Delays in Terminal Air Traffic Control

ROBERT M. OLIVER*
University of California, Berkeley, Calif.

In this paper delay problems where departure and landing operations are performed on a runway or within the glide path of a terminal air traffic system are studied. The distribution of delays, the number of delayed aircraft, and the effect of multiple streams feeding the service system are discussed. Delay models include cases where two priority classes are serviced by one runway. The conditions under which the highest priority leads to lowest expected costs or average delays are also discussed. Priorities are established such that aircraft having a service time less than a critical value are placed in the top priority. The analysis extends to more than two priority classes within a stream of landing and departing aircraft. Delay models are discussed which include, in addition to the usual assumptions about service times, the additional features that constant minimum spacings must be maintained between all users of the runway or glide path and that low-priority aircraft are interposed between high-priority aircraft. Delays due to self-clearing rules are discussed; long runs of one type of aircraft follow long runs of another type of aircraft until all queues are completely dissipated.

I. Introduction

WITHIN the last two decades, air traffic control has been the subject of increased attention at both a local and national level. One of the more important problems that arises is the scheduling and routing of aircraft in the terminal area. Although routing and stacking rules for an airport serving only one type of aircraft may be simple, the operations become increasingly complex as many types of aircraft demand service at one airport and as the effect of various costs and delays are considered. At the present time, congestion and delay problems are due as much to the increased demand for safe control procedures as to the over-all increase in air traffic itself. The advent of high-performance aircraft with high speeds and large fuel consumption characteristics makes timely scheduling rules an important aspect of flying, and it is a common occurrence to find that rules primarily designed for bad-weather flying are used to separate and distinguish aircraft during good weather. The case for instrument flight rules is succinctly described in a recent Project Beacon report[31] that predicts an increase of 300% in controlled aircraft flights by 1975, even though general aviation and air carrier flights in the USA are expected to have an increase of only 80% in the same period.

At the present time, most delays are experienced by aircraft waiting to land or by departing aircraft that wait until they can be interposed between landings. It is probably true that the cause and cure of delays will remain in terminal areas. The very nature of the terminal control process is one of making a random flow of arriving and departing aircraft into a highly controlled and regularly scheduled operation and of resolving potential conflicts in a common service facility.

Of all of the aspects of a modern air traffic control system which include communication, data processing, and equipment maintenance, the one that organizes and selects safe and timely routing and sequencing procedures will play one of the most important roles in future operations. There is every reason to believe that a control system with modern and precise tracking and computing equipment may fail in reaching its goal if scientific operating procedures are not implemented at the same time. The complexity and magnitude of these control systems require faster and surer decisions than the human being can obtain by ad hoc procedures.

A casual reader of air traffic control problems might come to the conclusion that few answers to these control and congestion problems are available in the scientific literature. This paper has been written in the hope of answering some of these questions and pointing out those areas in which theoretical answers to air traffic control problems are available under different names and in different contexts. It is not valuable to list operating rules for each terminal area with its own special problems; instead, one can point out aspects of control procedures which are common to many different airports and show how some of the congestion and delay problems can be remedied.

Many of the mathematical models of congestion have their origins in the theory of queues, where arrivals into a service system may be temporarily denied use of the facility. Queues and the accompanying costs and delays arise because the number of arrivals exceeds, for short periods of time, the number that can be serviced by the system. Since the servicing facility cannot "preservice" aircraft during those periods when the facilities are idle, a queue develops even when the average arrival rate is less than the theoretical service capacity of the facility. Queues disappear only to reappear again; although it is impossible to predict the precise appearance and size of these queues, it is often possible to calculate the statistical fluctuations in the process and thereby infer such important characteristics as average delay or probability of long delays to aircraft.

Perhaps the most revealing contribution of these statistical flow models is the dependence of extreme as well as average delays on the ratio of average flow rates to the theoretical capacity of the service facility. In general, average delays become very large, and large delays become more probable as the average arrival rate approaches the service capacity. For this reason, Blumstein, in a doctoral dissertation[2] and several papers,[1,3] has studied ways of increasing the service capacity of the terminal control area. His results have emphasized the effect of mixed operations, i.e., where both landings and departures use one facility.

In order to obtain expressions for delays, early authors made simplifying assumptions about service and arrival characteristics. For example, service times were assumed to be constant, whereas arrivals over the runway threshold were assumed to be Poisson. With certain airports and under certain operating conditions, such assumptions remain valid to this day; on the other hand, most of them have become

Received by IAS November 26, 1962; revision received March 9, 1964.

* Associate Research Engineer, Institute of Transportation and Traffic Engineering, and Associate Professor of Industrial Engineering.

Table 1 Asymptotic expansion $W(t)$

ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
α	0.33	0.43	0.51	0.59	0.66	0.73	0.80	0.87	0.94
β	3.62	2.66	2.06	1.62	1.26	0.95	0.68	0.43	0.21

outmoded as air traffic procedures were developed to handle large numbers and many types of aircraft.

II. Delays in Single-Runway Operations: Constant Service Times

The simplest queuing models of single-runway operations require fairly restrict assumptions about the service and arrival characteristics of aircraft. If we assume that the probability of counting n landing aircraft in an interval $(0,t)$ is the Poisson distribution,

$$p_n(t) = e^{-\lambda t}(\lambda t)^n/n! \qquad (1)$$

with average arrival rate λ, and if the service time of each aircraft is a constant Δ, the steady-state probability π_n of finding n aircraft in the system is given by Eq. (4). The derivation of this result follows the argument that there are n airplanes in queue at the completion of a service, if there are $n - j + \geqq 1$ in queue before service begins and if $j \geqq 0$ arrive. If there are initially zero in the system, n must arrive during the first service time to give a final count of n. These statements are summarized by the set of linear equations

$$\begin{aligned} \pi_0 &= \pi_0 p_0(\Delta) + \pi_1 p_0(\Delta) \qquad (2) \\ \pi_1 &= \pi_0 p_1(\Delta) + \pi_1 p_1(\Delta) + \pi_2 p_0(\Delta) \\ &\vdots \\ \pi_n &= \pi_0 p_n(\Delta) + \pi_1 p_n(\Delta) + \ldots \pi_{n+1} p_0(\Delta) \\ &\vdots \end{aligned}$$

and the requirement that the probabilities must sum to 1:

$$1 = \pi_0 + \pi_1 + \pi_2 + \ldots \qquad (3)$$

The solution of Eq. (2) subject to (3) was found by Fry,[12] Crommelin,[9, 10] Erlang,[5] and Pearcey[26]:

$$\pi_0 = 1 - \lambda\Delta$$

$$\pi_n = (1 - \lambda\Delta) \sum_{m=1}^{n} (-1)^{n-m} e^{m\lambda\Delta} \times \qquad (4)$$

$$\left\{ \frac{(m\lambda\Delta)^{n-m}}{(n-m)!} + \frac{(m\lambda\Delta)^{n-m-1}}{(n-m-1)!} \right\} \qquad n \geqq 1$$

This equation provides an explicit computation for π_n from which one can obtain the average value and higher moments of the number of airplanes in the service system. To obtain the average queue or stack size, we exclude one in service; mathematically, we multiply π_n by $(n - 1)$ and sum over n. The average delay in queue (exclusive of the constant service time) is

$$w = \lambda\Delta^2/[2(1 - \lambda\Delta)] \qquad (5)$$

since the average delay times the average arrival rate equals the average queue size.

By equating the average delay of aircraft to a function of the average arrival rate λ and the constant service time Δ, Eq. (5) forms the basis for much of the discussion in this paper. It is interesting to note that the term $\lambda\Delta = \rho$ is a number that must always be less than unity and thereby expresses the intuitive result that the average arrival rate must not exceed the service capacity Δ^{-1} over long periods of time. ρ is called the utilization or fractional capacity of the service system.

The first term in Eq. (4) also shows that the probability of an empty facility is

$$\pi_0 = 1 - \rho \qquad (6)$$

alternatively, the fractional capacity or utilization ρ equals the probability $1 - \pi_0$ that the service system is busy. The system utilization also plays a dominant role in the distribution of delays; the formula found by Erlang[5] in 1909 for the probability that the delay (exclusive of runway occupancy time) is less than or equal to t is

$$W(t) = (1 - \rho)e^{\lambda t} \sum_{j=0}^{k} \frac{(j\rho - \lambda t)^j}{j!} e^{-j\rho}$$

$$k\rho \leqq \lambda t \leqq (k + 1)\rho \qquad (7a)$$

Recently, Oliver[33] has computed tables of $W(t)$ for values of ρ in the range 0.01 to 0.99. This equation can be used to obtain the expression

$$W(t) = (1 - \rho)(1 + \lambda t) + 0(t^2) \qquad t < \Delta \qquad (7b)$$

for small t and the asymptotic expansion for the probability that the delay exceeds large values of t,

$$1 - W(t) \backsim \alpha e^{-\beta t \Delta^{-1}} \qquad (7c)$$

Representative values of α and β are shown as a function of ρ in Table 1. Equation (7c) shows the behavior mentioned earlier. The larger the t, the smaller will be the probability that the delay of an airplane will exceed this number. However, as the flow rate of aircraft into the service system approaches the capacity service rate, $\rho \rightarrow 1$, and the exponent in (7c) gets small. Hence, the probability of long delays increases.

Pearcey,[26] in 1948, derived an expression for the probability of successive delays to a group of aircraft, i.e., the number of aircraft that get delayed between two undelayed aircraft. Busy periods of the runway alternate with idle periods, the latter consisting of those points in time when no aircraft are in queue or in service. The probability that no aircraft will be delayed in the service time of the first airplane which begins a busy period is equal to the probability that no aircraft will arrive during Δ, i.e.,

$$f_0 = e^{-\lambda\Delta} = e^{-\rho} \qquad (8a)$$

Since the first airplane to begin the busy period does not get delayed, the distribution of n successive delays

$$f_n = \frac{(n + 1)^n}{(n + 1)!} \rho^n e^{-\rho(n+1)} \qquad n \geqq 0 \qquad (8b)$$

is the Borel distribution[4] for one less than the number serviced in a busy period.

For large values of n, it can be shown that f_n is asymptotically

$$f_n \backsim \frac{e^{n(1-\rho)}\rho^n}{n(2\pi n)^{1/2}} \left[1 - \frac{1}{12n} + \ldots\right]$$

Both this expression and the average value of the number of aircraft delayed in a busy period

$$\sum_{n=0}^{\infty} nf_n = \frac{\lambda\Delta}{1 - \lambda\Delta} = \frac{\rho}{1 - \rho} \qquad (8c)$$

can be obtained by power series expansion of the generating function of Eq. (8b).[14] The similarity between this last equation and the average delay in Eq. (5) is apparent. Both expressions become very large as the system utilization approaches unity.

III. More General Service and Arrival Times

As we have already suggested, the constant service time assumption for all aircraft becomes unduly restrictive as many aircraft types begin to use the runway. When arrivals are Poisson, the work of Pollaczek[29, 30] and Khintchine[18] and the novel arguments of Kendall[17] show that the average delay that elapses between arrival and the instant when service begins is, more generally,

$$w = \frac{\lambda b^{(2)}}{2(1 - \lambda b)} = \frac{b\rho(1 - \chi_b^2)}{2(1 - \rho)} \tag{9}$$

where λ is the average arrival rate, b the average service time, $\lambda b = \rho$ the system utilization, $b^{(2)}$ the second moment of the service time distribution, and χ_b^2 is called the coefficient of variation of the service distribution. This equation is one version of the Pollaczek-Khintchine formula.

Equation (9) points up at least two important characteristics of a single-operation runway: first, the average delay is a sharply decreasing function of the capacity service rate with an infinite delay predicted as the arrival rate approaches the runway capacity. The second important point is that, the more regular the service, the smaller will be average delay for a fixed arrival rate.

To demonstrate the effect of greater or less regularity in service (and arrival) times, it has been useful to consider the family of Erlang density functions:

$$b(t) = k\mu(k\mu t)^{k-1}e^{-k\mu t}/(k - 1)! \qquad k \geqq 1 \tag{10}$$

This function is normalized so that the average service time is μ^{-1} independent of k, the variance is $k\mu^{-2}$, and the coefficient of variation is $\chi_b^2 = k^{-1/2}$. Substituting into Eq. (9) for the average delay gives

$$w = \frac{\lambda}{2\mu} \cdot \frac{k + 1}{k} \cdot \frac{1}{\mu - \lambda} \tag{11}$$

As k goes from 1 to very large values, the distribution of service times goes from the exponential or random case to the constant service queue of Eq. (5) with $\Delta = 1/\mu$. Hence a variety of distributions corresponding to realistic service operations can be approximated by the choice of the two parameters k and μ.

The distribution of delays, as well as asymptotic expressions for large delays, may be difficult to obtain for any given service distribution. However, the case of Erlang services and arrivals has received much attention, and the reader is referred to books by Cox and Smith,[8] Morse,[21] Riordan,[33] and references contained therein. For Poisson arrivals and Erlang services, Morse shows that the probability that delay exceeds t is

$$W(t) = \Sigma_j \alpha_j e^{-\beta_j t} \tag{12}$$

where α_j and β_j are constants that can be calculated from Eq. (10).

One of the important steps in the derivation of Eq. (9) shows that the probability of an empty facility under these less restrictive service assumptions is still $\pi_0 = 1 - \rho$. Again we find that busy and idle periods of the runway alternate; both the probability and length of busy periods increases as the average arrival rate approaches the service capacity, i.e., as utilization $\rho \to 1$.

The distribution of the busy period and the number served in a busy period has been solved in several important cases.[4, 8, 14, 33] The average number of aircraft delayed in a busy period is identical to Eq. (8c) where we interpret $\rho = \lambda b$ rather than the special case $\lambda\Delta$ for constant service times. Cox and Smith[8] have obtained simple expressions for the probability that the length of the busy period exceeds a given value and have also shown conditions under which it can be well approximated by a function of the form $(\alpha/t^{1/2})e^{-\beta t}$, where α and β can be computed from pertinent data of the service times.

No discussion of mathematical queuing models of air traffic operations would be complete without mention of important results obtained by Lindley,[19] since he has obtained numerical solutions for the expected value and variance of delays when regular arrivals feed a facility with service times distributed according to the $k = 2$ case in Eq. (10).

The effect of landing waveoffs on Eq. (12) has been studied by Read and Yoshikawa.[32] Neglecting turn-around time of a waveoff, the queuing formulas still hold. If turn-around times are included, both the service capacity and the delay formulas are modified; the authors obtain good upper and lower bounds for the new expressions.

Blumstein[1] argues that runway operations may not create the dominant service times for landing aircraft and has shown, by a careful analysis of the effect of Instrument Flight Rules (IFR) control rules, that relative velocities of different aircraft types may create delays in the vicinity of the glide path. Our expressions for the delay distribution of landing aircraft are still valid if aircraft are serviced one at a time by the glide path instead of the runway. The major differences seem to be in the ease of measuring or computing service times and delays; some new methods are suggested in a paper by Oliver.[22]

Although the distribution of service times is arbitrary, it is well to remember that Eqs. (5) and (9) are based on the assumptions that arrival and service times are independent of one another and that the condition of the queue and servicing of an airplane in queue begins at the instant that a preceding airplane completes service. If the queue is empty, service begins at the instant of arrival. Each of these assumptions is subject to change as safe operating procedures are introduced into air traffic control system and as aerodynamic characteristics of aircraft restrict the timing and location of service.

IV. Delays with Two Priority Classes and Mixed Operations

In the queuing models of Secs. II and III, we have assumed service in the order of arrival, commonly known as the strict order or FIFO (first-in, first-out) rule. Servicing rules that permute the order of airplanes in queue, without reference to length of service times, do not lead to a reduction in average delay. Wishart[37] has shown that the average delays in a LIFO (last-in, first-out) system are identical to the FIFO systems that we study in this paper. Burke's results[6] for delays and queue sizes when random selections are made from the busy queue can also be used to show that the average queue size and delay are unaffected by the ordering rules. In both of these cases and by intuitive arguments, it seems evident that any ordering that simply permutes the occupants of a queue will increase the variance of the delay distribution.

On the other hand, if priorities for service are established on the basis of service times of different aircraft types, reductions in average delays and costs can be expected; as an example, we might consider the service characteristics of jet vs propeller-driven aircraft. Although airplanes may arrive at random instants of time and may represent independent samples from a population of many different aircraft types, modern detection and communication systems identify the particular aircraft type before actual servicing begins. Almost invariably, aircraft type is associated with service time on the runway or in the glide path, and, to a greater or lesser degree large, high-speed aircraft require long service times, and smaller, low-speed aircraft require shorter service times.

Since the average delay formulas are increasing functions of the moments of the service time distribution, it is not surprising that we can find reductions in average system delay by organizing priority classes to give preference to short and more regular service times. If, in a stream of landing aircraft, we make only two priority distinctions, the average delay to the highest-priority class can be shown to be[7]

$$w_1 = \{\lambda[\alpha b_1^{(2)} + (1 - \alpha)b_2^{(2)}]\}/[2(1 - \rho_1)] \tag{13}$$

where α and $1 - \alpha$ are the fractions of aircraft in the priority classes and the subscripts 1 and 2 refer to high-priority and low-priority classes, respectively. We see that the numerator of Eq. (9) is replaced by the second moment of the mixed distribution, i.e., $b^{(2)} = \alpha b_1^{(2)} + (1 - \alpha)b_2^{(2)}$. The average delay to the low-priority class is

$$w_2 = w_1/(1 - \rho) = w_1/(1 - \rho_1 - \rho_2) \tag{14}$$

It is interesting and useful to know that $\rho_1 = \alpha\lambda b_1$ and $\rho_2 = (1-\alpha)\lambda b_2$ are just the utilization of each priority class, whereas $\pi_0 = 1 - \rho = 1 - \rho_1 - \rho_2$ is again the probability of finding an empty system. Equations (13) and (14) point out the well-known result that an increase in the average arrival rate of the high-priority group increases the average delay of both classes; similarly, an increase in the variance of the low-priority service times increases the average delay of both groups.

Average delay for the system of both priority types is just the average of Eqs. (13) and (14) when weighted by the fractional arrival rates α and $1 - \alpha$:

$$w = \alpha w_1 + (1 - \alpha)w_2 \tag{15}$$

We can compare the priority system with the nonpriority system by comparing Eq. (15) with the simple expression we have already used in Eq. (9). We find that the *priority system* has smaller average delays if

$$\left[\frac{\alpha}{1 - \rho_1} + \frac{1 - \alpha}{(1 - \rho_1)(1 - \rho)}\right] < \frac{1}{1 - \rho}$$

or, alternatively, if

$$b_2 > b_1 \tag{16}$$

In other words, the average service time of the low-priority group must be greater than the average service time of the high-priority group if the average delay of the priority system is to be less than that of the system without priorities. It is interesting that we do not have to make any assumptions about the shape of the distribution of service time to reach this conclusion.

This simple analysis extends easily to these cases where we are interested in minimizing linear functions of the average delay and costs of delay; we might, for example, be interested in minimizing expected delay costs of a terminal system where it was assumed that airborne aircraft were several times as expensive to operate as aircraft awaiting departure. Cox and Smith[8] have discussed this problem in their simple and valuable book on queuing theory. Pestalozzi[27] has applied their analysis to several priority systems, including departing and landing aircraft. He shows that, by organizing priority classes on the basis of aircraft performance, it is possible to reduce average costs of terminal operations by a large factor without violating the accepted rule of performance to landings over departures. More recently, Little[20] has derived a general proof and an efficient computational algorithm based on linear programming techniques to find optimal priorities when average arrival rates are Poisson and when service times are independent of each other.

V. Optimal Priority Classes and Delays

Up to this point, we have found only an inequality statement that must hold if average delays of a priority system are to be less than those of a corresponding nonpriority system. As we have already mentioned, the air traffic control system is fortunate that service times can be identified with aircraft. One can therefore organize priorities on the basis of aircraft in queue whose service times can be divided into two groups: those with service times less than x (high priority) and those with service greater than x (low priority). If service times of all aircraft are sampled from a known distribution, the obvious question is, what is the best choice for x?

As before, we assume that service times are sampled from a cumulative distribution function $B(t)$ and substitute

$$\alpha = B(x) \qquad 1 - \alpha = 1 - B(x)$$

$$\rho_1(x) = \lambda \int_0^x tb(t)dt \tag{17}$$

$$\rho_2(x) = \lambda \int_x^\infty tb(t)dt$$

into Eqs. (13–15) to get the average delay of the priority system. The system utilization ρ and the second moment $b^{(2)}$ are independent of the choice of x, and the final result is

$$w(x) = \frac{\lambda b^{(2)}}{2} \frac{[1 - \rho B(x)]}{(1 - \rho)[1 - \rho_1(x)]} \tag{18}$$

The smallest average system delay is obtained by setting the derivative of $w(x)$ equal to zero. The result is an integral equation for x^*, the optimum value of x,

$$1 = (x^*/b)[1 + \lambda B(x^*)(x^* - b_1)]^{-1} \tag{19}$$

The solution of this equation has already been discussed by Cox and Smith[8] for the special case where $B(t) = 1 - e^{-\mu t}$, i.e., service times are exponentially distributed. Figure 1 is a plot of x^* vs the system utilization, i.e., the ratio of average arrival rate to average service capacity. The solutions correspond to the case where service times have an Erlang distribution of order two or four, service capacity is μ, and utilization is ρ. The graph shows that the optimum division occurs later and later as the arrival rate or the utilization of the system increases. Hence, a larger fraction of all aircraft is included in the top-priority system as the arrival rate approaches the theoretical service capacity of the system. Once the solution for x^* has been found, the minimum average delay can be written in the form

$$w(x^*) = \frac{1}{2} \cdot \frac{\rho}{1 - \rho} \cdot \frac{b^{(2)}}{x^*} \tag{20}$$

Although $w(x^*)$ always becomes large as the average arrival

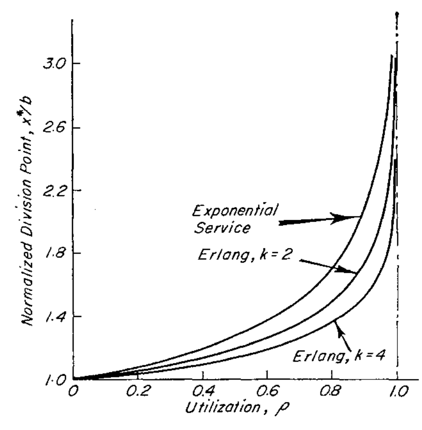

Fig. 1 Optimum division x^* of two priority classes.

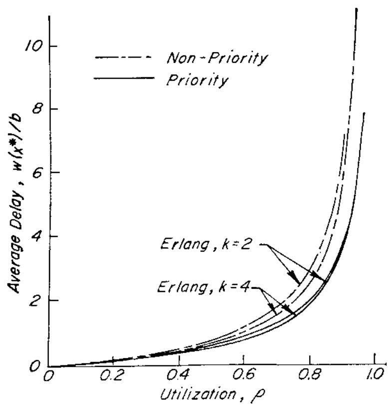

Fig. 2 Minimum average delay $w(x^*)$.

rate approaches the system capacity, it does so less rapidly than the nonpriority system or even any two-class priority system where $x \neq x^*$. Figure 2 is a plot of the (normalized) average delay for both priority and nonpriority systems. Even for values of medium utilization, the reduction in average delay may be of the order of 20 to 25%. An interesting property of this system is that the optimum division always occurs at a point greater than the average service time of all aircraft. Since we know that

$$w(x^*) = \frac{1}{2}\frac{b^{(2)}}{x^*}\frac{\rho}{1-\rho} \leqq \frac{1}{2}\frac{\lambda b^{(2)}}{1-\rho}$$

we must have

$$x^* \geqq b = \int_0^\infty tb(t)dt \qquad (21)$$

The procedure for organizing several priority classes follows easily from this example. If we introduce N priority classes, we can expect to find still further reductions in average system delay. We specify N points of separation such that the average system delay is made up of the weighted sum of $N + 1$ terms, each one being the delay of a particular priority class. The optimal location of the priority classes for the lowest expected delay costs are discussed in a paper by Oliver.[25]

As the number of priority classes increases, we can expect to find each aircraft in its own priority class. That is to say, first attention is given to the aircraft in queue which has the lowest service time. In this case, the limiting value of average system delay becomes

$$w = \frac{\lambda b^{(2)}}{2}\int_0^\infty a(t)b(t)dt \qquad (22)$$

where

$$a(t) = \left[1 - tB(t) - \int_0^t B(t)dt\right]^{-2}$$

This model is the so-called infinite priority system; it has been discussed by Phipps.[28]

A terminal system may not be able to control or want to delay more than several priority classes. Although the exact operational procedures suggested by Eq. (22) may be impractical, the general rules for organizing priorities that minimize linear functions of average delay are clear.

VI. Priorities with Mixed Operations

The effect of instrument flight rules on landing and departing aircraft is an important factor in many airport servicing systems. The queuing models that we have discussed in Secs. IV and V allow the service of several traffic streams at one facility, say, the runway or the glide path. However, in all of these models we have made the assumption that, whenever the queue is not idle, a customer from the highest available priority class is selected for service at the instant that another service is completed. If a high-priority unit arrives in the system at an instant when a lower-priority unit is already in service, we have always allowed the latter to complete its service. In these cases, the high-priority arrival has been delayed.

These assumptions can be modified in two ways. In the first place, the landing and departing aircraft have to maintain minimum separation standards between one another; hence, the facility may be busy but not actually servicing an airplane. In the second place, the lower-priority class may be denied use of the service facilities if it appears that a higher-priority unit will arrive during its service time. It has been quite common, for reasons that we have outlined in Secs. IV and V, to give priority to landing over departing aircraft. Radar detection and modern communication systems make it possible to calculate the arrival times of the landings over the runway threshold; hence, it is possible to foresee those cases where a landing aircraft and a departure might violate the dual occupancy rule mentioned in the introduction. In today's operations, these conflicts are resolved in favor of the landing aircraft by not allowing the interposition of a departure until minimum separation clearances to preceding as well as following landings are established.

The operating procedures mentioned in the preceding paragraph introduce new complications in the queuing models. Until recently, the problem remained unsolved; even now, explicit solutions are only available for the case of Poisson arrivals for landings and departures. In a slightly different context, namely, that of road traffic when a minor (departure) traffic stream intersects a major (landing) stream, the distribution of delays to a single random arrival in the minor stream has been thoroughly studied. One of the models of this process divides the landing stream into blocked and unblocked periods.

Figure 3 is a schematic diagram of these blocked periods when departures are forbidden use of a runway. Arrival times over runway threshold are marked by τ, with a subscript to denote the number of the airplane. Blocked periods correspond to those points in time when departures can be interposed between landings. The possible interposition of departures is shown by the regularly spaced dots lying between blocked periods; each dot is separated by a constant spacing Δ_d.

If a departure arrives during a block, his wait for a gap is at least as long as the remaining length of the block in progress; on the other hand, if the departure is fortunate enough to arrive during a gap in the stream, his wait for interposition will be zero if there are no departures in queue ahead of him and a positive number if there are one or more in the departure queue. In each case, the formula for average delay differs from Eqs. (9, 13, and 14) for nonpreemptive queuing models.

The criterion for deciding whether or not there is a blocked period may be complicated; hence, its construction and analysis may also be difficult. If one can assume that landing airplanes do not queue for service at the runway, the blocked periods are constructed by collecting all of those

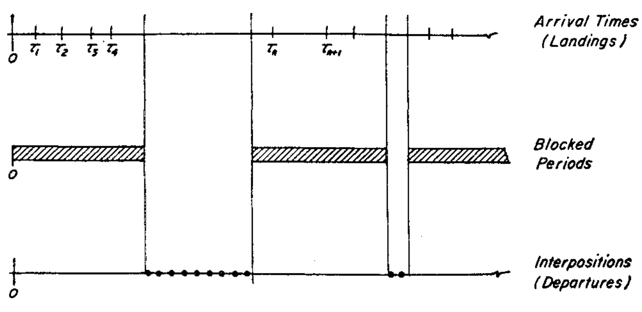

Fig. 3 Arrival times, blocked periods, and interpositions.

points in time which are closer than a constant separation Δ_{dl} to a following landing aircraft and Δ_{ld} from a preceding landing. This construction is shown in Fig. 3, where the ratio of Δ_{dl} to Δ_{ld} is approximately 3 to 5.

The probability that a single random departure does not have to wait is the probability that he lands within an unblocked period. In the case of Poisson landings, this probability, $e^{-\lambda_l\Delta}$, is simply the probability that the spacing between two landing aircraft exceeds $\Delta = \Delta_{dl} + \Delta_{ld}$.

In general, this single departure has to wait a random time. It has been shown[40, 36] that the probability that this wait exceeds t is

$$1 - W_d(t) = \sum_{i=0}^{j-1} (-1)^i e^{-(i+1)\lambda_l\Delta} \times \left\{\frac{[\lambda_l(t - i\Delta)]^i}{i!} + \frac{[\lambda_l(t - i\Delta)]^{i+1}}{(i+1)!}\right\} \qquad j\Delta \leqq t \leqq (j+1)\Delta \quad (23)$$

with average value equal to

$$w_d = \int_0^\infty t\, dW_d(t) = \lambda_l^{-1}e^{\lambda_l\Delta} - \Delta - \lambda_l^{-1}$$

Again we mention that the subscripts l and d refer to landings and departures, respectively.

If the queues of more than one aircraft develop in the departure stream, expressions for departure delays are further complicated. Gaver[13] has obtained general time-dependent formulations for the queue length and probability distribution of wait for service, and Galliher[39] and Oliver[24] have studied specific airport applications where landings have complete priority over departures. Approximations for certain high-flow-rate cases have also been obtained by Oliver and Bisbee[23] and by Jewell[16]; more recently, Yeo[41] has obtained solutions for a general class of these merging models.

In general, departure interposition rules that maintain minimum spacings to adjacent landings also come under the heading of "preemptive priority repeat" queuing models. That is to say, one can consider a fictitious operation where departures are interposed between landings only to find on certain occasions that their service is interrupted by the appearance of a landing. Hence, their actual runway occupancy times must be extended by assuming that they return to the departure queue to wait another interposition.

In Fig. 3 we see how blocked and unblocked periods alternate in the stream of landing aircraft. During the blocked periods, departures cannot be interposed between landings; during the unblocked periods, they can be released at constant separations that are specified by traffic control rules. An important characteristic of these models is that the capacity service rate of departures is not only a function of the departing aircraft but is also sensitive to the arrival rate of landing aircraft.

If we assume that the minimum separation between departing aircraft (in the absence of landing aircraft) also equals Δ, the capacity service rate of departure is

$$\mu_d = \lambda_l e^{-\lambda_l\Delta}/(1 - e^{-\lambda_l\Delta}) \qquad (24)$$

The numerator of this expression is simply the product of the average arrival rate of landing aircraft times the probability that the spacing between two landings exceeds the minimum value Δ. We multiply this product by a factor (>1) that allows multiple interpositions between those landings whose separations are greater than Δ. The expression for average delay to departures turns out to be

$$w_d = \frac{\lambda_l + \lambda_d}{\lambda_l\lambda_d} \cdot \frac{e^{\lambda_l\Delta} - \lambda_l\Delta - 1}{1 + \lambda_l\lambda_d^{-1} - e^{\lambda_l\Delta}} \qquad (25)$$

As the average arrival rate of departures approaches the service capacity in Eq. (24), the denominator of Eq. (25) goes to zero and w_d becomes large.

VII. Self-Clearing Rules

In addition to the priority structures that have already been discussed, there are additional operating rules for mixed operations which increase capacity and reduce expected delays or delay costs; unfortunately, the mathematical models are also more difficult. We owe thanks to a group of pioneers in the theory of road traffic, notably Tanner,[34] for studying the case where two priority classes vie for use of a common service facility, the first occupant of the facility representing the highest-priority class until all customers in that class are serviced. If, for example, a landing airplane finds the runway free, the runway continues to service landings until the queue of landings is exhausted. If any departing aircraft arrive during the period when landings have been serviced, they and all others who arrive during their own service periods are also serviced. In other words, the departure queue is serviced until it returns to zero. The runway again reverts to landing operations and so on until, at some instant, there are neither departures nor landings demanding service. At this point, the facility becomes idle; the next arrival—a landing or departure—determines the highest priority, and the process is repeated until the facility again becomes idle. In this operation, we can think of two blocked periods, one identified with a run of landings and the second with a run of departures.

Blumstein[2] has shown that the over-all service capacity of the runway increases as a result of this type of operation; average delays for the terminal system are reduced not only because of these capacity increases but also because large reductions in average delays of one aircraft type may be exchanged for small increases in average delay of a higher-priority group.

Assume the case where both landing and departing aircraft are Poisson. If only one landing begins a blocked period, then its length is that of a busy period; all landings arriving during the service of the first and subsequent arrivals must also be serviced. If no departures arrive during the landing block, operations revert to an idle period. If only one departure arrives during the entire landing block, the length of the ensuing departure block is also the length of a busy period for a departures-only service system. If more than one departure arrives, the departure block is an integral number of busy periods. Either an idle period or a landing block follows the departure block, the former appearing in the case where no landings arrive and the latter when there are one or more landings.

Using our earlier notation λ_l and λ_d for average Poisson arrival rates, Δ_l and Δ_d for their respective constant service times, and ρ_l and ρ_d for utilizations, it is easy to show that

$$\begin{aligned} Pr\,\{\text{idle runway}\} &= 1 - \lambda_d\Delta_d - \lambda_l\Delta_l = 1 - \rho_l - \rho_d \\ Pr\,\{\text{landing block}\} &= \lambda_l\Delta_l = \rho_l \\ Pr\,\{\text{departure block}\} &= \lambda_d\Delta_d = \rho_d \end{aligned} \qquad (26)$$

The delay in queue will be zero for a landing or departure that arrives during an idle period. If a landing arrives during a landing block, he must wait for service of all of those landings that entered the system before him; if he arrives during a departure block, he must wait for the service of all departures as well as landings ahead of him. From Tanner's results,[34] it can be shown that the average delay to landings is

$$w_l = \frac{1}{1 - \rho_l - \rho_d - \rho_l\rho_d} \times \left[\frac{(1 - \rho_l)(\rho_l\Delta_l + \rho_d\Delta_d)}{2(1 - \rho_l - \rho_d)} - \rho_l\rho_d\Delta_l\right] \qquad (27)$$

with a symmetric expression for w_d which is obtained by interchanging the role of the subscripts l and d. Tanner has studied the effect of this priority rule under a variety of constant-service assumptions that include, among others, the requirement that the service facility be cleared of both types of aircraft for a fixed period of time in order to allow the safe transition from a landing to a departure block.

Equations for the distribution of delays suggest that extremely long delays may occur for either group. Although it is not clear whether long runs of alternating types of aircraft operations can be allowed in practice, we conclude that average delays for the entire system can be reduced by the organization of rules for runway and glide-path usage; the extent to which any of these rules can be implemented will depend upon the costs of delays, safe operating procedures, and other factors.

VIII. Summary

Queuing models of airport capacity and delay problems give strong evidence that more regularly scheduled arrival and service times are desirable. However, given certain random characteristics of the terminal area, such as Poisson arrivals and unpredictable service times, it is also clear that average delays or costs of the system can be reduced by appropriate organization of priority classes. These priority classes should be based upon the service characteristics of particular aircraft involved in the terminal area and may not be as simple as the rule: priority to landings over departures. It may be desirable to feed one runway with several landing stacks or several queues of departing aircraft. It has also been shown that optimal rules of priority organization can be found to satisfy certain simple criteria based on average delay or cost. These rules and the expressions for expected delays are simple to compute; hence, the operating characteristics of a complex system can, to a certain extent, be predicted before it is put into actual operation.

It has also been shown that the major effect of including minimum separations between aircraft in the terminal area or the use of certain preemptive priority models for departing aircraft is to increase greatly the average and extreme delays of the low-priority class, i.e., the departing aircraft in today's operations. The advantages and disadvantages of self-clearing rules have also been discussed; there is reason to believe that they may be optimal under certain traffic conditions.

Expressions for average delays and the probability that delays exceed a given value are sensitive to the utilization of the service facility, i.e., the ratio of the arrival rate to the theoretical capacity service rate. Capacity service rates can usually be calculated by assuming those conditions where aircraft are separated by minimum clearances on the runway or glide path. Hence, it should be pointed out that present operations that do not make use of automatic computation techniques to calculate the arrival times of landing aircraft in the glide path or over the runway threshold, the priority classifications, or the construction of blocked periods in the landing stream either violate or exaggerate minimum separations in order to be "on the safe side." Inevitably, the result is either an unsafe operating condition or a reduction in capacity and an increase in delays and costs.

References

[1] Blumstein, A., "Landing capacity of a runway," Operations Res. **7,** 752–763 (1959).

[2] Blumstein, A., "An analytical investigation of airport capacity," Cornell Aeronautical Lab. Rept. TA-1358-G-1 (1960).

[3] Blumstein, A., "The operations capacity of a runway used for landings and take-offs," *Proceedings of the Second International Conference on Operations Research* (John Wiley & Sons, Inc., New York, 1961).

[4] Borel, E., "Sur l'emploi du theoreme de Bernoulli pour faciliter le calcul d'un infinite de coefficients," Compt. Rend. **214,** 452–456 (1942).

[5] Brockmeyer, E., Halström, H. L., and Jensen, A., "The life and works of A. K. Erlang," Trans. Danish Acad. Tech. Sci., no. 2, 1–274 (1948).

[6] Burke, P. J., "Equilibrium delay distribution for one channel with constant holding time, Poisson input and random service," Bell System Tech. J. **38,** 1021–1031 (1959).

[7] Cobham, A., "Priority assignment in waiting line problems," Operations Res. **2,** 70–76 (1954).

[8] Cox, D. R. and Smith, W. L., "Queues," *Methuen Monograph* (John Wiley & Sons, Inc., New York, 1961), pp. 1–180.

[9] Crommelin, C. D., "Delay probability formulae when the holding times are constant," Post Office Elec. Eng. **J. 25,** 41–50 (1932).

[10] Crommelin, C. D., "Delay probability formulae," Post Office Elec. Eng. J. **26,** 266–274 (1933).

[11] Feller, W., *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1957), 2nd ed., Vol. 1.

[12] Fry, T. C., *Probability and Its Engineering Uses* (D. Van Nostrand Co., Inc., New York, 1928), Chap. X.

[13] Gaver, D. P., "A waiting line with interrupted service, including priorities," J. Roy. Statist. Soc. **24B,** 73–90 (1962).

[14] Haight, F. A., "The Borel-Tanner distribution," Biometrika **47,** 143–150 (1960).

[15] Horonjeff, R., Read, R., and Ahlborn, G., "Exit taxiway locations," Special Rept. conducted for the Research Div., Bureau of Research and Development, Federal Aviation Agency, Berkeley, Institute of Transportation and Traffic Engineering, Univ. of California (September 1960).

[16] Jewell, W. S., "Multiple entries in traffic," J. Soc. Ind. Appl. Math. **11,** 872–885 (1963).

[17] Kendall, D. G., "Some problems in the theory of queues," J. Roy. Statist. Soc. **13B,** 151–185 (1951).

[18] Khintchine, A. Y., *Mathematical Models in the Theory of Queuing* (Hafner, New York, 1960).

[19] Lindley, D. V., "The theory of queues with a single server," Proc. Cambridge Phil. Soc. **48,** 277–289 (1952).

[20] Little, J. D. C., "A Markov decision process in a single channel queue with Poisson arrivals," mimeographed notes (1962).

[21] Morse, P. M., *Queues, Inventories and Maintenance* (John Wiley & Sons, Inc., New York, 1958).

[22] Oliver, R. M., "Delays to aircraft serviced by the glide path," Operations Res. Quart. **13,** 201–209 (June 1962).

[23] Oliver, R. M. and Bisbee, E. F., "Queuing for gaps in high-flow traffic streams," Operations Res. **10,** 105–114 (1962).

[24] Oliver, R. M., "Airport runway capacity and delays—I," Operations Research Center, Univ. of California, Res. Rept. 25 (August 1962).

[25] Oliver, R. M., "On a problem of optimum priority organization," Operations Research Center, Univ. of California, Res. Rept. 63-29RR (December 30, 1963); to appear Soc. Ind. Appl. Math (1964).

[26] Pearcey, T., "Delays in landing of air traffic," J. Roy. Aeronaut. Soc. **52,** 799–812 (1948).

[27] Pestalozzi, G., "Priority rules for runway use," Transportation Engineering Graduate Rept., Dept. of Civil Engineering, Univ. of California (1962).

[28] Phipps, T. E., "Machine-repair as a priority waiting line problem," Operations Res. **4,** 76–85 (1956).

[29] Pollaczek, F., "Über eine Aufgabe der Wahrscheinlichkeitstheorie," Math. Z. **32,** 64–100, 729–750 (1930).

[30] Pollaczek, F., "Uber das Wateproblem," Math. Z. **38,** 492–537 (1934).

[31] U. S. Federal Aviation Agency, *Project Beacon: Report of the Task Force on Air Traffic Control* (U. S. Government Printing Office, Washington, D. C., 1961).

[32] Read, R. R. and Yoshikawa, R., "Operations analysis of locating high-speed runway exits," Operations Res. **10,** 1–18 (1962).

[33] Riordan, J., *Stochastic Service Systems* (John Wiley & Sons, Inc., New York, 1962).

[34] Tanner, J. C., "A problem of interference between two queues," Biometrika **40,** 58–69 (1953).

[35] Tanner, J. C., "A theoretical analysis of delays at an uncontrolled intersection," Biometrika **49,** 163–170 (1962).

[36] Weiss, G. H. and Maradudin, A., "Some problems in traffic delay," Operations Res. **10**, 74–114 (1962).

[37] Wishart, D. M. G., "Queuing systems in which the discipline is 'last-come, first-served'," Operations Res. **8**, 591–599 (1960).

[38] Oliver, R. M., "Tables of the waiting time distribution of the constant service queue (M/D/1)," Operations Research Center, Univ. of California, Res. Rept. ORC 63-12 (September 1963).

[39] Galliher, H. P., "Airport runway traffic design and delays," Transportation Engineering Conference, American Society of Civil Engineers, Detroit, Mich. (October 1962).

[40] Raff, M. S., "The distribution of blocks in an uncongested stream of automobile traffic," J. Am. Statist. Assoc. **46**, 114 (1951).

[41] Yeo, G., "Single-server queues with modified service mechanisms," J. Australian Math. Soc. II, 499–507 (1962).

MAY-JUNE 1964 J. AIRCRAFT VOL. 1, NO. 3

Inertial Navigator Flight-Testing Experience with the Lockheed F-104

D. R. MEYER* AND K. E. HODGE†
Lockheed Aircraft Corporation, Burbank, Calif.

In this paper a description is given of the flight profiles and courses flown, and the technique used for the reduction of data, during flight testing of a lightweight inertial navigator installed in the F-104 interceptor. Program objectives were to evaluate accuracy of navigator outputs and determine system reliability. Equipment requirements for system accuracy were established in terms of a statistical parameter, circular error probable. The major program effort was centered about the flight testing of inertial navigator systems in five F-104 airplanes. Navigator distance outputs were checked photographically against an independent reference, points on the earth's surface; a simple data-correction technique was used which obviated need for stabilization of the downward-looking camera. Problems were discovered in several areas. One of these, runaway distance outputs because of condensed moisture, was revealed only by flight tests. Because of the maneuverability of the F-104 and its extensive flight envelope, it was necessary to conduct inertial navigator tests under a wide variety of conditions.

Introduction

A REVIEW is presented of Lockheed experience during flight-test evaluation of a lightweight inertial navigator installed in the F-104 Super Starfighter airplane. This airplane, a derivative of the USAF "Starfighter," is capable of Mach 2 flight, interception of targets at high altitudes, and low-level ground attacks. The general arrangement of the F-104 is shown in Fig. 1. Highly maneuverable, the F-104 provides strenuous environmental conditions for all airplane systems, including avionics.

Inertial navigator tests commenced in a DC-3 electronics system test bed in July 1960. The first quantitative inertial navigator data from an F-104 systems installation were obtained in January 1961. Final data flights were conducted in August 1962. With the exception of high-humidity tests made from Eglin Air Force Base, Fla., all Lockheed-conducted inertial navigator flights were made from USAF Plant 42, Palmdale, Calif. The program objective was to determine whether the inertial navigation system installation was satisfactory, i.e., to insure that the accuracy of navigator and platform outputs was consistent with design objectives and that system reliability was sufficient to assure a high probability of mission success.

Description of System

The LN-3 inertial navigation system, manufactured by Litton Industries, was specified for the F-104 to provide accurate self-contained navigation information and roll angle, pitch angle, and vertical acceleration data from transducers mounted on the navigator stable platform. The distance and heading readout is provided by a Position and Homing Indicator (PHI) System, which is also used to display Tactical Air Navigation (TACAN) and dead-reckoning navigation computer output data. The PHI is manufactured by Computing Devices of Canada, Ltd. and features a 12-station selector unit (SSU) to facilitate navigation to preselected destinations.

The inertial navigator is a lightweight, fully automatic, self-contained navigation system. The heart of the system is a four-gimbal platform assembly aligned to the local vertical.

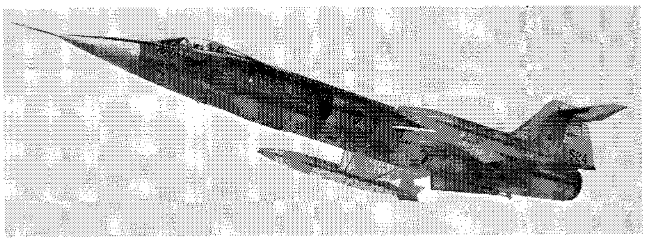

Fig. 1 Lockheed F-104 general arrangement.

This gimbal assembly mounts two floated gyros arranged in a dumbell configuration and three single-axis torque-balanced accelerometers with mutually perpendicular axes. Accelerations measured along the two axes tangent to the earth are used for navigation, whereas that measured along the local vertical is used by the autopilot for an inertial damping signal in the "altitude hold" mode.

Successive integrations are required to obtain ground velocity and position. When provided with manual inputs of initial and target or base position, the inertial navigator in conjunction with the PHI system continuously determines the position of the aircraft and indicates to the pilot the head-

Presented at the AIAA/AFFTC/NASA FRC Testing of Manned Flight Systems Conference, Edwards Air Force Base, Calif., December 4–6, 1963 (no preprint number; published in bound volume of preprints of the meeting); revision received March 17, 1964.

* Department Manager, Electronic Systems, Flight Test Division.

† Group Engineer, Systems Analysis, Flight Test Division. Member AIAA.