay the following vehicles in order to insure an even frequency and prevent bunching. Priority for buses at intersections may be used to reduce delay and better control adherence to schedules. In fact, with all the information collected with an AVLC system, even the timetables may be calculated in order to facilitate adherence to schedules. In heavily congested cities (e.g., Shanghai) it is also possible to imagine some real time scheduling of buses [Li et al., 1992] where at each departure of a bus, it is decided which stops will be served or skipped. As we can see, there are new and challenging problems ahead in this field.

7. Conclusion

This review has been brief, with very little detail. We have demonstrated, however, that in the transit area, operations research techniques are alive and well. They are being used to assist in the planning and operation of many transit networks around the world. All the problems have not been solved, and even for the ones that have been solved, better techniques could be found. New problems are arising from the implementation of new technology in transit. Better information will be available from the in-vehicle microcomputer and better information will be required by the public in general. There will be plenty of challenging operations research problems in the transit area for many years to come.

References

Part I

AACI/IATA (Airport Associations Coordinating Council/International Air Transport Association) (1981). *Guidelines for Airport Capacity/Demand Analysis*, Montreal, Que., and Geneva.

Abundo, S. (1990). An approach for estimating delays at a busy airport, S.M. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Alexander, B. (1970). Aircraft density and midair collision. *Proc. IEEE* 58, 377–381.

Ashford, N. (1988). Level of service design concepts for airport passenger terminals – a European view. *Transportation Planning Tech.* 12, 5–21.

Babic, O., C. Teodorovic and V. Tosic (1984). Aircraft stand assignment to minimize walking. *ASCE J. Transportation Engrg.* 110, 55–66.

Bandara, S., and S.C. Wirasinghe (1991). Optimum geometries for pier type airport terminals. *ASCE J. Transportation Engrg*, 117.

Bandara, S., and S.C. Wirasinghe (1992). Walking distance minimization for airport terminal configurations. *Transportation Res.* 26A, 59–74.

Benz, G.P. (1986). *Pedestrian Time-Space Concept – A New Approach to the Planning and Design of Pedestrian Facilities*, Parsons, Brinckerhoff, Quade and Douglas, New York.

Blumstein, A. (1959). The landing capacity of a runway. *Oper. Res.* 7, 752–763.

Blumstein, A. (1960). An analytical investigation of airport capacity, Report TA-1358-G-1, Cornell Aeronautical Laboratory, Ithaca, NY.

Dear, R. (1976). The dynamic scheduling of aircraft in the near-terminal area, Technical Report R76-9, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Dunlay, W.J. (1975). Analytical models of perceived air traffic control conflicts. *Transportation Sci.* 9, 149–164.

Endoh, S. (1982). Aircraft collision models, Technical Report R82-2, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Endoh, S., and A.R. Odoni (1983). A generalized model for predicting the frequency of air conflicts. *Proc. Conf. Safety Issues in Air Traffic Systems Planning and Design*, Princeton University, Princeton, NJ.

Flanagan, P.D., and K.E. Willis (1969). Frequency of airspace conflicts in the mixed terminal environment, Report of Department of Transportation Air Traffic Control Advisory Committee, Vol. 2, U.S. Dept. of Transportation, Washington, DC.

Fruin, J.J. (1971). *Pedestrian Planning and Design*, Metropolitan Association of Urban Designers and Environmental Planners, New York.

Graham, W., and R.H. Orr (1970). Terminal air traffic flow and collision exposure. *Proc. IEEE* 58, 328–336.

Green, L., P. Kolesar and A. Svoronos (1991). Some effects of nonstationarity on multiserver Markovian queueing systems. *Oper. Res.* 39, 502–511.

Harris, R.M. (1972). Models for runway capacity analysis, Report No. FAA-EM-73-5, The MITRE Corporation, McLean, VA.

Hockaday, S.L.M., and A. Kanafani (1972). A methodology for airport capacity analysis. *Transportation Res.* 8, 171–180.

Horangic, B. (1990). Some queueing models of airport delays, S.M. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

IATA (International Air Transport Association) (1981). Guidelines for airport capacity/demand management, Report AACC/IATA TRAP WG, Cointrin–Geneva, Switzerland.

Kivestu, P. (1976). Alternative methods of investigating the time-dependent M/G/K queue, S.M. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA.

Koopman, B.O. (1972). Air terminal queues under time-dependent conditions. *Oper. Res.* 20, 1089–1114.

Lee, A.M. (1966). *Applied Queueing Theory*, Macmillan, London, St. Martin's Press, New York.

MacDonald, B. (1987). A generalized model for the prediction of controller intervention rates in the en route ATC system, Ph.D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Mangoubi, R., and D.F.X. Mathaisel (1985). Optimizing gate assignments at airport terminals. *Transportation Sci.* 19, 173–188.

Marks, B.L. (1963). ATC separation standards and collision risk, Technical Note MATH 91, Royal Aircraft Establishment, Farnborough.

Newell, G.F. (1971). *Applications of Queueing Theory*, Barnes and Noble, Boston, MA, Chapman and Hall, London.

Odoni, A.R. (1972). Efficient operation of runways, in: A.W. Drake, R.L. Keeney and P.M. Morse (eds.), *Analysis of Public Systems*, MIT Press, Cambridge, MA.

Odoni, A.R. (1991). *Transportation Modeling Needs: Airports and Airspace*, U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA.

Odoni, A.R., and R. De Neufville (1992). Passenger terminal design. *Transportation Res.* 26A, 27–36.

Odoni, A.R., and E. Roth (1983). An empirical investigation of the transient behavior of stationary queueing systems. *Oper. Res.* 31, 432–455.

Psaraftis, H.N. (1980). A dynamic programming approach for sequencing groups of identical jobs. *Oper. Res.* 28, 1347–1359.

Ratcliffe, S., and R.L. Ford (1982). Conflicts between random flights in a given area. *J. Navigation* 35, 47–71.

Reich, P.G. (1966). Analysis of long-range air traffic systems: separation standards. *J. Navigation* 19, 88–98, 169–186, 331–347.

Siddiqee, W. (1974). Air route capacity models. *Navigation* 20, 296–300.

Swedish, W.J. (1981). Upgraded FAA airfield capacity model, Report MTR-81W16 (also Federal Aviation Administration Report FAA-EM-81-1), The MITRE Corporation, McLean, VA.

Transportation Research (1992). Special issue: Airport landside planning and operations, 26A, S.C. Wirasinghe and J.P. Braaksma, guest editors, pp. 1–92.

Transportation Research Board (1987). Measuring airport landside capacity, Special Report 215, National Research Council, Washington, DC.

Tosic, V. (1992). A review of airport passenger terminal operations analysis and modeling. *Transportation Res.* 26A, 3–26.

Venkatakrishnan, C.S., A.I. Barnett and A.R. Odoni (1993). Landings at Logan Airport: Describing and increasing airport capacity. *Transportation Sci.*, 211–227.

Part II

Andreasson, I. (1976). A method for the analysis of transit networks, in: M. Ruebens (ed.), *2nd European Congress on Operations Research,* North-Holland, Amsterdam.

Babin, A., M. Florian, L. James-Lefebvre and H. Spiess (1982). EMME/2: Interactive graphic method for road and transit planning. *Transportation Res. Record* 866, 1–9.

Babin, A., M. Florian, L. James-Lefebvre and H. Spiess (1984). EMME/2: Une méthode interactive graphique pour la planification du transport urbain multimodal. *Routes Transports* XIV(4), 22–31.

Ball, M.O., L.D. Bodin and R. Dial (1981). Experimentation with a computerized system for scheduling mass transit vehicles and crews, in: A. Wren (ed.), *Computer Scheduling of Public Transport: Urban Passenger Vehicle and Crew Scheduling,* North-Holland, Amsterdam, pp. 319–326.

Belletti, R., A. Davini, P. Carraresi and G. Gallo (1985). BDROP: A package for the bus drivers' rostering problem, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2,* North-Holland, Amsterdam, pp. 319–326.

Blais, J.-Y., J. Lamont and J.-M. Rousseau (1990). The HASTUS vehicle and manpower scheduling system at the Société de transport de la Communauté urbaine de Montréal. *Interfaces* 20(1), 26–42.

Blais, J.-Y., and J.-M. Rousseau (1982). HASTUS: A model for the economic evaluation of drivers' collective agreements in transit companies. *INFOR* 20, 3–15.

Bodin, L., B. Golden, A. Assad and M. Ball (1983). Routing and scheduling of vehicles and crews: The state-of-the-art. *Comput. Oper. Res.* 10(2), 65–211.

Bookbinder, J.H., and A. Désilets (1992). Transfer optimization in a transit network. *Transportation Sci.* 26(2), 106–118.

Carraresi, P., and G. Gallo (1984a). A multilevel bottleneck assignment approach to the bus drivers' rostering problem. *European J. Oper. Res.* 16(2), 163–173.

Carraresi, P., and G. Gallo (1984b). Network models for vehicle and crew scheduling. *European J. Oper. Res.* 16(2), 139–151.

Ceder, A. (1984). Bus frequency determination using passenger count data. *Transportation Res. A* 18A(5/6), 439–453.

Ceder, A. (1986). Methods for creating bus timetables. *Transportation Res. A* 21A(1), 59–83.

Ceder, A., and H. Stern (1981). Deficit function bus scheduling with deadheading trip insertions for fleet size reduction. *Transportation Sci.* 15, 338–363.

Chapleau, R. (1974). Réseaux de transport en commun: Structure informatique et affectation, Ph.D. Thesis, Dept. I.R.O., Univ. de Montréal, Que.
Chapleau, R., B. Allard and M. Canova (1982). Madituc: Un modèle de planification opérationnelle adapté aux entreprises de transport en commun de taille moyenne, Publication #265, Centre de Recherche sur les Transports, Univ. de Montréal, Que.
Chriqui, C. (1974). Réseaux de transport en commun: Les problèmes de cheminement et d'accès, Ph.D. Thesis, Dept. I.R.O., Univ. de Montréal, Que.
Chriqui, C., and P. Robillard (1975). Common bus lines. *Transportation Sci.* 9, 115–121.
Constantin, I. (1993). Optimisation des fréquences d'un réseau de transport en commun, Publication #881, Centre de Recherche sur les Transports, Univ. de Montréal, Que.
Constantin, I., and M. Florian (1993). Optimizing frequency in a transit network; A non-linear bi-level programming approach, Publication #914, Center for Research on Transportation, University of Montreal, Que.
Daduna, J.R., and A. Wren (eds.) (1988). *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems 308, Springer-Verlag, Berlin.
De Cea, J. (1986). Rutas y strategias en modelos de asignacion a redes de transporte publico, *IV Congreso Panamericano de Ingenieria de Transito y Transporte*, Chile, December 1986.
De Cea, J., J.P. Bunster, L. Zubieta and M. Florian (1988). Optimal strategies and optimal routes in public transit assignment models: an empirical comparison. *Traffic Engrg. Control* 29(10), 520–526.
Desrochers, M. (1986). La fabrication d'horaires de travail pour les conducteurs d'autobus par une méthode de génération de colonnes, Publication #470, Centre de Recherche sur les Transports, Univ. de Montréal, Que.
Desrochers, M., and J.-M. Rousseau (eds.) (1992). *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems 386, Springer-Verlag, Berlin.
Desrochers, M., and F. Soumis (1989). A column generation approach to the urban transit crew scheduling problem. *Transportation Sci.* 23(1), 1–13.
Desrosiers, J., F. Soumis and M. Desrochers (1982). Routes sur un réseau espace–temps. *Proc. of the Meeting of the Administrative Sciences Association of Canada*, Operations Research Section 3(28) [also Publication #236, Centre de Recherche sur les Transports, Univ. de Montréal, Que.].
Desrosiers, J., F. Soumis and M. Desrochers (1984). Routing with time windows by column generation. *Networks* 14, 545.
Dial, R.B. (1967). Transit pathfinder algorithm. *Highway Res. Record* 205, 67–85.
Fearnside, K., and D.P. Draper (1971). Public transport assignment – A new approach. *Traffic Engrg. Control*, November, 298–299.
Florian, M. (1977). A traffic equilibrium model of travel by car and public transit modes. *Transportation Sci.* 11(2), 166–179.
Florian, M. (1985). Le logiciel EMME/2 et son utilisation dans la région de Montréal. *Transports* 30, 141–148.
Florian, M., and H. Spiess (1983). On binary mode choice/assignment models. *Transportation Sci.* 17(1), 32–47.
Florian, M., and H. Spiess (1986). Models and techniques for the planning of urban and regional transportation networks, Publication #463, Centre de Recherche sur les Transports, Univ. de Montréal, Que.
Furth, P. (1986). Zonal route design for transit corridors. *Transportation Sci.* 20(1), 1–12.
Furth, P., and N.H.M. Wilson (1981). Setting frequencies on bus routes: Theory and practice. *Transportation Res. Record* 818, 1–7.
General Motors Transportation System Center (1980). Interactive graphic transit design system (IGTDS) demonstration study: Bellevue, Washington, Final Report, Vols. I and II (EP-80023A) (GMTSC, Warren, MI).
Hagberg, B. (1985). An assignment approach to the rostering problem: An application to taxi vehicles, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam, pp. 313–318.

Hamer, N., and L. Séguin (1994). The HASTUS vehicle and crew scheduling system: New version and features, in: J. Paixao, and J.R. Daduna (eds.), *Computer-Aided Transit Scheduling*, Springer-Verlag, Berlin, to appear.

Hasselström, D. (1981). Public transportation planning – A mathematical programming approach, Ph.D. Thesis, Department of Business Administration, University of Gothenburg, Gothenburg.

Hickman, M.D., H.N. Koutsopoulos and N. Wilson (1988). A strategic model for operator workforce planning in the transit industry. *Transportation Res. Record* 1165, 60–68.

Hoffstadt, J. (1981). Computerized vehicle and driver scheduling for Hamburg Hochbahn Aktiegesellschaft, in: A. Wren (ed.), *Computer Scheduling of Public Transport – Urban Passenger Vehicle and Crew Scheduling*, North-Holland, Amsterdam, pp. 35–52.

Jansson, K., and B. Ridderstolpe (1992). A method for the route-choice problem in public transport systems. *Transportation Sci.* 26(3), 246–251.

Kaysi, I., and N. Wilson (1990). Scheduling transit extraboard personnel. *Transportation Res. Record* 1266, 31–43.

Koutsopoulos, H. (1990). Scheduling for extraboard operations in transit systems. *Transportation Sci.* 24(2), 87–104.

Koutsopoulos, H., A. Odoni and N. Wilson (1985). Determination of headway as a function of time-varying characteristics on a transit network, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam, pp. 391–414.

Koutsopoulos, H., and N. Wilson (1987). Operator workforce planning in the transit industry. *Transportation Res. A*, 21A, 127–138.

Lampkin, W., and P.D. Saalmans (1967). The design of routes, service frequencies and schedules for a municipal bus undertaking: A case study. *Oper. Res. Q.* 18(4), 375–397.

Last, A., and S.E. Leak (1976). Transept: A bus model. *Traffic Engrg. Control* 17(1), 14–20.

Le Clerq, F. (1972). A public transport assignment method. *Traffic Engrg. Control* 14(2), 91–96.

Lessard, R., M. Minoux and J.-M. Rousseau (1989). A new approach to general matching problems using relaxation and network flow subproblems. *Networks* 19(4), 459–480.

Li, Y., J.-M. Rousseau and M. Gendreau (1994). Real-time scheduling for public transit operations with bus-location information, in: J. Paixao, and J.R. Daduna, *Computer Aided Transit Scheduling*, Springer-Verlag, Berlin, to appear.

Li, Y., J.-M. Rousseau and F. Wu (1992). Real-time scheduling on a transit bus route, in: M. Desrochers, and J.-M. Rousseau, *Computer Aided Transit Scheduling*, Springer-Verlag, Berlin.

Luedtke, L.K. (1985). Rucus II: A review of system capabilities, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam, pp. 61–110.

Mandl, C. (1980). Evaluation and optimization of urban public transportation networks. *European J. Oper. Res.* 5, 396–404.

Marguier, P., and A. Ceder (1984). Passenger waiting strategies for overlapping bus routes. *Transportation Sci.* 21(3), 207–230.

Maze, T.H., and S. Khasnabis (1985). Bus garage location planning with dynamic vehicle assignments: A methodology. *Transportation Res. B* 19B(2), 1–13.

Nguyen, S., and S. Pallottino (1988). Equilibrium traffic assignment for large scale transit networks. *European J. Oper. Res.* 37, 176–186.

Nicoletti, B. (1973). Automatic crew rostering. *AGIFORS Symp. Proc.*, Acapulco, pp. 157–182.

Nicoletti, B. (1975). Automatic crew rostering. *Transportation Sci.* 9, 33–42.

Osleeb, J.P., and H. Moellering (1976). TRANSPLAN: An interactive geographic information system for the design and analysis of urban transit systems, URISA 1976. *14th Annual Conference of the Urban and Regional Information Systems Association*, pp. 212–225.

Paixao, J., and J.R. Daduna (1994). *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, to appear.

Parker, M.E., and B.M. Smith (1981). Two approaches to computer crew scheduling, in: A. Wren (ed.), *Computer Scheduling of Public Transport – Urban Passenger Vehicle and Crew Scheduling*, North-Holland, Amsterdam, pp. 193–222.

Rapp, M.H., P. Mattenberger, S. Piguet and A. Robert-Grandpierre (1976). Interactive graphics system for transit route optimization. *Transportation Res. Record* 559, 73–88.

Rousseau, J.-M. (1984). Crew scheduling methods in the operation and planning of transportation systems, in: M. Florian (ed.), *Transportation Planning Models*, North-Holland, Amsterdam, 439–474.

Rousseau, J.-M. (ed.) (1985). *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam.

Rousseau, J.-M., and J.-Y. Blais (1985). HASTUS: An interactive system for buses and crew scheduling, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam, pp. 45–60.

Rousseau, J.-M., J. Desrosiers and Y. Dumas (1994). Results obtained with Crew-Opt, a column generation method for transit crew scheduling, in: J. Paixao, and J.R. Daduna (eds.), *Computer-Aided Transit Scheduling*, Springer-Verlag, Berlin, to appear.

Rousseau, J.-M., R. Lessard and J.-Y. Blais (1985). Enhancements to the HASTUS crew scheduling algorithm, in: J.-M. Rousseau (ed.), *Computer Scheduling of Public Transport 2*, North-Holland, Amsterdam, pp. 255–268.

Rousseau, J.-M., R. Lessard and A. Désilets (1988). Aliages: A system for the assignment of bus routes to garages, in: J.R. Daduna, and A. Wren (eds.), *Computer-Aided Transit Scheduling*, Springer-Verlag, Berlin, pp. 8–14.

Saha, J.L. (1970). Algorithms for bus scheduling problems. *Oper. Res. Q.* 21, 463–474.

Schéele, C.E. (1977). A mathematical programming algorithm for optimal bus frequencies, Institute of Technology, University of Linköping, Sweden.

Scott, D. (1984). A method for scheduling urban transit vehicles which takes account of operation labor cost, Publication #365, Centre de Recherche sur les Transports, Univ. de Montréal, Que.

Scott, D. (1986). Minimal fleet size in transhipment-type vehicle scheduling problems, Publication #487, Centre de Recherche sur les Transports, Univ. de Montréal, Que.

Shiftan, Y., and N. Wilson (1993a). Strategic transit workforce planning model incorporating overtime, absence and reliability relationships. *Transportation Res. Record* 1402, 98–106.

Shiftan, Y., and N. Wilson (1993b). Public transport workforce Sizing Recognizing the Service Reliability Objective, in: J. Paixao, and J.R. Daduna (eds.), *Computer Aided Transit Scheduling*, Springer-Verlag, Berlin, to appear.

Spiess, H. (1983). On optimal route choice strategies in transit networks, Publication #286, Centre de Recherche sur les Transports, Univ. de Montréal.

Spiess, H., and M. Florian (1989). Optimal strategies: A new assignment model for transit networks. *Transportation Res. B*, 23B(2), 83–102.

Turnquist, M. (1979). Zone scheduling of urban bus routes. *Transportation Engrg. J.* 105(1), 1–12.

UMTA/FHWA (1977). *UTPS Reference Manual*, U.S. Department of Transportation, Washington, DC.

Ward, R.E., P.A. Durant and A.B. Hallman (1981). A problem decomposition approach to scheduling the drivers and crews of mass transit systems, in: A. Wren (ed.), *Computer Scheduling of Public Transport – Urban Passenger Vehicle and Crew Scheduling*, North-Holland, Amsterdam, pp. 297–312.

Wilhelm, E. (1975). Overview of the Rucus package driver run cutting program, Preprints, in: L. Bodin, and D. Bergman (eds.), *Workshop on Automated Techniques for Scheduling of Vehicle Operators for Urban Public Transportation Services*, Chicago, IL.

Wilson, N. (1986). Bus operations planning, Preprints, *Management and Planning of Urban Transport Systems from Theory to Practice*, Montréal, Que.

Wilson, N., and M. Banasiak (1984). The role of computer models in transit service allocation decisions. U.S. Department of Transportation, Report MA-11-0041, Washington, DC.

Wren, A. (ed.) (1981). *Computer Scheduling of Public Transport – Urban Passenger Vehicle and Crew Scheduling*, North-Holland, Amsterdam.

Wren, A., and J.-M. Rousseau (1994). Bus driver scheduling, in: J. Paixao, and J.R. Daduna (eds.), *An Overview in Computer-Aided Transit Scheduling*, Springer-Verlag, Berlin, to appear.