NPS-54-88-007

NAVAL POSTGRADUATE SCHOOL
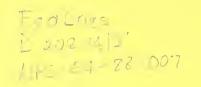## Monterey, California

SCHEDULING FLIGHTS AT HUB AIRPORTS

DAN TRIETSCH

June 1988

Approved for public release; distribution unlimited.

Prepared for: Naval Postgraduate School
Monterey, CA 93943-5100

NAVAL POSTGRADUATE SCHOOL
Monterey, California

RADM. R. C. Austin   Kneale T. Marshall
Superintendent   Acting Provost

Reproduction of all or part of this report is authorised.

This report was prepared by:

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
UNCLASSIFIED	

SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
DECLASSIFICATION/DOWNGRADING SCHEDULE	Approved for public release; distribution is unlimited.

PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
NPS-54-88-007	

NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
Naval Postgraduate School		

ADDRESS (City, State, and ZIP Code)	7b ADDRESS (City, State, and ZIP Code)
Monterey, CA 93943-5000	

NAME OF FUNDING/SPONSORING ORGANIZATION NPS Foundation Research Program	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

ADDRESS (City, State and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO	PROJECT NO	TASK NO.	WORK UNIT ACCESSION NO
Monterey, CA 93943-5000				

TITLE (Include Security Classification)

Scheduling Flights at Hub Airports (UNCLASSIFIED)

PERSONAL AUTHOR(S)
Dan Trietsch

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final Report	FROM _____ TO _____	June 1988	32

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Hub Operations; Airline Scheduling; Stochastic Arrivals

ABSTRACT (Continue on reverse if necessary and identify by block number)

In a typical hub airport, incoming flights from many origins feed outgoing flights to many destinations. If an incoming flight is late, outgoing flights which are fed by it may also be delayed eventually. Alternately, planes may leave before some feeding flights arrive, thereby incurring high misconnection penalties. Clearly, if we plan for very long scheduled ground time between the last incoming flight and the first outgoing one, we can reduce the risk of unscheduled delays or misconnections. However, such a schedule may cost the airline too much in terms of idle personnel and equipment and will not be attractive to the passenger either. On the other hand, if we plan for very short scheduled ground time, we run the risk of excessive unscheduled delays, and/or misconnection penalties. In this paper we develop models designed to optimize the scheduled ground time under two pure policies: (i) to wait as long as necessary to ensure all connections,

DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED

NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dan Trietsch	408-646-2456	54Tr

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete.

UNCLASSIFIED

19. and (ii) not to wait at all (i.e., pay misconnection penalties rather than delay penalties). The models can also be applied to similar problems such as express parcel deliveries and ground transportation hubs.

SCHEDULING FLIGHTS AT HUB AIRPORTS

by

Dan Trietsch[*]

June 1988

----

* Department of Administrative Sciences, Naval Postgraduate School, Monterey, CA 93943

SCHEDULING FLIGHTS AT HUB AIRPORTS

Abstract

In a typical hub airport, incoming flights from many origins feed outgoing flights to many destinations. If an incoming flight is late, outgoing flights which are fed by it may also be delayed eventually. Alternately, planes may leave before some feeding flights arrive, thereby incurring high misconnection penalties. Clearly, if we plan for very long scheduled ground time between the last incoming flight and the first outgoing one, we can reduce the risk of unscheduled delays or misconnections. However, such a schedule may cost the airline too much in terms of idle personnel and equipment and will not be attractive to the passengers either. On the other hand, if we plan for very short scheduled ground time, we run the risk of excessive unscheduled delays, and/or misconnection penalties. In this paper we develop models designed to optimize the scheduled ground time under two pure policies: (i) to wait as long as necessary to ensure all connections, and (ii) not to wait at all (i.e., pay misconnection penalties rather than delay penalties). The models can also be applied to similar problems such as express parcel deliveries and ground transportation hubs.

Introduction:

Our concern in this paper is scheduling connections at a hub facility, where the timing of arrivals and departures is subject to stochastic variation. It is presented in terms of passenger air service, but can also be applied to various other hub networks operations; e.g., express parcel deliveries, central warehouses break-bulk/consolidation operations etc.

The basic premise is that the arrivals at the hub feed the departures, and hence the scheduling of the arrivals and of the departures has to be coordinated. A complicating factor is that the arrivals may run behind schedule--an issue which is usually not taken into account explicitly in airline scheduling models. In fact, the arrival time is a stochastic variable influenced by a multitude of conditions, both in and out of the airline's control. For instance, weather conditions at the hub itself and in the other airports may impact the on-time performance. Similarly, when many planes are scheduled to land or take off within an interval of two or three minutes then even under perfect weather conditions some stochastic queueing delays will certainly result.

We present two basic models: (i) where all departures which need to be fed by late arrivals wait as long as necessary--and thus may get delayed as well and incur a lateness penalty, and (ii) where departures never wait, and instead there is a misconnection penalty. The objective in both cases is to minimize the total expected time and penalty expenditure.

In both cases we assume that the arrival time delays are independent of each other. In reality, this assumption is not totally justifiable, since conditions at the hub may impact all arrivals, and since the weather across several airports may be correlated. However, we need the assumption to achieve tractability. In practice, if we have adverse weather conditions at the hub, we

may have to resort to a revised schedule which allows more time between flights, and at the same time accept more misconnections. On the other hand, if the problems at the hub are due to conjestion[1, 6, 9] we may expect some degree of dependence among the delays, but we do present a scheduling policy which tends to minimize this effect by specifying gaps between adjacent arrivals or departures.

The literature on airport scheduling to date is concerned mainly with deterministic models, and is thus not directly relevant for us here. For instance, see Etschmaier and Mathaisel[3]. A stochastic model similar to ours was presented by Hall[4], and is concerned mainly with transit passengers, where several feeder lines serve a single train. The policy adopted is no-waiting, assuming there will be a later train. The objective is to minimize the total expected time in the system. The focus is on the exponential distribution, and the results are in terms of this special case.

The first model in this paper--where we assume departures wait for all connecting passengers--is an extension of a project purchase order scheduling model developed by Ronen and Trietsch[7, 8], which in turn is a generalization of the classic newsboy problem. The project purchase order model is concerned with ordering stochastic lead time project items just in time to minimize the sum of their holding cost and the expected project lateness penalties incurred if any items are late. The project items are analog to our feeder flights, and the project lateness is analog to departure delays. However, in the project problem we feed just one project, while in the hub connection problem we feed several departures. In addition, there is no analog there to the conjestion issue. Thus the model requires considerable modification for our purpose.

In Section 1, we briefly describe the main results of [8], as a basis for our model here. Section 2 introduces our own first model formally. Section 3

gives an approximate solution when the arrival and departure times (our decision variables) are not constrained. This solution calls for almost simultaneous departures (but not necessarily simultaneous arrivals). Since such a practice would lead to certain queueing delays in the departures, its applicability is questionable. Therefore, in Section 4 we present a solution for a case where the arrivals and departures are constrained to be staggered by a specified amount (say three minutes), and we look for the optimal scheduled ground time between the last arrival and the first departure. Section 5 is devoted to the second model where departures take place regardless of pending arrivals, and late passengers are accommodated by other flights, at a fixed penalty cost. In Section 6, the conclusion, we discuss the interaction of our models with other operational decisions and with each other. Most of the issues in Section 6 require further research at this stage, but cannot go unmentioned here. Four of the points we discuss are (i) how to use the results from the two pure models to support a more balanced policy which prevents long waiting and excessive misconnections at the same time; (ii) the mutual influence between the speed choice of the aircraft and the scheduling problem; (iii) the impact of our solution on the gate assignment problem; and (iv) how to assign aircraft to flight segments in and out of the hub so as to minimize the expected misconnection penalties.

1. The Generalized Newsboy Model:

Consider the following problem: The completion of a project hinges upon receiving all the purchased components by the time they are scheduled to be used. Ordering too soon will cause excessive inventory holding costs, while ordering too late is likely to delay the whole project, and thus incur the project delay penalty cost. In order to minimize the total expected cost of the project, each of the orders has to be scheduled optimally.

If there is only one purchased component, the optimal ordering time is calculated by an almost direct application of the newsboy model, in such a manner that the probability of delay multiplied by the delay penalty will equal the complementary probability multiplied by the inventory holding cost of the component per time unit. When more than one item is involved, the problem is more complex. On one hand, it is enough that one item will be delayed to delay the whole project, and thus incur the penalty cost <u>in addition to</u> the holding cost of those items which did arrive in time. Intuitively, this may push us to order even earlier than in the one item case. On the other hand, if several items are late we should not penalize each of them with the full delay cost, but rather have them "share" the burden. This in turn may push us to order later. It becomes quite clear that intuition alone is not likely to produce a good solution here. We introduce the following notation:

- P lateness penalty for the project ($/day)
- C_i holding cost for item i, i=1,...,n ($/day)
- t^* project due date. (For simplicity we assume that this is also the due date of the items. See [7, 8] for a more sophisticated version.)
- T_i time item i is ordered (our decision variable)
- T_i^* optimal value of T_i

- $F_i()$ cummulative distribution function (CDF) of item i's lead time. That is, $F_i(t - T_i) = Pr\{$item i will arrive before time t$\}$. Similarly, $F_i(t^* - T_i)$ is the probability item i will arrive in time. $F_i()$ is assumed stationary and independent of $F_j()$; $\forall$ $j \neq i$

- $f_i()$ density function of item i's lead time

- F_i^* $F_i(t^* - T_i^*)$, i.e., the optimal probability item i will be in time

- $\overline{F}_i$ Upper bound on F_i^*

- $\underline{F}_i$ Lower bound on F_i^*

- S $P + \sum C_i$. (S is defined for convenience in presenting the results.)

Then in order to minimize the total expected holding costs and project lateness penalty the following set of nonlinear equations must be satisfied by T_i (i=1,...,n)

$$C_i = S \int_{t^*}^{\infty} f_i(t-T_i) \prod_{j \neq i} F_j(t-T_j)dt \quad ; \quad i = 1,...., n. \tag{1.1}$$

We also have the following expressions for the bounds $\underline{F}_i$ and $\overline{F}_i$

$$\overline{F}_i = 1 - C_i/S \quad ; \quad i = 1, ... , n, \tag{1.2}$$

$$\underline{F}_i = 1 - C_i/(S \prod_{j \neq i} \underline{F}_j) \quad ; \quad i = 1, ... , n. \tag{1.3}$$

As shown in [8], if we define $A_i = C_i/S$ and $x = \prod \underline{F}_j$; j = 1,...,n, then

$$\underline{F}_i = x / (x + A_i), \tag{1.4}$$

and (1.3) has a valid solution iff the polynomial

$$g(x) = \prod_{i=1}^{n}(x + A_i) - x^{n-1} = 0 \tag{1.5}$$

has a positive root $x < 1$. Thus the solution of the set of equations (1.3) is reduced to a search over a single variable, x. Once x is found, (1.4) is used to obtain $\underline{F}_i$. If more than one such root exists, we pick the largest. Loosely speaking we may say that the bound (1.3) exists for relatively low values of C_i/S. Furthermore, for such low C_i/S values the upper bound and the lower bound tend to be close to each other.

Since $F_i()$ is monotone, and assumed given, it is straightforward to compute T_i^* given F_i^*. Similarly, given $\overline{F}_i$ we can easily find a lower bound for T_i^*, and given $\underline{F}_i$ an upper bound for T_i^*.

2. The Hub Scheduling Model:

Suppose we have n+1 origin/destination points indexed by i=0,1,...,n, where 0 is the index of the hub itself. We use the following notation

- c time unit value per average passenger

- p delay penalty per time unit per passenger ($p \geq c$)

- $D_{i,j}$ average satisfied demand from origin i to destination j

- S_i scheduled arrival time for flight (i,0)--a decision variable

- T_j scheduled departure time for flight (0,j)--a decision variable

- S_i^* optimal S_i value

- T_j^* optimal T_j value

- $F_i()$ CDF for delays in flight (i,0). Assumed independently distributed $\forall$ i

- $F_0()$ CDF for delays in takeoff for flight (0,j) <u>not</u> due to waiting for

incoming flights. Assumed identically distributed and independent $\forall$ j

- P_j $\sum_{i=0,n} pD_{i,j}$; j=1,....,n
- P $\sum_{j=1,n} P_j$
- C_i $\sum_{j=1,n} cD_{i,j}$; i=1,....,n
- d(i)   final destination of the aircraft assigned to (i,0)
- o(j)   origin of the aircraft assigned to (0,j), i.e., the inverse of d(i)
- b_i time unit value of the aircraft and crew assigned to the (i,0) segment

  (thus the value on the (0,j) segment is $b_{o(j)}$)

We assume that F_i is given, and that the departure of (i,0) from i is scheduled ahead of S_i by a prespecified period. A more sophisticated version is discussed in Section 6. Our objective function is to minimize the total scheduled time costs plus the total unscheduled delay penalties, i.e.,

$$\text{Min } Z = \sum_{j=1}^{n} b_{o(j)} T_j - \sum_{i=1}^{n} b_i S_i + \sum_{i=1}^{n} \sum_{j=1}^{n} cD_{i,j}(T_j - S_i) +$$

$$\sum_{j=1}^{n} (P_j + b_{o(j)}) \int_{T_j}^{\infty} (1 - \prod_{k=0}^{n} F_k(t - S_k)) dt + \sum_{i=1}^{n} pD_{i,0} \int_{0}^{\infty} (1 - F_i(t)) dt \quad \text{(2.1)}$$

s.t.

$$\min\{T_j\} \geq \text{Max}\{S_i\} \quad ; \quad i, j = 1,....,n \tag{2.2}$$

where $S_0 = T_j$; $\forall$ j. Note that the first part of the objective function is a deterministic cost associated with the scheduled gaps between arrivals and departures, while the second part adds stochastic penalties for unscheduled delays. In particular, there is no incentive to arrive ahead of schedule; such

early arrivals would be recorded as arrivals on time as far as the objective function is concerned. Also note that (2.2) implies that connections require zero time. In reality, it is necessary to add an appropriate constant minimal connection time to the model. A simple way to do so would be to redefine the scheduled departures as the model's T_j^* plus the constant. To determine the magnitude of this constant is a separate problem similar to our bigger model-- we need to allow enough time so that the probability the connection will be large enough to avoid a large expected misconnection penalty, yet not so large that the cost of the allotted time itself will become excessive.

To continue, assuming that Z is differentiable, and using the Leibnitz method, we obtain the following partial derivatives for (2.1)

$$\frac{\partial Z}{\partial S_i} = \sum_{j=1}^{n} [-cD_{i,j} + (P_j + b_{o(j)}) \int_{T_j}^{\infty} f_i(t-S_i) \prod_{\substack{k=0 \\ k \neq i}}^{n} F_k(t-S_k) \, dt] - b_i; \; \forall \, i, \quad (2.3)$$

$$\frac{\partial Z}{\partial T_j} = \sum_{i=1}^{n} cD_{i,j} - (P_j + b_{o(j)})(1 - \prod_{k=0}^{n} F_k(T_j-S_k)) + b_{o(j)} \; ; \; \forall \, j. \quad (2.4)$$

Though we have 2n equations for 2n decision variables, it is clear by observation of the objective function that if we add a constant to all the variables, Z will not change. Therefore, we can fix one of the variables to our convenience. Furthermore, due to (2.2), some S_i (or alternately T_j) values may have a binding constraint. In all other instances the partial derivatives should be set to zero. This can only be done numerically. Though (2.3) involves integrals, when we resort to numeric methods there is no conceptual difficulty in representing them by appropriate sums and still achieve any required degree of accuracy. (Of course it is much more convenient to avoid

having to deal with integrals if possible. Indeed, the solution we present in Section 4 does not involve direct computation of integrals.)

In the next two sections (3 and 4) we discuss two special cases, namely (i) if it is possible to schedule many flights to take off at the same time--a clear impossibility (though one can still encounter timetables which look as if this assumption was made); and (ii) when the schedule is constrained in such a manner that planes are scheduled to arrive and leave at fixed intervals, and we only need to determine the optimal interval between the last arrival and the first departure (above and beyond the minimal connection time).

3. The Uncapacitated Case--An Approximate Solution:

Consider the (imaginary) case where the airport itself has an infinite capacity to support any number of takeoffs/landings simultaneously. In this case it makes perfect sense to schedule all departures (and/or all arrivals) at the same time --if minimizing the objective function calls for it. We proceed to show that for the departures this is approximately the case. For this purpose we resort to a simplified case where the b_i values are assumed to be zero. The real expenses represented by b_i are assigned equally to all passengers and thus become part of c and p. Rewriting (2.4) for this case and rearranging the terms slightly we get

$$\frac{\delta Z}{\delta T_j} = P_j [\sum_{i=1}^{n} cD_{i,j}/P_j - (1 - \prod_{k=0}^{n} F_k(T_j - S_k))] \; ; \; \forall \; j. \tag{3.1}$$

Now observe the term $\sum cD_{i,j}/P_j$. By definition $P_j = \sum_{i=0,n} pD_{i,j}$; ignoring $D_{0,j}$, it can be approximated by $\sum_{i=1,n} pD_{i,j}$. But if we do this, then (3.1) becomes

- 9 -

$$\frac{\partial Z}{\partial T_j} = P_j[c/p - (1 - \prod_{k=0}^{n} F_k(T_j - S_k))] \; ; \; \forall j. \tag{3.2}$$

Since our objective is to find the value of T_j which drives (3.2) to zero, we may observe now that the same value of T will solve (3.2) $\forall$ j. Furthermore, we are allowed to set one variable arbitrarily, so we can state $T_j^* = t^*$; $\forall$ j!

The latter approximation will not be significant if $D_{0,j} / \sum_{i=1,n} D_{i,j}$ is small $\forall$ j. Even if $D_{0,j} / \sum_{i=1,n} D_{i,j}$ is not small, but is approximately constant $\forall$ j, we would just have to replace c/p with a smaller positive constant, and the major result--namely that $T_j^* = t^*$ $\forall$ j--will remain valid. Note also that (3.2) has a solution only if $F_k(T_j - S_k) > 0$; $\forall$ k, i.e., if constraint (2.2) is satisfied.

To continue, if we substitute $T_j^* = t^*$; $\forall$ j for T_j in (2.3), set $b_i = 0$ and open the brackets we obtain

$$\frac{\partial Z}{\partial S_i} = \sum_{j=1}^{n} -cD_{i,j} + \sum_{j=1}^{n} P_j \int_{t^*}^{\infty} f_i(t - S_i) \prod_{\substack{k=0 \\ k \neq i}}^{n} F_k(t - S_k) \, dt \; ; \; \forall i. \tag{3.3}$$

And by the definition of C_i and **P**, we can write it as

$$\frac{\partial Z}{\partial S_i} = -C_i + P \int_{t^*}^{\infty} f_i(t - S_i) \prod_{\substack{k=0 \\ k \neq i}}^{n} F_k(t - S_k) \, dt \; ; \; \forall i. \tag{3.4}$$

Note that if we substitute **P** by **S** (3.4) is almost identical to (1.1)--the only difference being that in (3.4) we have to take account of $F_0(t - t^*)$ which has no analog in (1.1). This implies that all the results stated in Section 1

- 10 -

for the project case, hold approximately for the hub scheduling problem.

Indeed, once we set $T_j^* = t^*$; $\forall$ j, we can speak of our problem in terms of a "project" which is to clear the departures at time t^*. The incoming flights are analog to the purchased materials. The only difference is that even if all incoming flights are in, it may still happen that some departures will be delayed due to specific problems relevant to them. Therefore, the expected penalty attributable to delays in incoming flights is slightly smaller--which is where F_0 comes in.

In particular, we may write the following analogs to the bounds (1.2) and (1.3) respectively

$$\overline{F}_i = 1 - C_i/P \quad ; \quad i = 1, \ldots , n, \tag{3.5}$$

$$\underline{F}_i = 1 - C_i/(P \prod_{j \neq i} \underline{F}_j) \quad ; \quad i = 1, \ldots, n, \; j = 0, 1, \ldots, n. \tag{3.6}$$

Though we have the term F_0 as part of (3.6), the solution is exactly as described in Section 1, with $A_i = C_i/P$ and x defined with j = 0,....,n (instead of j = 1,....,n).

The results in this section are approximate due to the fact that we allocate the aircraft's time value to the passengers, and ignore the fact that the fraction of passengers bound to the hub itself may be significantly different for some origins. However, since the basic assumption of infinite capacity is such a strong assumption, we do not require an exact solution for this case anyway. The major contribution of this section is the insight that the departures are likely to be close to each other in the uncapacitated case. Since this is a super-optimal solution, we can conclude that the gap between adjacent departures should be kept as low as possible.

4. The Capacitated Case:

Needless to say, airports do have limitations on their capacities. There-fore, the approximation presented in Section 3 cannot be seriously recommended. Still, some airlines tend to schedule many departures at practically the same time (e.g., 15 flights between 6:00PM to 6:05PM). It seems that the reason they do this is in order to list attractive departure times (i.e., departures which are likely to appear on the first screen monitored by travel agents). The results are long queueing delays, and an obvious decrease in the total system utility.

Fortunately there are indications that this practice will be curtailed in the near future. Indeed, airports today start to crack down on practices such as moving the plane a few feet from the gate to create an "on-time" departure even before the plane is allowed to taxi; they also do not allow planes to taxi when too many other planes are in line ahead. Such policies will force airlines which insist on scheduling under the infinite capacity assumption to have to report many late departures. Also, some airlines may realize that passengers spending 25 minutes or so in a takeoff queue, and observing that 90% of the planes in line are operated by their own carrier, cannot help but blame it for the delay, thus they are likely to avoid such practices even if they are not forced to do so officially.

A general finite capacity model for our problem might incorporate (2.1) and (2.2) as before, plus an additional set of constraints such as

$$|S_i - S_k| \geq \triangle_A \; ; \; \forall \, k \neq i, \tag{4.1}$$

$$|T_j - T_k| \geq \triangle_D \; ; \; \forall \, k \neq j, \tag{4.2}$$

where $\triangle_A$ and $\triangle_D$ specify minimal scheduled gaps between arrivals and departures respectively (we do not assume $\triangle_A = \triangle_D$). Unfortunately, such a model would be very difficult to implement, due to the large number of constraints. Therefore we proceed to look for simpler models. We propose a basic model which is very simple to compute, and an improvement based on it which requires more computation but is still quite tractable.

In both the bacic and the improved models we assume that the sequence of the arrivals and the sequence of the departures are determined separately (we'll discuss how in Section 6). This reduces the number of constraints considerably, and makes them simpler as well (no need to deal with absolute values). In the basic model we also assume that the scheduled gap between adjacent flights in each sequence is set to a constant; e.g., if flight $(k,0)$ immediately precedes flight $(i,0)$ then $S_i - S_k = \triangle_A$, and if flight $(0,k)$ immediately precedes flight $(0,j)$ then $T_j - T_k = \triangle_D$. The result is that we only have to find the optimal gap between $\text{Max}\{S_i\}$ and $\min\{T_j\}$, and all the other variables will be determined.

Note that the actual arrivals and departures will still be stochastic variables, distributed "around" the scheduled times. Also note that the actual sequence will not necessarily be identical to the planned sequence, especially if $\triangle$ is small relative to the standard deviation of the actual arrivals/ departures. Incidentally, if $\triangle$ is indeed small relative to this standard deviation, the actual process will be approximately Poisson. The scheduled gap, $\triangle$, determines the rate of the process, and should be chosen in such a manner that when summed over all operators in the airport the queue will not become excessive. This can be controlled by using basic queueing formulas (or see [9]). Note that for a particular airline $\triangle$ may be smaller during off-peak hours, while during peak hours it will have to be larger. In the future, it may

become necessary for airports to assign $\triangle$ to each airline, especially during peak time. Such an assignment could perhaps be based on the number of gates each airline operates, or by bidding. (See [1, 6, 9] for other approaches to this issue, including dynamic control as a function of the weather conditions at the hub.)

To continue with the basic model, we can still set one arrival or departure time arbitrarily, and in this case it is most convenient to set $Max\{S_i\} = S_{Max}$. In effect this reduces our problem to a search over a single variable, namely $min\{T_j\}$. Note now that if we increase (decrease) $min\{T_j\}$ by a small increment, all the other T values are increased (decreased) by the same increment. Therefore, the derivative $dZ/dmin\{T_j\}$ is obtained by summing (2.4) $\forall$ j, i.e.

$$\frac{dZ}{dmin\{T_j\}} = \sum_{j=1}^{n} \sum_{i=1}^{n} cD_{i,j} + \sum_{j=1}^{n} b_{o(j)} - \sum_{j=1}^{n} (P_j + b_{o(j)})(1 - \prod_{k=0}^{n} F_k(T_j - S_k)). \qquad \textbf{(4.3)}$$

After setting all S_i and T_j values as per their respective sequences and the assumed $min\{T_j\}$ value, (4.3) lends itself to a very easy search to find the $min\{T_j\}$ value which drives it to zero. The constraint (2.2) is easily taken care of by restricting the search to the appropriate domain. That is, if the derivative for $min\{T_j\} = S_{Max}$ is positive, the solution is at this point (recall that we add a constant to take care of the connection time); otherwise, $min\{T_j\}$ will be positive. As it happens, this more realistic version of the problem is much easier to solve numerically than the one discussed in the previous section --a welcome fringe benefit indeed.

It is interesting to note that had we chosen to set $min\{T_j\} = T_{min}$ (instead of setting S_{Max}), we would then have to evaluate the derivative $dZ/dMax\{S_i\}$ instead of $dZ/dmin\{T_j\}$. If we would proceed to do that by summing (2.3); $\forall$ i,

- 14 -

analog to our procedure above, the resulting expression would be much more cumbersome to evaluate than (4.3), due to the integrals involved. Of course, (4.3) could still be used instead, since when evaluated for the same value of $\min\{T_j\} - \text{Max}\{S_i\}$ the two derivatives sum to zero.

To complete the description of the basic model it remains to discuss how to set $\triangle_A$ and $\triangle_D$. As for $\triangle_D$, it makes sense to set it as low as possible, to better approximate the unconstrained solution--which sets $\triangle_D = 0$. However, it may happen that by setting $\triangle_A$ to a small value we actually increase the value of the objective function.

Assume we set $\triangle_D$ to the lowest feasible value, and we now wish to search for the optimal value $\triangle_A$. If it would be easy to evaluate Z for different values of $\triangle_A$, then a simple search could be performed in order to choose the value which minimizes it. Unfortunately, Z is quite cumbersome, due to the integrals involved. Therefore we may want to minimize an alternative objective function which is likely to move in the same direction as Z and is simpler to compute. A viable candidate for this is the sum of squares of the values of (2.4) computed for the S and T values associated with each value of $\triangle_A$. The solution for each $\triangle_A$ value ensures that the sum of the (2.4) values--i.e., (4.3)--will be zero, but in the uncapacitated solution the sum of their squared values is zero as well. Thus our alternative objective function tends to strive to this unconstrained optimum criterion. The search should be conducted for feasible values of $\triangle_A$ only, i.e., at least the minimal value allowed.

The method described above to search for $\triangle_A$ can now be extended and provide us with an improved model which does not specify an equal gap between adjacent arrivals or departures ($\triangle_A$ or $\triangle_D$), but rather specifies minimal gaps and allows larger gaps where appropriate. We still assume the same sequences, and strive to minimize the squared sum of (2.4) instead of minimizing

Z directly. Formally we should perhaps add the squared sum of (2.3) to the new objective function, since we allow the gaps between departures to vary as well as the gaps between arrivals. However, there are two reasons why not to do this: (i) (2.3) is much more difficult to compute than (2.4), and (ii) as discussed above, it is quite likely that the gaps between departures will assume the minimal $\triangle_D$ values, so their respective partial derivative values are less relevant. Denoting the index of the kth arrival by i(k) and the index of the kth departure by j(k), the model is:

$$\min \sum_{j=1}^{n} (\partial Z / \partial T_j)^2 \tag{4.4}$$

s.t.

$$(T_{j(1)} - S_{i(n)}) \sum_{j=1}^{n} (\partial Z / \partial T_j) = 0 \tag{4.5}$$

$$S_{i(k+1)} - S_{i(k)} \geq \triangle_A \; ; \; \forall \; k = 1, \ldots, n-1 \tag{4.6}$$

$$T_{j(k+1)} - T_{j(k)} \geq \triangle_D \; ; \; \forall \; k = 1, \ldots, n-1. \tag{4.7}$$

$$T_{j(1)} \geq S_{i(n)}. \tag{4.8}$$

$(\partial Z / \partial T_j)$ is as per (2.4). This model can be solved by standard NLP search programs. Note that if (4.8)--which is equivalent to (2.2)--is strictly positive, then (4.5) can only be satisfied by setting (4.3) to zero. The basic model becomes a special case by setting (4.6) and (4.7) as equalities and searching for the optimal $\triangle_A$ and possibly $\triangle_D$ values.

- 16 -

5.   The No-Waiting Case:

So far we have assumed that all departures wait if necessary for any number of passengers who have to make connections and whose arrivals have been delayed. In practice some airlines accommodate such passengers differently, by assigning them to alternative flights--including flights offered by competitors--and/or have them stay overnight at the airline's expense.   Virtually every airline would do so if a flight is so late that to wait for it would disrupt the schedule beyond some acceptable level; especially since sometimes such delays have a domino effect in other airports where passengers may still have to make further connections.   In this section we introduce a model for this policy.   We will consider a pure no-waiting policy, where departures do not wait even for a single minute, and the penalty per misconnection is a given constant, denoted by **p** (representing the average additional cost involved plus the imputed value of the passenger's dissatisfaction).   We will consider the finite capacity case only, following the general assumptions and approach of Section 4.

Even though our policy is not to wait, all the passengers whose final destination is $d(i)$ will still be delayed if flight $(0,i)$ arrives after $T_{d(i)}$, and thus the penalty $P_{d(i)} + b_i$ will have to be borne.   In addition, delay penalties may be incurred due to F_0, but these need not be considered here because they are not influenced by our decision variables.   However, F_0 also has a beneficial effect on the expected number of misconnections which we have to take into account.

To that end, let $F_i(T_j - S_i)$ denote the probability that the connection will be made if $(0,i)$ is scheduled to arrive at S_i and $(0,j)$ is scheduled to leave at T_j.   Also let $f_i(T_j - S_i)$ be the density function associated with $F_i(T_j - S_i)$.   Given F_i and F_0, we obtain

$$F_i(z) = F_i(z) + \int_0^{} F_i(z + t)(1 - F_0(t))dt \; ; \; \forall \; z \geq 0. \qquad (5.1)$$

Recall that under our assumptions F_0 is identically distributed for all departures, hence $F_i(T_j - S_i)$ is also identically distributed to $F_i(T_k - S_i)$; $\forall \; k \not= j$, $k \not= d(i)$. Therefore the single index i is sufficient to identify **F**.

Our objective function, Z, includes the value of the scheduled passengers' time, the expected number of misconnections multiplied by **p**, and the appropriate delay penalties for late departures. As mentioned above, the expected delays due to problems at the hub itself are not included. Also, the time value of passengers whose final destination is 0 is not taken into account--since it is not influenced by our decision variables. Note that passengers going to d(i) are not subject to misconnections. Thus

$$Z = \sum_{\substack{i=1 \\ j \not= d(i)}}^{n} \sum_{j=1}^{n} D_{i,j} \, [c \, (T_j - S_i) + p \, (1 - F_i(T_j - S_i))] + \sum_{j=1}^{n} b_{o(j)} T_j -$$

$$\sum_{i=1}^{n} b_i S_i + \sum_{i=1}^{n} [P_{d(i)} + b_i] \int_{T_{d(i)}}^{\infty} [1 - F_i(t - S_i)]dt. \qquad (5.2)$$

Taking the partial derivative of Z by T_j we obtain

$$\frac{\partial Z}{\partial T_j} = \sum_{\substack{i=1 \\ i \not= o(j)}}^{n} D_{i,j} \, [c - p \, f_i(T_j - S_i)] + b_{o(j)} -$$

$$[P_j + b_{o(j)}][1 - F_{o(j)}(T_j - S_{o(j)})] \; ; \; \forall \; j. \qquad (5.3)$$

To complete the model, all that remains is to plug (5.3) into (4.4) and (4.5) instead of (2.4).

6. Related Issues:

In this section we discuss several related issues, which should be taken into account while implementing scheduling methods such as the ones described above. Some of these issues merit further research, and a thorough treatment is beyond the scope of this paper.

In particular we'll discuss (i) the optimal speed choice for a plane as a function of its expected lateness; (ii) how to determine the sequences of arrivals and departures (which we have assumed given above)--an issue which turns out to be connected with the speed choice policy; (iii) how to devise a method comprising both delay penalties and misconnection penalties--where we wait only if the wait is expected to be short enough; (iv) how does the model impact the gate assignment problem; (v) how to assign the aircraft to the flight segments so as to minimize the number of passengers who have to change planes in the hub; and (vi) how to determine the scheduled duration of the flights, i.e., if flight (i,0) is scheduled to arrive at S_i, when should it leave i.

The Optimal Speed as a Function of Expected Lateness

Delays in flights are generally due to problems at the origin and at the destination. At the origin such delays may be the result of weather conditions, queueing, problems in processing all boarding passengers in time (especially if stand-by passengers are involved), problems in preparing the plane for takeoff (fueling, loading, mechanical problems), etc. At the destination, delays may be caused by weather conditions or by queueing. Assume now that a flight has departed behind schedule, and has a considerable distance to go toward the destination--then it may be possible to recapture some of the delay by utilizing a higher speed. The question is whether or not this is economically justifiable, and how much to increase the speed if so. Note that in this case, the CDF

of the arrival time given the speed choice is a lot tighter then the original CDF, since all the variability in departure time has already been realized. Thus it may become possible to specify a fairly tight updated ETA for the flight as a function of the average airspeed we choose.

The average speed and the variable operating costs for a particular flight are determined by the flight trajectory, which can be viewed as a trajectory in a three dimensional space comprising distance from origin, altitude, and time. This trajectory has to be optimized afresh for each distance/airtime combination, and is also influenced by the payload and by weather conditions. In this paper we ignore the weather and payload factors, and we will assume that for each flight we have a function which yields the variable cost associated with any possible value of the flight duration. This function is conceptually obtained by optimizing the trajectory for each value of the argument. The domain of the function is restricted from above by the slowest economical speed (i.e., the speed which maximizes the range), and from below by the maximal speed. Obviously, this is a monotone decreasing function, and we will assume that it is convex over its domain. Our main concern is with flights into the hub, and we'll denote the function for flight $(i,0)$ by $g_i(t)$.

Denote the value of each time unit for flight $(i,0)$ by s_i. This includes--but is not restricted to--the value of the passengers' time and b_i. The optimal time allotted to the flight should minimize the total cost for the flight--TC_i

$$TC_i = s_i t + g_i(t). \qquad \qquad (6.1)$$

By taking the derivative of TC_i by t, we obtain the following first order condition for the optimal time

$$d[g_i(t)]/dt = -s_i, \qquad\qquad (6.2)$$

If $d[g_i(t)]/dt \leq -s_i$; $\forall$ t in the domain, we choose the minimal speed; and if $d[g_i(t)]/dt \geq -s_i$; $\forall$ t in the domain, we choose the maximal speed. Otherwise, under our convexity assumption, there is exactly one solution for (6.2), strictly within the domain. It is easy to see that if we increase s_i, we should also increase the speed, unless we were already using the maximal speed.

Suppose now we operate under the first policy, and our ETA given the regular speed for the flight is such that some departures are going to be delayed waiting for flight (i.0), then the time value for the flight should be increased and include the time of all delayed passengers at the penalty rate p, as well as the time value of the waiting planes and crews. If several planes are waiting, it is quite likely that we'll have to use the maximal speed. Similarly, under the second model, by increasing the speed of an already delayed flight we may reduce the expected number of misconnections, and should choose a higher speed.

The details of the exact speed choice model under expected lateness require further research, but it is clear even at this stage that the amount of time which can be recaptured economically (and technologically) is higher for long flights than for short ones. This implies that long flights are less likely to be excessively late in practice than short ones, but they may also incur higher operating costs than our model accounts for.

Sequencing the Flights In and Out of the Hub

The sequencing problem exists in the capacitated cases only, since in the infinite capacity case the sequences are determined by the solution itself. The problem is nontrivial because of the stochastic elements. To wit, if we ignore these we get the following simplistic version of the sequencing problem

$$\min\left\{\sum_{j=1}^{n}T_j[b_{o(j)} + c\sum_{i=1}^{n}D_{i,j}] - \sum_{i=1}^{n}S_i[b_i + c\sum_{j=1}^{n}D_{i,j}]\right\}. \qquad \textbf{(6.3)}$$

(6.3) lends itself to decomposition, and is optimized by minimizing the sum involving T_j and maximizing the sum involving S_i. The values within the brackets measure the time value of each flight, and are easy to calculate. All that remains is to sort the flights by these time unit value measures. In the case of arrivals, they should be sequenced in ascending order, while the departures should be sequenced in descending order. This solution can be easily adapted to cases where the value of time unit per passenger is a function of the origin/destination pair, in which case the time unit measures will include weighted values of $D_{i,j}$.

When we do take the stochastic elements into account, the solution above may be far from optimal. Indeed, looking into the uncapacitated versions of the problem may give us some insight to that effect. Recall that the departures take place at the same time, but, since we do not assume identical CDFs for all flights, we may expect the optimal solution of the arrival times to define sequences which have more to do with the variances of the delays in the origin airports than with the number of passengers or the time unit value of the aircraft. It is quite likely that this order will be better than the one suggested above. Similarly, under the no-waiting policy, the order suggested by the sorting tends to increase the number of misconnections, since flights with many passengers tend to have high time unit values, and are sequenced with less ground time.

Solving the sequencing problem optimally may require actually comparing all the possible sequences in terms of the resulting objective function values. Even if we resort to simpler alternative objective functions such as (4.4) this

is a tall order, since the number of sequences is exponential. Therefore, we suggest a heuristic approach based on the discussion above. We describe the heuristic in terms of the waiting model (Section 4), but the same idea will also work for the no-waiting policy. We start with the uncapacitated solution, which defines a sequence for the arrivals. In order to avoid having to solve (2.3) for all i, we suggest here to use the bound (3.5) instead. Next, fixing the scheduled arrivals as per this sequence, at intervals of $\triangle_A$, we solve for each T_j separately so as to drive (2.4) to zero. This will define a sequence for the departures.

As discussed above, flights from afar can compensate for delays by a higher speed. This implies that it may be expedient to schedule these flights to arrive later than flights from nearer origins. Likewise, we may wish to schedule flights to farther destinations to depart earlier than the others. Such a policy constitutes a second heuristic approach to the sequencing problem.

Looking at this second heuristic, the order it defines may be completely different than the one implied by the first heuristic. Indeed, this heuristic ignores the differences in the original CDFs completely. The question is, can we take both into account at the same time? One method to do that is to "cheat" the first heuristic by using inflated values for the time unit values on long segments. That is, we increase the value of C_i in (3.5) for the longer segments, and similarly decrease the value of P_j in (2.4) for longer segments. This will cause the first heuristic to shift in the direction of the second one.

Combining Delay and Misconnection Penalties

It is easy to devise examples where the waiting policy will cause hundreds of passengers to wait unbounded periods for a few delayed passengers. On the other hand, the no-waiting policy may imply that a plane leaves one minute

- 23 -

before a large group of passengers could board. Therefore, neither of the pure policies we examined should be followed too religiously. We need to devise a combined policy, where we wait, but not indefinitely. That is, if an incoming flight is late, we have to decide which outbound flights should wait for it. This decision should be based on the late flight's ETA and revised CDF, and may have to be updated dynamically.

Theoretically, the fact that the final decision about waiting is not based on a predetermined pure policy, makes the model very complex to optimize analytically. We propose a very simple heuristic approach instead:

(i) Compare the two pure methods in terms of the objective function,

(ii) schedule the flights as per the pure method which costs less, and

(iii) improve the actual performance by local optimization in case of delays.

Note that step (i) also determines the sequencing of the arrivals and of the departures.

The immediate decisions outlined in step (iii) may justify the development of a decision support system (DSS), since they have to be made in a short time. Such a DSS is not likely to require much in terms of computing power and will use data readily available on the network. Thus it should not be too difficult to implement.

Implications for the Gate Assignment Problem

The existing literature on gate assignment strives to minimize the total walking distance for all passengers[2, 5]. While this is a reasonable objective, one must also consider the time allotted for making a connection when calculating the cost involved. Clearly, if your flight is about to leave, it is more important for you that it be at a nearby gate than otherwise. This establishes a link between the gate assignment problem and the scheduling problem.

An obvious heuristic suggested by this consideration, is to assign the latest arrivals to the more central gates, and if possible at the same time to assign the first departures to the more central gates. This approach, however, does not take into account the traditional objective function (total walking distance) at all.

An alternative approach may be, again, to "cheat" the regular gate assignment algorithms in such a manner that they will favor the tight connections. Given the schedule, this can perhaps be done by dividing $D_{i,j}$ (a required input for the regular algorithms) by the probability of connection (one minus the probability of misconnection) raised to some power not less than one (e.g., squared). When the probability of connection is close to one, $D_{i,j}$ will not be altered by much, but when the probability is low, $D_{i,j}$ will be significantly inflated.

In our uncapacitated model, the probability of connection is likely to be roughly equal across all connections, and hence it is not likely to influence the gate assignment by much. However, in the capacitated model, some connections are forced to assume lower probabilities of connection, while others are forced to longer waits instead. In this environment, the gate assignment is likely to be significantly different. Note also that we may choose to change the assignment dynamically; e.g., if a plane is known to be late, and its original gate is not favorable, we may arrange a better gate for it to reduce the expected penalty.

Finally, when determining the minimal connection time which should be added to all departures, we can use a lower value if we know that the tight connections are going to be made from favorable (central) gates. In this way the gate assignment problem may influence the exact scheduling--as well as the scheduling influencing the gate assignment.

Assigning Flight Segments to Aircraft

As discussed in Section 5, the final destination of the $(i,0)$ aircraft ($\forall$ i) is an important factor vis a vis the expected number of misconnections. It also impacts the gate assignment problem, since the $(i,d(i))$ passengers enjoy zero walking distance to their departure. So far we have assumed that the assignment of aircraft to flight segments is given. Let us now discuss how this assignment may be carried out.

Our approach to the problem is by minimizing the number of passengers who have to change planes, subject to constraints on the allowed pairs. The constraints may be used to promote good load levelling among the planes.

We assume all aircraft are of the same type and capabilities. In practice this is not likely. However, if we assign the aircraft type to each segment in advance--as a function of the total number of passengers and the distance involved--we can solve the segment assignment problem for each aircraft type and the origin/destination airports associated with it separately. Therefore this assumption is not restrictive.

We proceed by creating a square matrix where the element in the ith row and j^{th} column is $\sum_{k \neq j} D_{i,k}$ (i.e., the number of passengers who will have to change planes if we choose $d(i) = j$). Next, if certain pairs have a total distance which is too low or too high we can prevent them by adding large penalties to their entries. Now, we can use the Hungarian method to assign the destinations to the origins.

A major problem with this solution is that it may take up to n days for a plane to return to its original origin. In practice the policy may be to have symmetry, i.e., $d(d(i)) = i$; $\forall$ i. This policy ensures that each plane will return to its home base every other day (assuming one flight per day per plane from i to d(i) through 0). We can optimize the assignment under this policy by

comparing the $n(n-1)/2$ combinations explicitely. This is possible only for an even n; if n is odd we have to choose one flight which returns to its origin, or select three flights which rotate on a three days basis. If we choose to have one flight return, then exhaustive enumeration requires comparing $n(n-1)(n-2)/2$ alternatives, and if we choose to select three flights to rotate every three days the number increases to $n!/(12(n-5)!)$--which is $O(n^5)$. In these cases we may wish to reformulate the problem as an ILP model. Though formally the worst case perfomance of such an ILP model can be worse than $O(n^5)$, it may be better on average.

Scheduling the Durations of Flights

In the main body of the paper, we assumed that the duration of each flight is predetermined. Thus if flight $(i,0)$ is scheduled to arrive at S_i, then in order to schedule its departure from i we simply subtract this duration from S_i. We also assumed above that F_i is given $\forall$ i. Note now that if we allow more time without changing the speed choice, we can shift F_i to the left, thus increasing the probability of arrival, and vice versa. Some reflection reveals that any such change in the assigned duration would cause either of our models to change S_i in such a manner that the probability of connection will remain exactly the same, i.e., S_i is shifted by the same amount. Note that only the $D_{i,0}$ passengers would be impacted by such a change--their time value increases to p if the flight is delayed beyond S_i as per the last element of (2.1); the other participants only care if subsequent departures are impacted. Therefore we have to optimize the assigned duration in terms of the $D_{i,0}$ passengers only. This is achieved by setting the probability of delay times p equal to the complementary probability times t, as per the newsboy model. The same method should be used to determine the nominal durations of the $(0,j)$ flight segments.

References

[1] G. F. Andreatta and G. Romanin, Aircraft Flow Management under Conjestion, _Trans. Sci._ **21**, 249-253 (1987).

[2] O. Babic, C. Theodorovic and V. Tosic, Aircraft Stand Assignment to Minimize Walking, _J. Trans. Engineer_ **110** (No. 1, January), 1984.

[3] M. M. Etschmaier and D. F. X. Mathaisel, Airline Scheduling: An Overview, _Trans. Sci._ **19**, 127-138 (1985).

[4] R. W. Hall, Vehicle Scheduling at a Transportation Terminal with Random Delay en Route, _Trans. Sci._ **19**, 308-320 (1985).

[5] R. S. Mangoubi and D. F. X. Mathaisel, Optimizing Gate Assignments at Airport Terminals, _Trans. Sci._ **19**, 173-188 (1985).

[6] A. R. Odoni, The flow Management Problem in Air Traffic Control, pp. 269-288 in _Flow Control of Conjested Networks_, A. R. Odoni, L. Bianco and G. Szego (eds.), Springer-Verlag, New York 1987.

[7] B. Ronen and D. Trietsch, A Decision Support System for Purchasing Management of Large Projects, to appear in _Opns. Res._

[8] B. Ronen and D. Trietsch, Optimal Scheduling of Purchasing Orders for Large Projects, New York university, Graduate School of Business Administration, Center for Research on Information Systems, CRIS #127, 1987.

[9] R. C. Rue and M. Rosenshine, The Application of Semi-Markov Decision Processes to Queueing of Aircraft for Landing at an Airport, _Trans. Sci._ **19**, 154-172 (1985).

Distribution List

Agency	No. of copies
Defence Technical Information Center Cameron Station Alexandria, VA 22314	2
Dudley Knox Library, Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Office of Research Administration Code 012A Naval Postgraduate School Monterey, CA 93943	1
Library, Center for Naval Analyses 4401 ford Avenue Alexandria, VA 22302-0268	1
Dan Trietsch Code 54Tr Naval Postgraduate School Monterey, CA 93943	40