

AIRLINE NETWORK DESIGN

PHILLIP J. LEDERER

University of Rochester, Rochester, New York

RAMAKRISHNAN S. NAMBIMADOM

Susquehanna Investment Group, Chicago, Illinois

(Received April 1995; revision received September 1996; accepted July 1997)

The goal of this paper is to understand choices of networks and schedules by a profit maximizing airline. By "network" we mean the routing pattern for planes and by "schedule" we mean the frequency of service between cities *and* the amount of time put into the schedule to assure on-time arrival. This paper analyzes network and schedule choice using an "idealized" model that permits derivation of analytic, closed form expressions for airline and passenger costs. Many important conclusions are obtained. It is optimal for a profit maximizing airline to design its network and schedule to minimize the sum of airline and passenger costs. Profit maximizing choice of schedule frequency depends on the network. Direct service has lower schedule frequency than other networks.

Parametric studies are performed on the effect of distance between cities, demand rate, and the number of cities served on the choice of the network. Some conclusions are: (1) If the distance between cities is very small, then direct service is optimal; otherwise, other networks, such as hub and spoke are optimal. (2) Similarly, for very high demand rates, direct service is optimal; and for intermediate values, hub and spoke is optimal. (3) If the number of cities is small, direct service dominates; and if it is large, hub and spoke is optimal.

We note that any airline's schedule includes safety time as a buffer against delays, and we demonstrate that schedule reliability is highest for direct routing. Surprisingly, the amount of time that is added to the schedule to buffer delays is relatively less in direct networks than in other networks. This can explain the superior on-time performance and high equipment utilization of direct carriers such as Southwest Airlines.

There have been dramatic changes in the routing structure of airlines in the past fifteen years. Since deregulation in the United States and elsewhere, hub and spoke systems have emerged as the dominant network design. Recently however, the business press has noted the financial success of carriers offering direct service. One of these carriers, Southwest Airlines, has been judged America's most profitable airline and the industry's leader in on-time performance in 1992 (*N.Y. Times*, July 15, 1993, D1). Also, many airlines offering direct service have recently begun operation. Although the economic advantages of hub and spoke networks have been studied in the transportation and economics literatures, there has been limited examination of other types of airline networks. With the emergence of nonhub networks, study of alternate networks is required. In particular, questions such as "when do we expect direct service to be more profitable than hub and spoke?" and "do we expect direct service schedules to be more reliable than those of other networks?" need to be addressed.

In this paper we answer these questions by studying the design of an airline's network and its schedule. We interpret the "network" as the routing pattern for planes and the schedule as the frequency of service between cities *and* the amount of time put into the schedule to assure on-time arrival. Our goal is to examine how network choice and scheduling decisions affect airline cost and passenger service quality. Airline cost is the carrier's cost of operating

its network, and passenger service quality is measured by the cost borne by passengers due to travel time, *schedule delay* (the time difference between the ideal departure time and the actual departure time), and incidents of late arrival. These costs are important because a necessary condition for an airline's profit maximizing choice of network and schedule is that the sum of these costs are minimized.

Four different types of networks are studied, which we call *direct*, *hub and spoke*, *tour*, and *subtour* routings. Direct routing is a network where all cities are directly connected. Hub and spoke routing is a network where all passengers travel by airplane to a hub and then catch a flight to their final destinations. A tour routing is a network where each plane travels from city to city until all have been visited. A subtour routing is a network where passengers take a plane from their origin city, and travel on this plane to the hub, generally making some stops at other cities before reaching the hub. At the hub, they catch another plane that takes them to their destinations, perhaps making some stops along the way. We show that each routing can maximize airline profit under some conditions.

This paper analyzes network and schedule choice using an "idealized" model. The idealized model permits derivation of analytic, closed form expressions for airline and passenger costs. With the model, important characteristics of the real system can be captured, while the model remains simple enough to derive formulae for the costs. Using our understanding of the relationship between

Subject classifications: Networks: design of airline networks. Transportation: economics. Economics: consumer choice, production cost.
Area of review: SERVICES.

network/schedule choices and airline profit, we study how such factors as distance between cities, number of cities served, and demand rates affect the choice of profit maximizing network and schedule.

Many important conclusions are obtained. It is optimal for a profit maximizing airline to design its network and schedule to minimize the sum of airline and passenger costs. This is because changes in the network or schedule that decrease passenger costs can be captured by the airline though price increases. We show that profit maximizing choice of schedule frequency depends on the network, and that direct service has lower schedule frequency than other networks. It is also shown that if passengers are very sensitive to schedule frequency because schedule delay costs are high, then a tour is the profit maximizing network choice.

Parametric studies are performed on the effect of distance between cities, demand rate, and the number of cities served on the choice of the profit maximizing network. Some conclusions are: (1) If the distance between cities is very small, then direct service is optimal; if it is very large, then a tour is optimal; and for intermediate distances, other networks, such as hub and spoke are optimal. (2) Similarly, for very high demand rates, direct service is optimal; for very low, a tour is optimal; and for intermediate values, hub and spoke is optimal. (3) If the number of cities is small, direct service dominates; if it is large, hub and spoke is optimal; and for intermediate values, other networks are best.

We note that any airline's schedule includes safety time as a buffer against delays. This planned delay time increases passengers' travel time and airline cost since more craft are required, but reduces passengers' chances of arrival after the scheduled time. It is shown that if schedule reliability is chosen to minimize total airline and passenger costs, schedule reliability is highest for direct routing. Surprisingly, the amount of time that is added to the schedule to buffer delays is relatively less in direct networks than other networks. That is, relatively less time is included in direct schedules to protect against delays, but the reliability of direct networks is highest. This can explain the superior on-time performance and high equipment utilization of non hub carriers such as Southwest Airlines. Finally, an example shows that congestion at the hub does not change optimal network choice unless delays at the hub are very much larger than at spoke cities.

Literature Review

Study of airline networks spans several disciplines. Economics researchers, operations researchers, and transportation engineers all have made contributions to understanding airline networks.

In the economics literature, several analytical papers consider the effect of network topology on competition within the airline industry. Lederer (1993) models competition as a noncooperative game where airlines select network designs and prices for transportation between any

two nodes. Oum et al. (1993) study strategic use of hub networks to discourage entry. Brueckner and Spiller (1991) model economies of scale and study the public welfare effects of competition between airlines due to cost externalities.

Several empirical economics papers consider airline cost economics. Bailey et al. (1985) show that a hub and spoke system can decrease airline cost compared to direct routings by exploiting economies of scale by using larger, lower cost aircraft. Caves et al. (1984) study passenger airlines demonstrating economies of traffic density and constant returns to scale for the number of cities served. Oum and Tretheway (1990) demonstrate the evolution of several airline networks from linear to hub systems, and describe a hub's effects on airline cost and passenger travel time. McShan and Windle (1989) measure the change in hub and spoke routing in the United States since deregulation and estimate the effect of this change on airline costs. Transportation engineers predict consumer choice of transportation services using random utility models; see for example Ben-Akiva and Lerman (1985). Morrison and Winston (1985) present a detailed econometric analysis of business traveler's modal choice as a function of factors such as transportation price, transit time, and time between departures. They find that price, travel time, and *schedule delay* are the most important passenger choice factors.

Several operations research papers present analytic models of airline cost. The most relevant to our approach is Jeng (1988), who analyzes an idealized airline routing problem. His purpose is to study the mix of hub and direct routing flights chosen by an airline. He shows how network parameters such as demand level, geographical distance between cities, and number of cities served affect the routing choices. Unlike this paper, he assumes that passenger demand is inelastic and does not consider aircraft schedules or the effect of networks on schedule reliability. Kanafani and Ghobrial (1985) study the effect of congestion at hub airports. Using a case study and an analytic model, they show that hubs are efficient even when landing fees are used to reduce hub delays. Kuby and Gray (1993) study package delivery systems and generalize hub networks to allow stopovers and feeders. They formulate a mixed integer mathematical program to solve for the cost minimizing network. Similarly, Marsten and Muller (1980) formulate a mathematical program to design an air cargo carrier's route and plane assignment for spider-shaped networks. Trietsch (1993) uses a generalized newsboy model to study optimal choice of ground time in hub systems with the objective of minimizing airline and passenger costs due to delay.

The rest of this paper is organized as follows. In Section 1, we discuss the notation used in the paper and develop the basic model. Detailed expressions for passenger and airline costs are presented in Section 2. Section 3 obtains the optimal frequency for different types of networks. Section 4 discusses the nature of optimal network and its cost under various conditions. Section 5 studies the optimal

choice of schedule reliability, which is the probability of on-time arrival. Section 6 generalizes one of the assumptions in the paper by considering a situation where congestion at the hub causes greater delays than at spoke cities. Finally, the paper concludes with a review of the major conclusions and a discussion of future research opportunities.

1. THE MODEL: VARIABLES AND ASSUMPTIONS

The following variables are used in the paper:

- n — number of cities on the circumference of a circle,
- R — radius of circle (miles),
- v — velocity of aircraft (mph),
- f — frequency of service per day,
- k — number of cities on subtour in a subtour network,
- m — ($= n/k$) number of subtours in a subtour network,
- ρ — demand for transit from one city to any other per day served by the airline,
- i — city index, $i = 1, \dots, n$,
- p — subtour index for a subtour network, $p = 1, \dots, m$,
- g — ground time per flight leg; $\frac{1}{2}g$ per takeoff and landing (hrs),
- α — probability of delay at takeoff or landing,
- λ — delay parameter; $1/\lambda$ is the expected delay in hours per takeoff or landing,
- ω — passenger's cost per hour due to schedule delay,
- ν — passenger's cost per hour due to trip travel time,
- a, b — parameters of an airplane's variable cost (see Equation (1-1)),
- c, d — parameters of an airplane's fixed cost (see Equation (1-2)).

Models of network design and cost are based on the following assumptions.

Assumption A1. *There are n cities equispaced on the circumference of a circle of radius R served by several airlines.*

Assumption A2. *There is demand for transit between each of the n cities. The demand per day is the same for all origin and destination pairs and equal to ρ . The ideal departure times of customers are uniformly distributed throughout the day.*

Assumption A3. *There is no demand for transit originating or destined to the hub, which is located at the center of the circle.*

Assumption A4. *The frequency of scheduled flights from each pair of cities is the same per day, and the scheduled departures are evenly spaced.*

Assumption A5. *Aircraft are available in a continuum of capacity. The airline schedules enough capacity to satisfy demand for the highest demand segment of any routing.*

Assumption A6. *The following assumptions about airline cost are made. The variable cost of flying an airplane is*

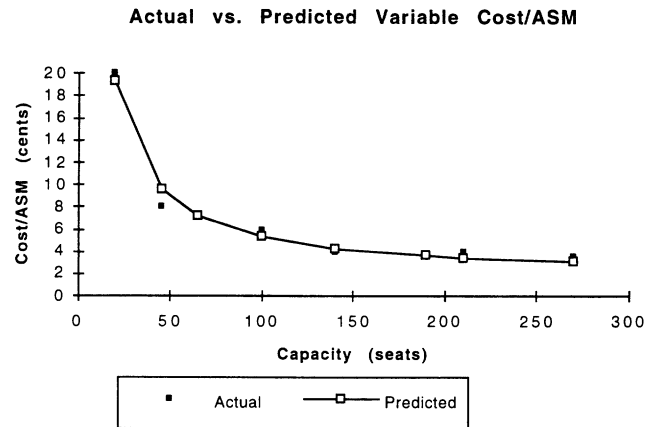


Figure 1. Variable cost per available seat-mile (ASM): actual and predicted.

$$\text{Predicted model cost/ASM} = 1.79 + 351.93 \frac{1}{\text{Seats}^2}; R^2 = 0.98. \quad (4.3) \quad (17.6)$$

(Numbers in parentheses are t -statistics.)

linear in distance traveled per seat, but this rate depends on airplane size. The variable cost per seat-mile displays economies of scale with respect to aircraft size. The fixed cost per seat of possessing an airplane displays economies of scale with respect to aircraft size.

Assumption A7. *Aircraft require time for takeoff, landing, and ground time.*

Assumption A8. *Delay of takeoff or landing is not related to the number of aircraft landing or taking off. This assumption is relaxed in an extension, where differing delay characteristics are considered.*

Assumption A9. *For each origin and destination, ω and ν are identical.*

The assumptions facilitate analysis, without losing the properties of real networks.

Assumptions A1, A2, A4, and A9 are symmetry assumptions that allow computation of analytic expressions for costs. Assumption A3 is made so that the choice network is independent of demand originating or destined for the hub. This assumption clearly biases our results against hubs. If there is significant hub demand, the attractiveness of a hub structure would certainly increase. For many existing hubs, 30% (or more) of demand is to or from the hub. If significant hub demand is assumed, then a hub network is likely to be optimal. This paper seeks to study conditions under which other networks are optimal, so assumption of significant hub demand prevents study of this question. Note that if we find that a hub and spoke network is optimal, then it is optimal even with hub demand. Assumption A5 is approximately true as far as aircraft sizes: there is a range of available aircraft capacities, although in discrete sizes (see Figures 1 and 2 for the craft sizes owned by one carrier). Our assumption that aircraft sizes match the demand might be difficult to accomplish. If

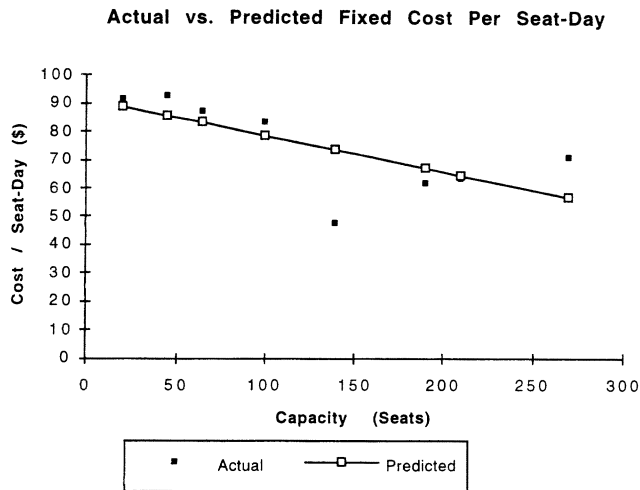


Figure 2. Fixed cost per airline seat-day: actual and predicted.

$$\text{Predicted model cost/airline seat-day} = 91.45 - 0.13 \text{ Seats}; \quad (10.7) \quad (-2.3)$$

$$R^2 = 0.47.$$

(Numbers in parentheses are *t*-statistics.)

demand varies through the day, or aircraft are employed in different sized markets, the size and demand will not always match. This part of A5 is clearly an approximation.

Assumption 6 is supported by empirical data. A fleet manager at a major U.S. passenger airline provided the airline cost data found in Figures 1 and 2. The figures report variable cost per available seat-mile and fixed cost per available seat-day as a function of aircraft size. Both variable and fixed costs display significant economies of scale. The variable cost per available seat-mile closely follows the function,

$$f(\# \text{seats}) = a + b/(\# \text{seats}), \quad (a, b > 0), \quad (1-1)$$

so total variable cost *per mile* displays constant marginal cost with respect to capacity. Fixed cost per available seat-day follows the function:

$$g(\# \text{seats}) = c - d/(\# \text{seats}), \quad (c, d > 0). \quad (1-2)$$

Although fixed cost per available seat-day is monotone decreasing in aircraft size, we assume it does not decline beyond 300 seats for the sake of realism. Assumption A7 implies that a takeoff or landing generates cost for the airline because the time consumed increases aircraft requirements. Thus, the total cost of operating an aircraft will be a concave increasing function of stage length (distance between origin and destination). This accords with empirical observation as reported by Jeng (1987).

Assumption A8 is made for simplicity and is generalized in Section 6 to networks where congestion is greater at hubs than at spoke airports. The assumption ignores the problem of congestion at hub airports, and, thus, seemingly biases our analysis toward hubs. However, it is shown that congestion at the hub does not change the optimal network unless delays at the hub are much larger (three or

four times larger) than spoke airport delays. Thus, A8 is a good first approximation.

1.1. Description of Networks

We next describe the networks in more detail. Each network is assumed to be operated with a schedule that has a service frequency f per day between any two cities and with aircraft large enough to satisfy expected demands. The routings described below provide service frequency of f per day between any city pairs and ensure that the routings can be repeated with some period.

In a *direct* network all spoke cities are directly connected. (See Figure 3a). Every $1/f$ days a plane leaves city 1 and travels to city 2, returns, leaves for city 3 and returns, and continues this pattern until it has visited and returned from city n . Then, the aircraft travels (empty) to city 2, leaves, and travels to city 3, returns, leaves for city 4 and returns, and continues this pattern until it has visited cities 5 through n . Then the aircraft travels (empty) to city 3, leaves and travels to city 4, returns, leaves for city 5 and continues in this pattern, until the aircraft has returned to city 1. Once complete this (slightly strange) routing assures that all spoke cities have been directly connected, and that the schedule can be repeated. In this paper, a direct network is denoted $k = 0$.

In a *hub and spokes* network a plane is dispatched from each spoke city every $1/f$ days, the planes arrive at the hub, wait for all passengers to change between planes and then the planes depart to the spoke cities. In this paper hub and spoke network is denoted $k = 1$. (See Figure 3b.)

In a *tour* network a plane leaves city 1 every $1/f$ days traveling in a designated direction (say counterclockwise) for the neighbor city, arrives at that city, proceeds to the next city, say city 2, and continues the tour of the spoke cities on the circumference. On reaching city n , the plane retraces its path back. This routing ensures that all spoke city pairs are served with a frequency f . In this paper a tour network is denoted $k = n$. (See Figure 3c). Although tours are rare for airlines, in mass transit and freight trucking they are not. It is an extreme version of the subtour, which is described next.

In a *subtour* network the n spoke cities are divided into m subsets, each containing k ($= n/m$) adjacent cities. It is required that k is an even divisor of n . A plane leaves the first spoke city in each subtour every $1/f$ days, travels to the nearest spoke city in a designated direction, say counterclockwise, and stops at the k spoke cities in the subtour. After departing the k th city, the plane travels to the hub, where it meets other arriving planes, passengers are interchanged, and then the plane departs to the last city visited on the subtour, tracing the cities on the subtour in the opposite direction, finally arriving at the first city on the subtour. A subtour is constructed so that the routing ends where it begins and each city pair is served at the same frequency. In this paper, a subtour network is denoted by the number of cities on a subtour ($k = n/m$), where k and m are both integer divisors of n . (See Figure 3d.) Subtour

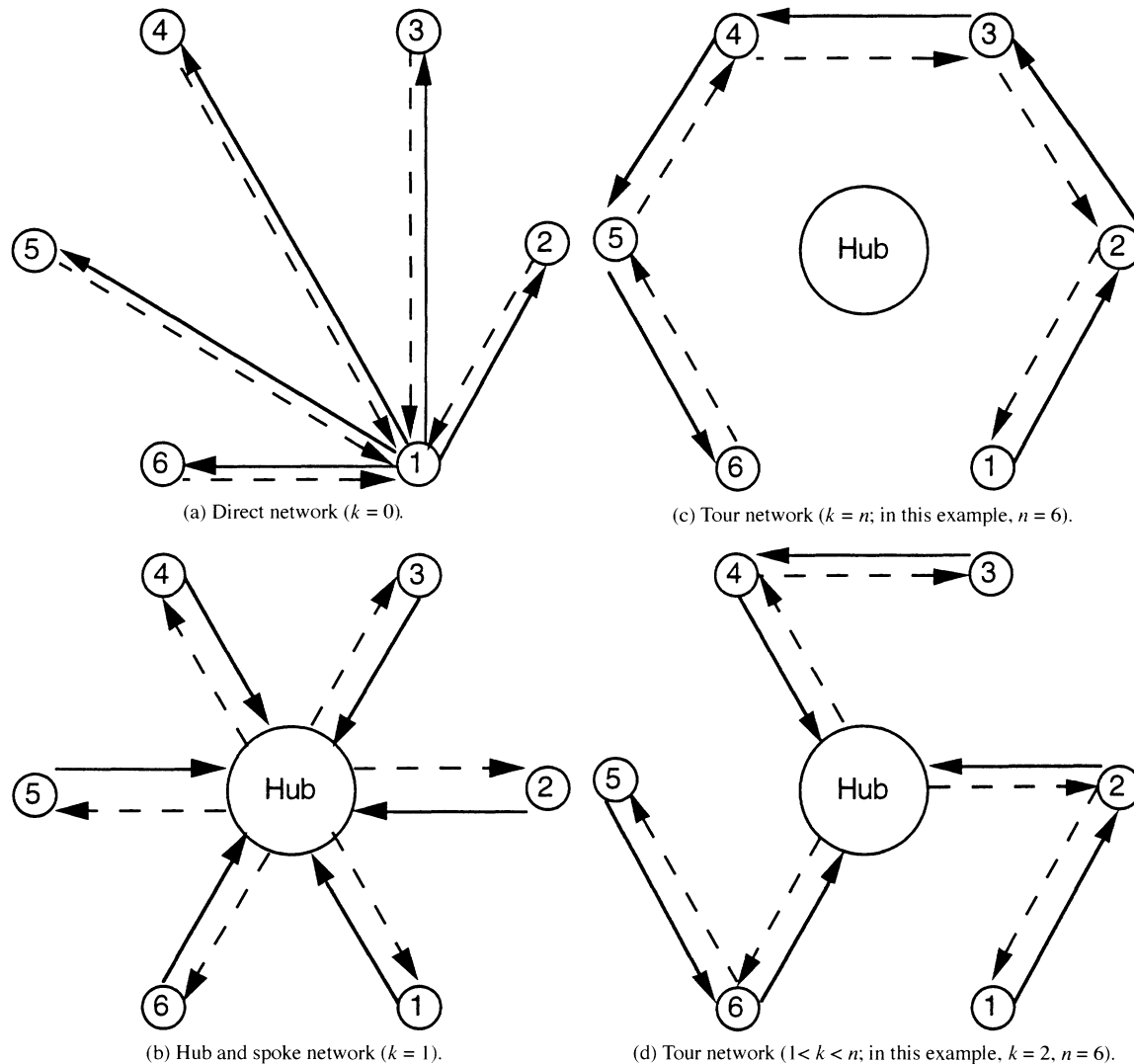


Figure 3. Network connections. (For clarity, only the direct connections between city 1 and the other cities are shown; all the cities are connected by direct flights.

networks are used frequently by airlines. Note that a *tour* is an extreme version of a subtour where the path directly connects spoke cities without going to the hub at all!

Thus, the set of networks is conveniently parameterized by the integer k . For any n , the set of networks considered are $k = 0$, and all k that are even divisors of n . For example, if $n = 24$, k takes values 0, 1, 2, 3, 4, 6, 8, 12, and 24.

1.2. Airline Profit and Consumer Choice

This paper studies network choice by profit maximizing airlines. We show that an airline's profit is maximized by the network design and schedule that minimizes the sum of airline and passengers' costs. In particular, we demonstrate that the airline's problem of maximizing profit by choosing network, schedule and prices, subject to meeting a demand vector d is equivalent to *minimizing the sum of airline and passengers' costs by choosing the network, schedule and prices subject to meeting fixed demand d .*

To understand this we examine customer choice of alternate routings. We assume that consumers choose between

alternative routings on the basis of the routings' utilities, and each customer's utility function is linear and separable in various attributes of service. This is often assumed in the literature; see Ben-Akiva and Lerman (1985), for example. In particular, customer's utility declines by p when the fare is p dollars. Customer's utility declines by $t\nu$ when t is the travel time. If a passenger desires to leave at some ideal time, but instead departs at a time h hours earlier (or later), he/she suffers a utility decline of $h\omega$. Customer's utility is often described in terms of average schedule delay (see Morrison and Winston 1985, for example). There is a flight every $1/f$ days so that the average schedule delay cost is $\omega/(2f)$. A customer's utility is thus:

$$U(p, t, f) = \text{Constant} - p - \nu(t) - \omega/(2f). \quad (1-3)$$

(Customer's utility also declines when arrivals are late, but we postpone study of this issue for now. Clearly, linear terms in expected late-arrival-time can be added.)

The key idea is that if changes in travel time and frequency are compensated for by changes in price so that

total utility is constant, then demand remains unchanged. Suppose a market “ ab ” has initial price, travel time and frequency, (p_o, t_o, f_o) , with corresponding total utility $U_{ab}(p_o, t_o, f_o)$ as given by (1-3). For any changes in the network that affect customer utility, assume price is adjusted as follows. If in the new network the travel time and frequency for ab are t_{ab} and f_{ab} , respectively, then the price for this market, p_{ab} , is chosen to satisfy

$$p_{ab} = \text{Constant}_{ab} - \nu(t_{ab}) - \omega/(2f_{ab}) - U_{ab}(p_o, t_o, f_o). \quad (1-4)$$

We could view (1-4) as an identity that defines prices. Then, the airline will attract the same number of passengers no matter the network or frequency. Using these prices, if demand from origin to destination is d_{ab} , then the airline’s profit as a function of its network is

$$\text{Profit} = \sum_{ab} p_{ab} d_{ab} - \text{Cost}(\text{network}, d), \text{ or}$$

$$\begin{aligned} \text{Profit} = \sum_{ab} & \left(\text{Constant}_{ab} - \nu(t_{ab}) - \omega \left(\frac{1}{2f_{ab}} \right) \right. \\ & \left. - U_{ab}(p_o, t_o, f_o) \right) d_{ab} \\ & - \text{Cost}(\text{network}, d). \end{aligned} \quad (1-5)$$

where $\text{Cost}(\text{network}, d)$ is the airline’s operating cost for the fixed demand vector d for the network chosen (we assume the network decision implies f and t for every market). Therefore, airline profit is maximized by choosing the network that minimizes the sum of passenger travel time cost, passenger schedule delay cost and the airline operating cost. Reductions in cost will be profitable only when they do not exceed the increase in passenger cost.

We also conclude that changes in the network that degrade passengers costs must be compensated for by decreases in price. There are clearly limits to such decreases: For example, the large travel time in a 50-city tour could never be compensated for without negative prices. This implies that certain networks with very high passenger costs will not be feasible alternatives for an airline!

Identity (1-4) ensures that passengers’ utilities for service are fixed, and we optimize the airlines profit for these utilities. However, a strategic decision that the airline can make is to change the utilities, and thus change demand. This can be done by varying price in ways different from (1-4). For any fixed passengers’ utilities, we have demonstrated that the airline’s profit maximizing network and prices must minimize total airline and passenger costs. There is a vector of initial utilities that are optimal in the sense that the airline’s profit is maximized over the set of all passengers’ utilities. Clearly, for optimally chosen utilities, the profit maximizing network design and frequency *must* minimize the total airline and passenger cost. Adjustment of utilities by pricing in ways different from (1-4) is an important research issue that is beyond the scope of this paper. This analysis is complex because it requires study of the airline’s demand function as well as its cost structure.

In the following section we develop analytic expressions for passenger and airline costs.

2. PASSENGER AND AIRLINE COST

This section analyzes passenger and airline cost. All costs are computed as the cost per day. We begin with passenger cost. Passenger costs can be grouped into two categories: schedule delay cost and travel cost. The total passenger schedule delay cost per day is the average schedule delay for all passengers multiplied by the schedule delay cost per hour, ω :

$$\frac{24(n)(n-1)\rho(\omega)}{2f}.$$

Passenger travel cost is the total time for all trips times the passenger’s cost per hour, ν . The time taken by any trip is made up of the following components.

- Expected time required by passengers for boarding and takeoff time.
- Delay time in takeoff.
- Flying time between cities.
- Expected time required for landing and unloading passengers.
- Delay time in landing.

The in-flight time is given by the flying distance divided by the velocity of the aircraft. This paper assumes that there can be random delays associated with takeoff and landing, but to compensate for these delays, airlines incorporate *planned delay time* into the schedule.

Planned delay time is extra time that is added to published schedules in excess of the expected flying and ground time due to delays in takeoff and landings. The purpose of planned delay time is to ensure that delays do not cause disruptions in the published schedule. Planned delay time increases passengers’ travel related costs and airline cost because it increases passenger travel time and the number of aircraft required to meet the schedule. Airlines in the United States average approximately 80% on-time performance, which is defined as arriving within 15 minutes of scheduled arrival time.

Planned delay time adds time to the published schedule, and we assume that it increases the perceived passengers’ travel time of a routing. Passengers at the time of purchase evaluate service quality based upon the published schedule. Also, air and ground controllers sometimes forbid planes gate access until the scheduled arrival time. Therefore, the planned delay becomes actual built-in time.

In the following, passenger travel time components are grouped into three categories:

- In-flight time.
- Ground time, which includes the time required for planes taking off and landing and the time required for passengers to board and leave the airplane.
- Planned delay time included in the schedule as a buffer against random delays.

A passenger's travel cost is his/her travel time times the hourly value of his/her time. Later, passenger's cost of real delay, which is the cost of arrival after the scheduled time, is considered. However, for simplicity, this cost is ignored for the time being.

Airline costs are partially fixed and partially dependent upon number of passengers transported and distance traveled. Airline costs are grouped as follows.

- *Fixed aircraft cost.* This is a function of the number of aircraft and the capacity of the aircraft, but not flying time. This cost includes fleet lease cost, but not operating cost.
- *Variable aircraft cost.* This cost is a function of the capacity of aircraft and the distance flown and includes the expense of fuel, maintenance, and crewing.
- *Per passenger cost.* This cost is directly driven by the number of passengers flown and includes the expense of reservations, ticketing, passenger ground service, and on-board meals.

Empirical data of Figures 1 and 2 are used as an estimate of the fixed cost of owning aircraft and variable costs per seat-mile. For simplicity, per passenger costs are ignored. This is realistic even though there are costs to transfer passengers and baggage at the hub, and customers may be served more than one meal on a routing. Kanafani and Ghobrial (1985) report CAB data on airline station cost, which is the cost of serving passengers at any type of station. (These costs include some expenses that hubs do not have such as that of ticketing transferring passengers. Therefore, they are higher than experienced at hubs.) The data show that airline station cost ranges from \$4 to \$22 per passenger, with a mean of \$10 (approximately). For realistic problems, this cost is only a small fraction of total cost and is ignored. Because the same number of passengers travel in each of the network designs, other per passenger costs for reservations, ticketing, and a meal are ignored. Although a passenger may be served more than one meal in a routing, the incremental cost of additional meals is small compared to other cost and is ignored. Although we do not consider these costs, they are easily added to the model.

The following subsections study these costs. Our cost analysis leads to the following conclusions. Passenger cost typically increases in k until $k > n/2$ and then decreases for $k = n$, with direct service ($k = 0$) having the lowest passenger cost. It is more difficult to make clean statements about how airline cost depends on networks. We show that drivers of airline costs are total seat-miles flown and the total number of airline seats required to provide sufficient capacity. Both are monotone increasing in k (at least up to some $k < 0.5n$). The airline's cost is sum of the total seat-miles multiplied by the airline's variable cost per seat-mile, plus the total number of seats required multiplied by the fixed cost per seat. But variable cost per seat-mile and fixed cost per seat are inversely related to the total number of seat-miles and total seats required, respec-

tively. Variable cost per seat-mile and fixed cost per seat fall with k , but total-seat miles and total number of seats required rise with k . The airline's cost (which is the sum of these products) demonstrates no general pattern as a function of k . Despite this, more can be said about the effect of k on total cost if factors such as the market radius, passenger demand rates, number of cities served and the frequency of service are examined. This is done in Sections 3 and 4.

2.1. Passenger Costs and Passenger In-Flight Time Cost

Next we study the passenger costs, for which we adopt the notation:

- $F(k)$ — Cumulative flying time per day over all passengers for network k ,
- $G(k)$ — Cumulative ground time per day over all passengers for network k ,
- A_α — Planned delay time at a spoke city to ensure a probability of delay of no more than α ,
- B_α — Planned delay time at the hub to ensure a probability of delay of no more than α ,
- $D(k)$ — Cumulative planned delay time per day over all passengers for network k .

For cities on the same subtour, in-flight time is the time required to travel the circumference of the circle between the cities; for cities on different subtours, in-flight time is that required to travel to the hub plus the time to travel from the hub to the destination city. For cities directly connected, in-flight time is the time to travel on the cord connecting the cities. Appendix A.1 reports the total passenger in-flight time as a function of network, $F(k)$, which is summarized in Table I.

This analysis shows that passenger in-flight time increases with k until $n/2$ and then decreases for $k = n$. Passenger in-flight time is least for direct routing, that is, $k = 0$. By examining the partial derivative of F with respect to k , and comparing the values of $F(1)$ and $F(n)$, we can see that for $k > 0$, $k = 1$ minimizes total in-flight time. In fact, the derivative of F is

$$\frac{\partial F}{\partial k} = \frac{2\pi R\rho}{v} \left\{ n + 1 - \frac{4k}{3} \right\} - \frac{2R}{v} \rho n,$$

and F increases with respect to k if $0 < k \leq 0.51n + 0.75$. Using the small angle approximation for $\text{Cot}(x)$ ($= \frac{1}{x}$), if n is moderately large, $F(1)$ is about 1.57 times $F(0)$ and $F(n)$ is about 1.64 times $F(0)$. Figure 4 shows an example of F . F increases with k until $n/2$ and then decreases for $k = n$.

2.2. Passenger Cost: Ground Time

Planes require time for landing and unloading passengers as well as for loading and takeoff. The total ground time in a trip between two cities with one intermediate stop is $2g$. Appendix A.2 computes the total passenger ground time, $G(k)$, as a function of network and is summarized in Table I.

Table I
Summary of Passenger and Airline Costs as a Function of Network Design

Network Attribute/ Network Type	Direct $k = 0$	Hub and Spoke and Subtour $1 \leq k < n$	Tour $k = n$
Schedule delay cost (\$/day)		$\frac{24(n)(n-1)\rho(\omega)}{2f}$	
Flying time, $F(k)$ (per day)	$\rho n \frac{2R}{v} \text{Cot}\left(\frac{\pi}{2n}\right)$	$\frac{2\pi R \rho}{v} \left\{ \frac{k^2-1}{3} + (k-1)(n-k) \right\} + \frac{2R}{v} \rho n(n-k)$	
Ground time, $G(k)$ (per day)	$\rho n(n-1)g$	$\rho \frac{ng}{3} (k^2-1) + \rho ng(n-k)(k+1)$	
Planned delay time, $D(k)$ (per day)	$\rho n(n-1)A_\alpha$	$\rho A_\alpha \left\{ \frac{n(k^2-1)}{3} + n(n-k)k \right\} + \rho B_\alpha n(n-k)$	
Required size of an airplane, $\text{SIZE}(k)$	$\frac{\rho}{f}$	$\frac{k(n-k)\rho}{f}$	$\frac{n^2\rho}{4f}$
$N(k)$	ρ	$k(n-k)\rho$	$n^2\rho/4$
Total Seat Miles, $\text{SM}(k)$ (per day)	$2R\rho \left(n \text{Cot}\left(\frac{\pi}{2n}\right) + \pi \right)$	$2R\rho \left(\frac{\pi}{n} 2(k-1) + 1 \right) n(n-k)$	$\pi R\rho \frac{n-1}{n} n^2$
Travel time per trip, $T(k)$ (days)	$\frac{2nR}{v} \text{Cot}\left(\frac{\pi}{2n}\right) + \frac{2\pi R}{v}$ $+ n(n-1)(g + A_\alpha)$	$\frac{4\pi R}{nv} (k-1) + \frac{2R}{v}$ $+ (2k-1)g + (2k-1)A_\alpha + B_\alpha$	$\frac{4\pi R}{nv} (n-1) + 2g(n-1)$ $+ 2A_\alpha(n-1)$
Total Seats, $\text{TS}(k)$ (per day)		$(n/k) \text{SIZE}(k) \lceil fT(k) \rceil$	

This analysis shows that ground time increases with k until $n/2$ and then decreases for $k = n$. Total ground time is least for direct service ($k = 0$). Examining the partial derivative of G with respect to k , and comparing $G(0)$ and $G(1)$ we can see that for $k > 0$, G is minimized by $k = 1$. The derivative is

$$\frac{\partial G}{\partial k} = \rho ng \left(\frac{2k}{3} + n - 2k - 1 \right),$$

which is positive for $k < 0.75n - 0.75$. $G(1)$ is twice $G(0)$ and $G(n)$ is $(n + \frac{1}{3})$ times $G(0)$. Figure 4 shows the average ground time as a function of k and shows that G is monotone increasing until $k = \frac{n}{2}$ and then decreases for $k = n$.

2.3. Passenger Cost: Planned Delay Time

Random delays occur during takeoff and landing. To guard against delays, airlines build time buffers into their schedules, so that there is only a small probability of real delay, α . Later we describe how α is selected.

It is assumed that for takeoff or landing, delay time is exponentially distributed with mean $\frac{1}{\lambda}$ and that delays are independent. In reality, weather- and congestion-related delays can be dependent events. However, the independence is an approximation that greatly simplifies the analysis. Also, more heavily used airports have longer delays.

This is typical for hubs. However, for simplicity, we ignore this issue for now and assume the delay time has the same distribution at all airports. We do study of the effect of hub congestion on networks in Section 6. Thus, the delay in arrival is the sum of two exponential random variables, and is Erlang(2, λ) distributed. At the hub, we assume that all flights wait until the last flight arrives before they exchange passengers. Hence the total delay when taking off from a spoke city, landing at the hub and then exchanging passengers is the maximum of m ($= n/k$), Erlang(2, λ) distributed random variables. The delay in taking off at the hub and landing at a spoke city is Erlang(2, λ) distributed. Let B_α be the number satisfying Probability $[Y \geq B_\alpha] = \alpha$, where Y is the maximum of m independent Erlang(2, λ) random variables, and let A_α be the number satisfying Probability $[X \geq A_\alpha] = \alpha$, where X is an Erlang(2, λ) random variable. Appendix A.3 reports the total planned delay time, $D(k)$, and is summarized in Table I.

Generally, direct routing has the lowest planned delay time of any network. It can be shown that for moderately large values of n ($n > 10$), and $\alpha = 0.2$, the ratio of planned delay time for hub and spoke to direct routing is about 2.5. The expression for $D(k)$ shows that planned delay time for the tour structure is approximately $(n + 1)/3$ times that for direct routing. Figure 4 shows the

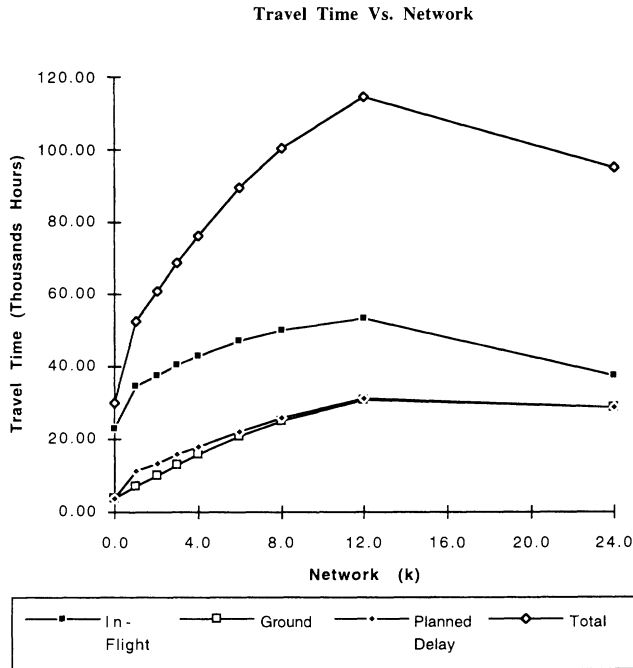


Figure 4. Total passenger travel time versus the network. For this example, $R = 500$ miles, $\rho = 25$ passengers/day, $v = 400$ mph, $g = 0.25$ hours, $\alpha = 0.2$, $\lambda = 12$, $f = 12$ dispatches, and $n = 24$ cities.

planned delay time for a typical problem: planned delay is monotone in k until $k = n/2$, declining for $k = n$. Figure 5a displays a typical pattern of delay time as a function of α and the network: for any α , planned delay time increases with k until $k = n$ (which is a tour). However, for problems with only a few cities, the planned delay time for a tour can be smaller than for hub and spoke; see Figure 5b. It can be analytically shown that the planned delay time is always monotone in k up to $k = n/2$. Figure 6 reports the critical value of α such that for lower values of α , the hub and spoke has less planned delay time than the tour. For reasonable values of n , the threshold value of α is much larger than that would be acceptable in practice. Thus from a perspective of planned delay time, direct service dominates the hub network and hub routing dominates tour and subtour networks.

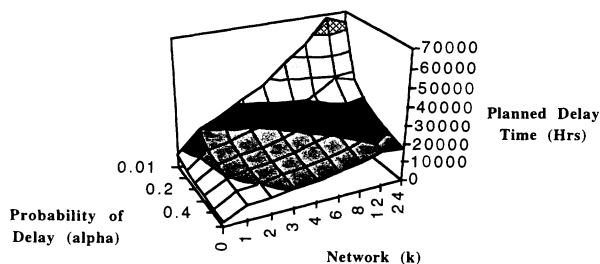


Figure 5a. Planned delay time versus probability of delay and the network. For this example, $n = 24$ and $\lambda = 12$, and $\rho = 25$ passengers/day.

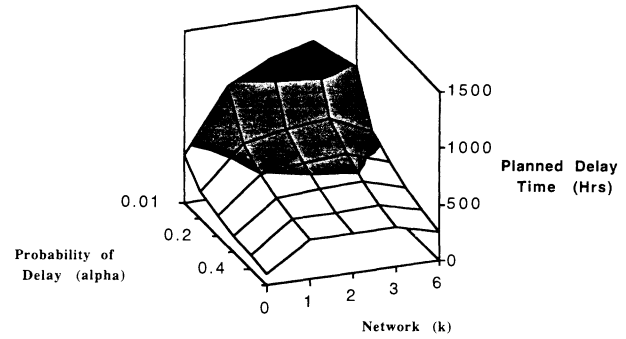


Figure 5b. Example where the tour network has lower planned delay time than hub and spoke for any α . For this example, $n = 6$ and $\lambda = 12$, and $\rho = 25$ passengers/day.

Critical Value of Alpha As A Function of the Number Of Cities Served

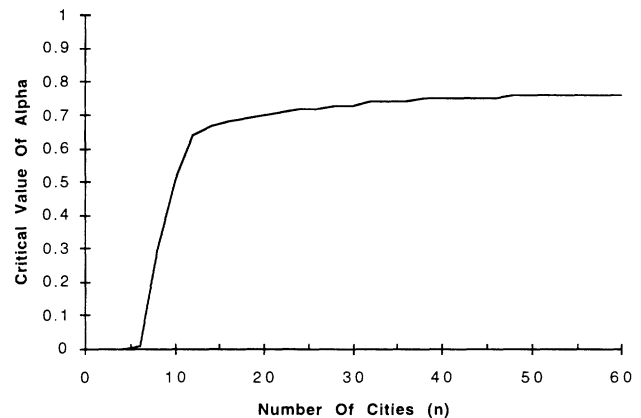


Figure 6. Critical value of α versus number of cities served. For each value of n , if α is less than the critical value of α , then hub and spoke minimizes planned delay time for the networks for all k such that $0 < k \leq n$.

The data of Figure 4 show that the percentage of in-flight time for direct service is substantially higher than for the other networks. For this example, in-flight time is 76% of scheduled time, while it is only 64% for hub and spoke and 39% for tour. This is consistent with published reports that direct carriers such as Southwest Airlines have higher in-flight time percentages than others. However, due to airline costs, we will show that direct service is not always the most efficient network.

2.4. Airline Cost and Variable Aircraft Cost

In the following subsections we study the airline cost and we adopt the following notation:

- SIZE(k) — Number of required seats for each aircraft in network k ,
- f^* — Optimal frequency per day,
- $N(k) = f(\text{SIZE}(k))$ — A variable defined for convenience,

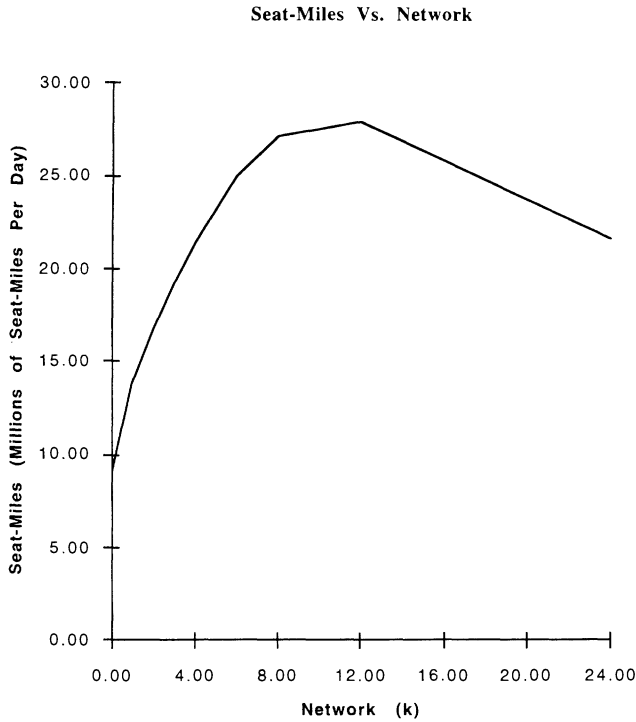


Figure 7. Total seat-miles flown versus the network. For this example, $R = 500$ miles, $\rho = 25$ passengers/day, $v = 400$ mph, $g = 0.25$ hours, $\alpha = 0.2$, $\lambda = 12$, $f = 12$ dispatches, and $n = 24$ cities.

- $SM(k)$ — Cumulative seat-miles flown per day by all aircraft in network k ,
- $T(k)$ — Total time for an aircraft to complete its route in network k ,
- $TS(k)$ — Total airline seats needed to meet the schedule in network k .
- $VC(k)$ — Airline variable cost for network k .

We begin with a study of the airline variable cost. The airline variable cost is the product of the total seat-miles flown and the cost per seat-mile, which is a function of the seat-size of the aircraft. The required aircraft size, $SIZE(k)$ is derived in A.4 and is summarized in Table I. It will be convenient to define $N(0) = \rho$; $N(k) = k(n - k)\rho$ for $0 < k < n$, and $N(n) = n^2\rho/4$, so that $SIZE(k) = N(k)/f$.

The expression shows that required aircraft size is monotone increasing in k . Thus, if the cost per seat-mile is as described in Figure 1, then the cost per aircraft seat-mile decreases with k .

A computation in Appendix A.4 finds the total seat-miles flown, $SM(k)$ as a function of network and is reported in Table I. Total seat-miles flown increases as a function of k and then decreases. Examining the expression for seat-miles as a function of k , $SM(0)$ is the fewest number of seat-miles, $SM(1)$ is the next at about 1.56 times $SM(0)$, and $SM(n)$ is about 2.45 times $SM(0)$. Differentiating $SM(k)$ with $0 < k < n$ with respect to k , it is seen

that the number of seat-miles flown increases with respect to k , up to $k = 0.3408n + 0.5$, and then decreases. Figure 7 reports the total seat-miles flown versus the network for a sample problem. Seat-miles are monotone for k up to $k = n/2 = 12$ and then decline. Seat-miles are minimized for direct service.

The network that minimizes the airline variable cost is, however, not necessarily direct routing. Airline variable cost is the product of the cost per seat-mile and the total seat-miles and is denoted by $VC(k)$. Assuming (2-1), $VC(k) = (a + bf/N(k)) SM(k)$. Comparing the airline variable costs for $k = 0$ and $k = 1$,

$$\frac{VC(1)}{VC(0)} = \frac{(a\rho(n-1) + bf)n}{(a\rho + bf)\left(\frac{2}{\pi}n^2 + \pi\right)},$$

which tends to

$$\frac{\pi}{2} \frac{a\rho}{(a\rho + bf)} \quad \text{for large } n.$$

Substituting estimates of a and b from Figure 1, $VC(1)$ is less than or equal to $VC(0)$, whenever $\rho/f \leq 344.47$. Thus, for aircraft sizes of 300 seats or less, airline variable cost for hub routing will always be smaller than direct routing. For $k > 1$, the derivative of $VC(k)$ with respect to k is given by

$$2R(11.25 + 9.44n - 22.49k)\rho + 703.86/k^2(2\pi - n)f.$$

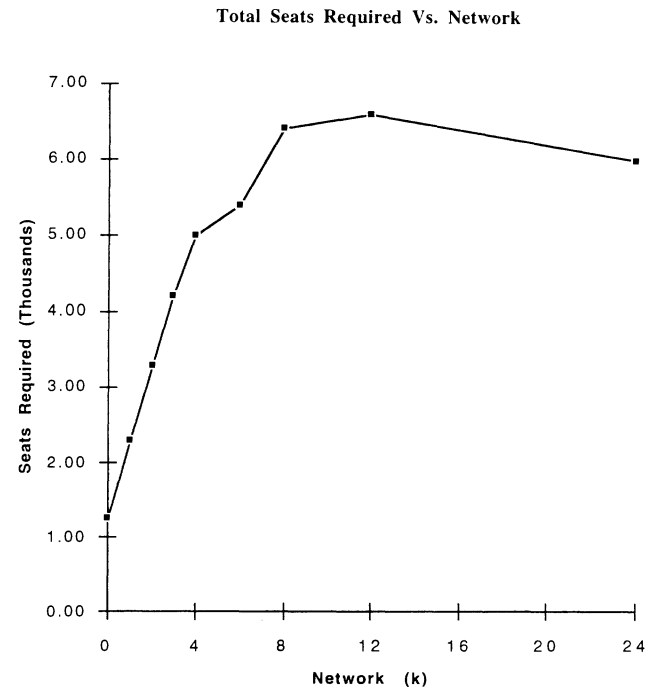


Figure 8. Aircraft capacity required versus the network. In this example, the hub and spoke system minimizes capacity cost. For this example, $R = 500$ miles, $\rho = 25$ passengers/day, $v = 400$ mph, $g = 0.25$ hours, $\alpha = 0.2$, $\lambda = 12$, $f = 12$ dispatches, and $n = 24$ cities.

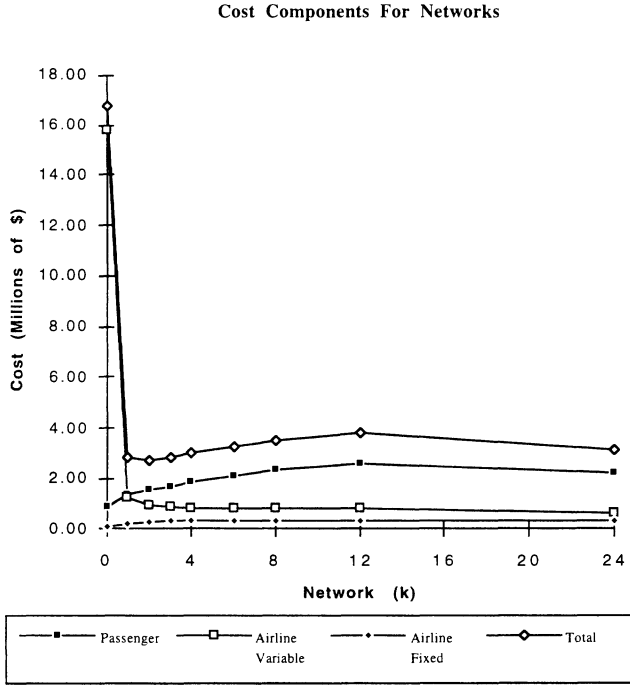


Figure 9. Total cost versus the network when frequency is fixed. The network $k = 2$ minimizes total cost. For this example, $\nu = \$20/\text{hour}$, $\omega = \$23/\text{hour}$, $R = 500$ miles, $\rho = 25$ passengers/day, $\nu = 400$ mph, $g = 0.25$ hours, $\alpha = 0.2$, $\lambda = 12$, $f = 12$ dispatches, and $n = 24$ cities.

The sign of the derivative depends on ρ , f , and k . The first term is linear monotone decreasing in k , while the second term is negative and monotone increasing in k for $n > 2\pi$. Together, the sum can be always negative (if $f \gg \rho$), or positive and then negative as k increases (if $\rho \gg f$), or negative then positive then negative as k increases (for intermediate relative values of f and ρ). For the example of Figure 9, $VC(0) > VC(1)$. The derivative of VC is negative for $1 < k < 6$, then at $k = 6$ and 8 it is positive and it is negative again for $k > 8$. The minimum value of VC is at $k = 24$. Figure 9 shows airline variable cost for a sample problem.

2.5. Airline Cost: Fixed Aircraft Cost

The fixed aircraft cost is the fixed cost per seat-day times the total number of required aircraft seats. The latter is the product of the required size of an individual aircraft and the number of planes needed to meet the schedule.

If a plane's scheduled time from departure from a spoke city to its return is T (possibly fractional) days and this time is greater than $1/f$, then additional planes are required to meet the schedule. The travel time for one full trip is denoted $T(k)$ and is found by summing up the time for travel, ground and delay time to complete the trip:

$$T(k) = \begin{cases} \frac{2nR}{v} \cot\left(\frac{\pi}{2n}\right) + \frac{2\pi R}{v} + n(n-1)(g + A_\alpha) & \text{if } k = 0, \\ \frac{4\pi R}{nv} (k-1) + \frac{2R}{v} + (2k-1)g & \text{if } 1 \leq k < n, \\ \frac{4\pi R}{nv} (n-1) + 2g(n-1) + 2A_\alpha(n-1) & \text{if } k = n. \end{cases} \quad (2-1)$$

Total aircraft seats needed is the product of the number of airplanes needed and the seats required for each aircraft. In the T days that one plane needs to complete its route and return to the point of origin, $\lceil fT \rceil$ (the least integer greater than fT) dispatches have to be made. Thus, the airline requires $\lceil fT \rceil$ planes for a direct network and $m\lceil fT \rceil$ planes for other networks, where $m = n/k$. Thus the total seats needed per day are

$$TS(k) = m \text{SIZE}(k) \lceil fT(k) \rceil = m \frac{N(k)}{f} \lceil fT(k) \rceil. \quad (2-2)$$

If (fT) is not small, $\lceil fT \rceil$ can be approximated by fT . The error in computing TS will be a small upward rounding error, and the average percentage error will be less than $0.5/fT(k) \times 100\%$. Using this approximation, we have

$$TS(1) - TS(0) = \rho(R/V)((2 - 4/\pi)n^2 - 2n + 2\pi) + \rho n(n-1)B_\alpha,$$

which is greater than 0 for $n \geq 2$. In other words, $TS(0) < TS(1)$ for $n \geq 2$. In addition, for $k \geq 1$, the partial derivative of $TS(k)$ with respect to k is given by

$$\rho n \left\{ \frac{R}{V} \left[(4\pi - 2) + \frac{4\pi}{n} (1 - 2k) \right] + (2n - 4k + 1)(g + A_\alpha) - B_\alpha \right\}. \quad (2-3)$$

$TS(k)$ is an increasing function of k at least until $0.42n$. We can also show that $TS(n)/TS(1)$ tends to

$$\left(\frac{n}{2}\right) \left(\frac{g + A_\alpha}{g + A_\alpha + B_\alpha} \right),$$

which is greater than 1. Thus $TS(n) > TS(1)$. Figure 8 reports $TS(k)$ as a function of k ; its behavior is as predicted.

As in the case of aircraft variable cost, it is not clear which network minimizes total fixed cost because the cost per seat depends upon aircraft size (see Figure 2). Networks with smaller required aircraft size may have a greater number of required seats. Since an aircraft's fixed cost decreases with size there may be a tradeoff: aircraft fixed cost per seat decreases, but the number of seats increases.

Next, the effect of the cost components on the optimal frequency of dispatch is considered.

3. OPTIMAL FREQUENCY

The optimal frequency for a network is the f that minimizes total cost for that network. Frequency affects customer schedule delay cost, airline variable and airline fixed

cost. Holding k fixed, higher frequency lowers customer schedule delay cost but raises airline costs. Airline costs increase because aircraft are smaller, and thus are more expensive per seat-mile. Thus, analysis of the optimal frequency involves a conflict between passenger and airline costs. On the other hand, the total cost minimizing frequency differs as a function of k because customer schedule delay cost, airline variable cost, and airline's fixed cost are also functions of k . This section studies the interaction of network and frequency, and properties of the optimal frequency.

If integrality of planes is ignored, the optimal frequency is found by differentiating the sum of these costs, and solving for a stationary point. Schedule delay cost is $\frac{12}{f} \omega(n)(n-1)\rho$. Fixed cost is $TS(k)[c - dN(k)/f]$, and variable cost is $SM(k)[a + b f/N(k)]$. The sum is

$$\frac{12}{f} \omega(n)(n-1)\rho + TS(k) \left[c - d \frac{N(k)}{f} \right] + SM(k) \left[a + b \frac{f}{N(k)} \right].$$

Note that only some of the terms depend on f . We call the sum of the terms that depend on f *frequency related cost*, $FRC(k, f)$ and

$$FRC(k, f) = \frac{12}{f} \omega(n)(n-1)\rho - TS(k)d \frac{N(k)}{f} + SM(k)b \frac{f}{N(k)}.$$

Differentiating $FRC(k, f)$ with respect to f and solving for the stationary point in f ,

$$f^* = \sqrt{\frac{12\omega(n)(n-1)\rho - TS(k)dN(k)}{\frac{SM(k)b}{N(k)}}}. \quad (3-1)$$

Using this optimal frequency,

$$FRC(k, f^*) = \sqrt{4(12\omega(n)(n-1)\rho - dTS(k)N(k)) \frac{bSM(k)}{N(k)}}.$$

Making some realistic assumptions simplifies the analysis of f^* . Our numerical experiments suggest that for most realistic values, $dN(k)TS(k) \ll 12\rho\omega(n-1)$. Thus, the term $dTS(k)N(k)$ within the radical can be ignored, so that f^* is proportional to:

$$f^* \propto \sqrt{\frac{n(n-1)N(k)}{SM(k)}} = \begin{cases} \frac{1}{\sqrt{\frac{4Rn}{\pi(n-1)} + \frac{2\pi R}{n(n-1)}}} & \text{if } k = 0, \\ \frac{1}{\sqrt{\frac{2R}{(n-1)k} + \frac{4\pi R}{n(n-1)} \frac{k-1}{k}}} & \text{if } 0 < k < n, \\ \frac{n}{\sqrt{4\pi R}}, & \text{if } k = n. \end{cases} \quad (3-2)$$

For any network, service frequency is monotone, convex decreasing in R and monotone concave increasing in n . (Note that $n \geq 2$). As distance R becomes larger, frequency declines, and as the number of cities served increases, the frequency increases for all networks.

The intuition is as follows. Increasing R raises the relative importance of variable airline cost in the sum of total costs. This encourages the use of larger aircraft, which can be done by reducing frequency. Why does f^* increase with n ? Assume that f^* is the optimal frequency for n^* cities. Now, hold f^* fixed and increase the number of cities served. This increases the number of passengers on aircraft for any $k > 0$. At these higher volumes there are decreasing economies of scale in variable cost compared with the initial n^* . This encourages the airline to trade off lower passenger cost for higher variable cost by increasing frequency.

The optimal service frequency for direct service is lower than for hub and spoke, and f^* is monotone decreasing in k . In particular, the ratio of optimal service frequency of hub and spoke to direct service is

$$\frac{f^*(1)}{f^*(0)} = \sqrt{\frac{2n^2/\pi + \pi}{n}}, \quad (3-3)$$

or $\sqrt{2n/\pi}$ as n becomes large.

The relationship between k and f^* can be understood as follows. Holding f fixed, as k increases the required size of aircraft ($SIZE(k)$) increases. This lowers the average cost of transport per mile. The number of seat-miles increases with k , it is optimal to raise f : there are diminishing economies of plane size compared with the gains of reducing passenger schedule delay cost. Total cost is minimized by reducing passenger schedule delay cost by increasing frequency.

The frequency related cost, $FRC(k, f^*(k))$ with optimal frequency f^* is:

$$FRC(k) = \begin{cases} \sqrt{4(12\omega(n)(n-1)\rho - dTS(k)N(k)) \left(\frac{4Rn^2}{\pi} + 2\pi R \right) b} & \text{if } k = 0, \\ \sqrt{4(12\omega(n)(n-1)\rho - dTS(k)N(k)) \left(2R \frac{n}{k} + 4\pi R \frac{k-1}{k} \right) b} & \text{if } 0 < k < n, \\ \sqrt{4(12\omega(n)(n-1)\rho - dTS(k)N(k)) \left(4\pi R \frac{n-1}{n} \right) b} & \text{if } k = n. \end{cases} \quad (3-4)$$

Costs With Optimal Frequency

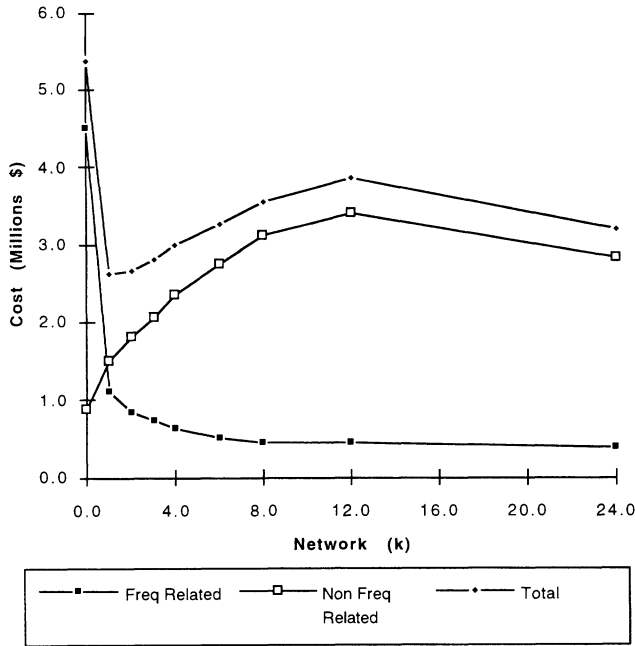


Figure 10. Total cost versus the network when optimal frequency is used. The parameters are the same as in Figure 9; now hub and spoke minimizes total cost.

Frequency related cost is a concave and increasing function of schedule delay cost, ω . If $12\omega(n)(n-1)\rho \gg dTS(k)N(k)$, simple differentiation shows that frequency related cost is monotone *decreasing* and convex in k when $n > 2\pi$. Thus, as customers' taste for convenient departure increases (ω rises), the k that minimizes total cost increases. At an extreme, as ω becomes very large, the tour network minimizes frequency related cost. A referee has pointed out to us that a recent startup, Vanguard Airline, schedules some of its aircraft in a tour.

Figure 10 plots frequency related cost, nonfrequency related cost (which is just total cost minus FRC), and the total cost versus the network as parameterized by k . As expected, FRC is a convex decreasing function of k , and nonfrequency cost is a concave single peaked function of k . The total cost is neither concave nor convex, and the minimum occurs, for this example at an interior point. The optimal frequency ranges from $f^* = 3$ for $k = 0$, to $f^* = 12$ for $k = 1$, to $f^* = 15$ for $k = 24$.

4. EFFECT OF RADIUS, DEMAND RATES, AND NUMBER OF CITIES ON OPTIMAL NETWORKS

Next we study the effect of distances between cities, demand rates between cities, and the number of cities on the network that minimizes total cost, the *optimal network*.

For any network, total cost minimizing frequency is used, so that

$$\begin{aligned} \text{Total Cost} = & \text{FRC}(k) + \nu(G(k)) + D(k) \\ & + \{a\text{SM}(k) + c\text{TS}(k) + \nu F(k)\}. \end{aligned}$$

The first term is *frequency related cost*. The next two terms are the passengers' cost of ground time and planned delay time. Their sum will be referred to as *fixed cost* because they are fixed with respect to the radius, R . The terms within the curved brackets sum components of airline variable cost, components of airline fixed cost, and passenger in-flight cost, and are referred to as *other cost*.

Summarizing results from Sections 1 and 2, *fixed cost* and *other cost* are monotone *increasing* in k , at least for $k < 0.34n$. This is because travel time and total seat-miles decrease as k decreases. *Frequency related cost* is monotone *decreasing* in k . Thus, *fixed cost* and *other cost* cause the optimal k to fall, while *frequency related cost* makes the optimal k rise.

Frequency related cost increases as a function of $n^2\sqrt{R\rho}$ for $k = 0$, $n^{3/2}\sqrt{R\rho}$ for k such that $0 < k < n$, and $n\sqrt{R\rho}$ for $k = n$. Other cost increases as a function of $n^2R\rho$. Fixed cost increases as a function of $n^2\rho$. Therefore, for a fixed network, total cost is a concave increasing function of R , ρ , and n^2 . Thus, there are economies of scale in distance, demand rate, and number of cities served.

These results lead to important insights about airline strategies. For example consider the difference between a hub carrier and a direct carrier. To increase profits, the hub carrier will have great incentive to increase the number of cities served (compared to direct carriers) because its cost rises less (a factor of $n^{3/2}$ vs. n^2). The direct carrier must rely upon increasing demand and profit by reducing price.

Next, the effect of distance between cities as measured by the market radius is studied. When $R = 0$, total cost equals *fixed cost*. This cost is minimized by direct routing and is a single peaked function of k ; see Figure 8 for a graph of TS, which drives fixed cost. Because of *fixed cost*, a direct network always minimizes total cost for very short distances.

As R increases, the frequency related cost increases as \sqrt{R} and the other cost rises as R . As R increases, fixed cost becomes less important in total cost (which forced direct service to be the optimal solution when $R = 0$), so that the optimal k rises. As R increases further, *other cost* rises faster than the *frequency related cost*, so the k of the optimal network rises or the optimal network becomes a tour ($k = n$). In the example of Figure 11, as R increases the optimal network changes from $k = 0$, to $k = 2$ to $k = 3$ to $k = 24$. As is clear from this example, hub and spoke need not be the optimal network.

Similar parametric analysis of the effect of ρ on the optimal network is possible. When $\rho = 0$, all networks have zero total cost and are optimal. The *frequency related cost* increases as $\sqrt{\rho}$ while the remaining terms increase as ρ . For values of ρ less than 1, *frequency related cost* increases faster than the remaining terms so the optimal network minimizes *frequency related cost* and is a tour. As

Optimal Network As A Function of Radius

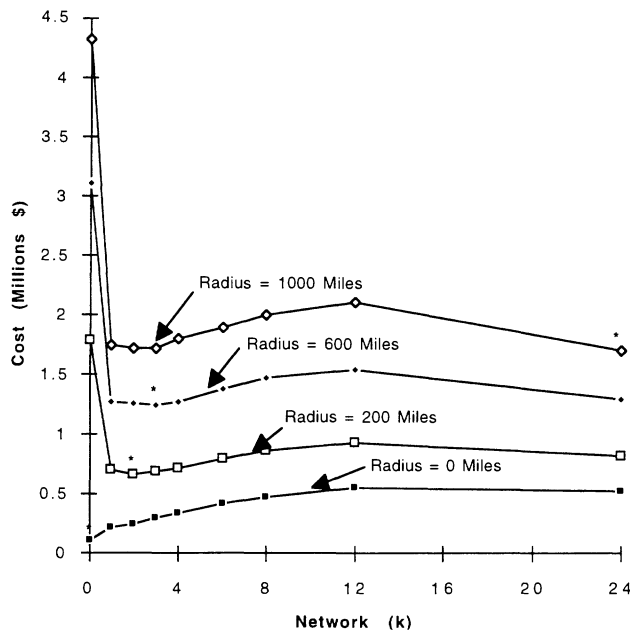


Figure 11. Total cost as a function of radius. For each radius, the total cost minimizing network is indicated with an asterisk. Except for the radius, the parameters are the same as for Figure 9.

ρ increases, the optimal k decreases as *other cost* and *fixed cost* become more important. When ρ is very large, the optimal network is direct service. Figure 12 is an example of these effects. For $\rho = 1$, a tour is optimal; for $\rho = 4$, the optimal design is $k = 2$; for $\rho = 15$, the optimal network is hub and spoke. For much larger values of demand rate, the optimal network is direct service. This result and the result of Section 1 help to explain routing for nonairline transportation markets. The tour is the network that minimizes cost when passengers' demand rates are low and passengers' schedule delay costs are high. This helps to explain the use of bus lines in mass transit systems, a network that is topologically equivalent to a tour as we have defined it (see Figure 3c).

The effect of the number of cities served, n , on the optimal network can also be considered. Suppose $R > 0$ and $\rho > 0$. For small n , *frequency related cost* increases slower than the other terms in n . For small values of n , the *frequency related cost* term will be relatively important in minimizing total cost. As n increases, it will be less important for $k > 0$ and the optimal k will fall. In the example of Figure 13, for $n = 6$ or less, the optimal network is a tour. As n increases, the optimal network becomes $k = 3$ and for $n = 24$, the optimal network is hub and spoke.

5. OPTIMAL CHOICE OF SCHEDULE RELIABILITY

Parameter α , the probability of delay on a single leg, is a measure of schedule reliability that has been held fixed

Optimal Network As A Function of Demand Rate

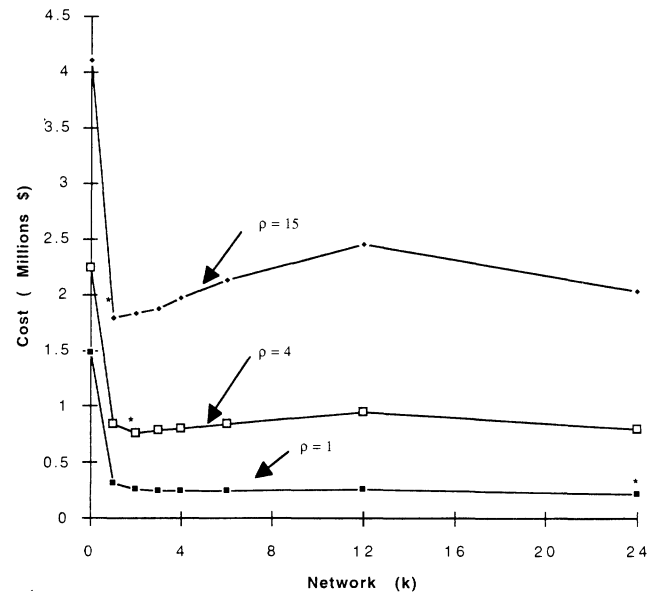


Figure 12. Optimal network as a function of demand rate. For each demand rate, the total cost minimizing network is indicated with an asterisk. Except for the demand rate, the parameters are the same as for Figure 9.

until this section. Now, the effect of network on choice of schedule reliability is studied.

Increasing schedule reliability (decreasing α) requires increasing planned delays which raises passenger's scheduled (and actual) travel time. Increasing planned delays also increases the time needed to complete trips. As a result, increasing schedule reliability raises the total aircraft required to meet the schedule, and thus the airlines' fixed cost. On the other hand, low schedule reliability means that the actual delay is often larger than the planned delay. When this happens, passengers arrive after the scheduled arrival time. In other words, passengers experience *real delay*.

For a fixed network, the optimal probability of delay is the α that minimizes delay related costs, which is the sum of passenger planned travel time cost, passenger real delay cost, and the airline's fixed cost of aircraft. To find the optimal α and to study the effect of network design on the optimal α , we calculated a passenger's expected real delay by simulating passengers actual delays for networks with different values of α , and k . As in the previous examples, $n = 24$ and $\lambda = 12$. Figure 14 shows these calculations for $k = 0$, $k = 1$, and $k = 24$. A conclusion is that the impact of schedule reliability on passengers' expected real delay depends on the network. For direct routing ($k = 0$), a passenger's expected real delay increases very rapidly with an increase in the probability of delay, but a passenger's expected real delay increases very slowly for hub and spoke ($k = 1$). All other networks were found to have

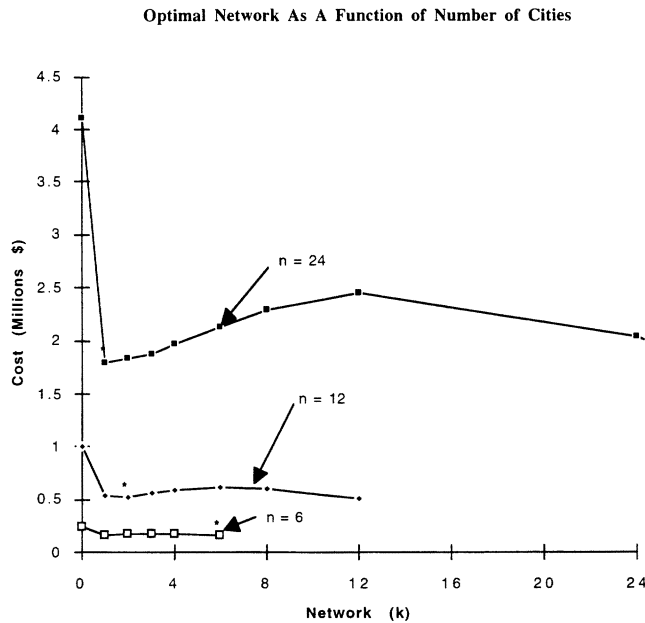


Figure 13. Optimal network as a function of number of cities served. For each number of cities, the total cost minimizing network is indicated with an asterisk. Except for the number of cities served, the parameters are the same as for Figure 9.

intermediate performance between the tour ($k = 24$) and the hub and spoke. What is striking is that a passenger's real expected delay in a hub and spoke system is relatively indifferent to the schedule reliability. Figure 14 shows that for hubs expected real delay does not increase much as α increases. This is because of two reasons: the nature of delays at the hub, and also the "buffering" that multiple legs on a trip provides. Delay at the hub is an order statistic as delays are caused by the last arriving flight. Increasing α does not cause the expected real delay to rise proportionally. Also, hub system trips have two legs that both buffer the schedule, so that the probability of real delay declines due to uncertainty pooling. Increasing α does not cause expected real delay to rise much due to this pooling.

Next, we consider optimal choice of α for different networks. Figure 15 shows delay related cost as a function of the probability of delay (α) for networks with $k = 0$, $k = 1$, and $k = n$. For this example, a passenger's cost for real delay was taken to be \$50/hour, which is significantly higher than passengers' travel time cost of \$20/hour. A direct routing's delay related cost is a different function of α than the other networks. For direct routing, the cost associated with passengers' real delay rises very rapidly, and becomes the dominant cost term, as the probability of real delay increases. In this example, the rapid rise in real delay causes the optimal value of α to be 0.15 for direct routing. On the other hand, for the other networks ($k = 1, \dots, n$), passenger scheduled travel time and airline fixed cost fall rapidly with α and are the dominant terms. As a result, the optimal value of α is much higher in these cases

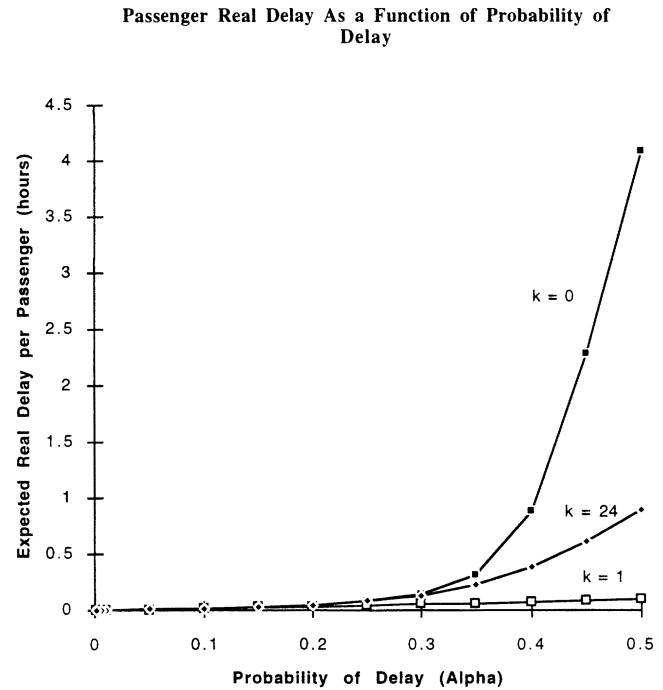


Figure 14. Passenger real delay as a function of probability of delay. Passengers' real delay rises more quickly for direct routing than for hub and spoke or for tour networks. In this example, $\lambda = 12$ and $n = 24$ cities.

($\alpha = 0.4$ for $k = 1$ and $\alpha = 0.3$ for $k = 24$). Note, too, that the airline's fixed cost falls rapidly with α , but then is nearly constant. This is because the airplane schedule has enough slack in it to support the schedule, but planes cannot be eliminated.

The optimal value of α was computed for all other values of k . Except for $k = 0$, the optimal α was always within the interval (0.3, 0.4). Because airline's cost is nearly constant except when the probability of delay is relatively small, the optimal value of α is approximately unchanged when α is chosen to minimize just the passengers' costs.

We conclude that the optimal value of α for direct routing is less than that for other networks. This result explains why non-hub airlines like Southwest Airlines have superior on-time arrival statistics compared with hub based carriers. Even though α is smaller for direct routing, the actual total planned delay time is less than that for other networks. For the example, planned delay time is 12% of the schedule for direct service (with $\alpha = 0.15$), while it is 19% for hub and spoke (with $\alpha = 0.40$)! Thus, direct carriers have higher reliabilities and higher capacity utilization of aircraft even ignoring hub's additional time for takeoff, landing, and ground services!

A manager for a major U.S. hub-based carrier told us that he believed that his firm's schedule reliability was too high and could economically be reduced. Our results show that this argument might be made based purely upon trading off passengers' costs and not the airline's. That is,

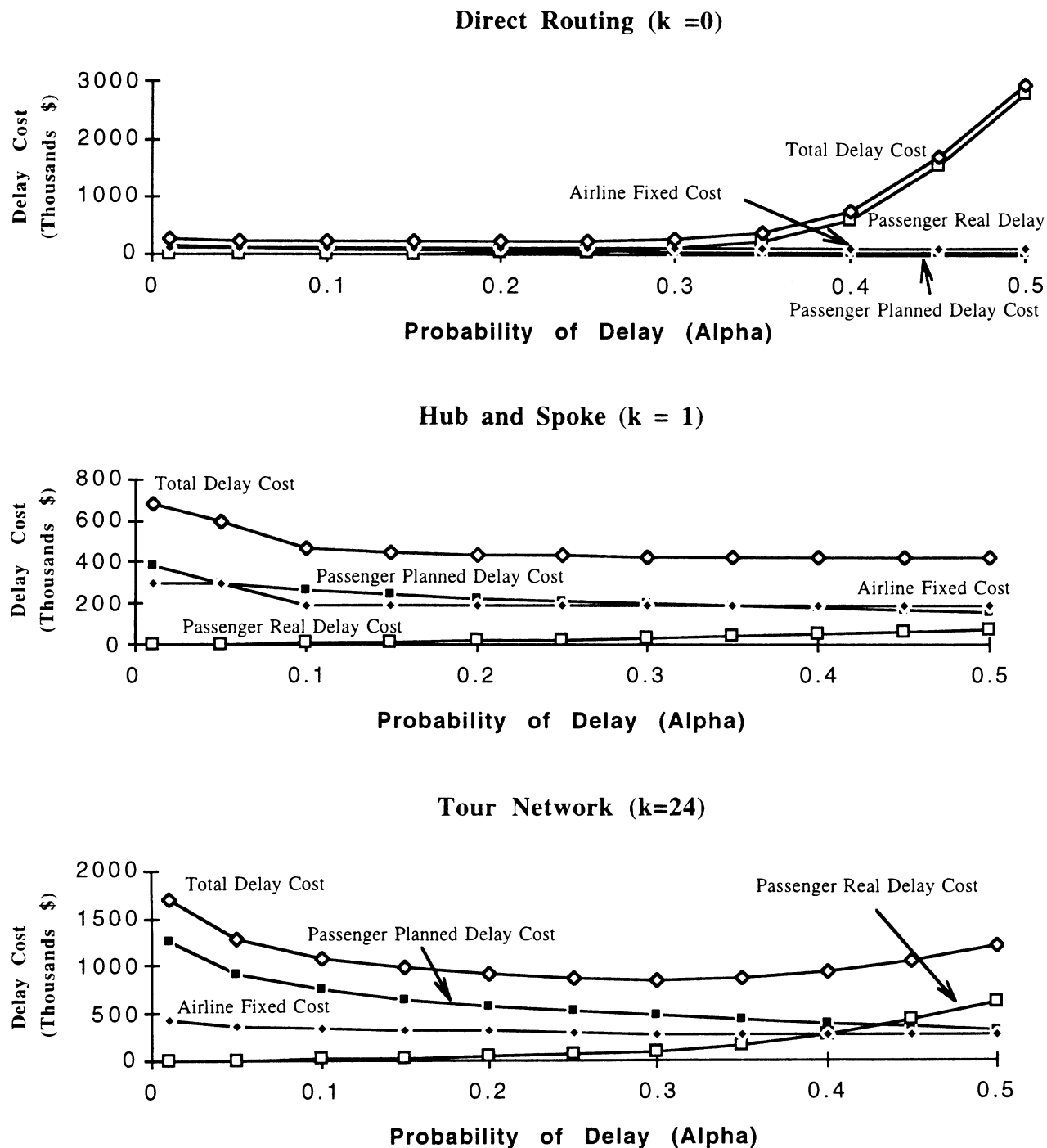


Figure 15. Total delay cost as a function of probability of delay. The alpha that minimizes delay cost depends on the network, and is at least for direct service. The data of this example are the same as for Figure 9, except that flight frequencies are optimally chosen.

passengers would prefer planned delay time to be removed from the schedule despite the cost of occasional delay. Or, said another way, marketing considerations could motivate schedules with larger delays because of the savings for *most* passengers in reduced travel times. Also, our analysis shows that it is not economically optimal to operate hub and spoke systems at too high reliability. The perceived decrease in airline timeliness over the last 15 years can be

partially explained by this analysis; increasing airport congestion is another major cause.

6. CONGESTION AT THE HUB

Congestion at hubs causes delays at hubs to exceed those at spoke cities. This section explores the effect of congestion at the hub on network choice. In particular we ask if

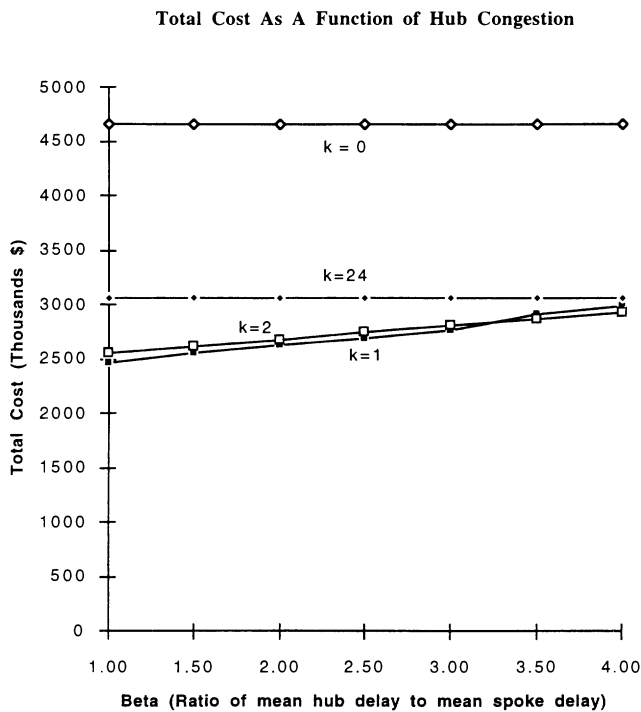


Figure 16. Total cost as a function of hub congestion. Total cost is computed for optimal values of f and α . Here the hub and spoke network is optimal for a beta less than 3.25. For larger beta, $k = 2$ is the optimal network. The data for this example are the same as for Figure 9.

long delays at the hub can cause the optimal network design to become a nonhub and spoke network. In the following analysis, the average delay during takeoff and landing at the hub is assumed to be some multiple of that at spoke airports. Specifically, delay at the hub is assumed to be exponential with a mean of β/λ , (where β is a constant greater than 1). In contrast, the delay during takeoff and landing at a spoke airport is assumed to be exponential with a mean of $(1/\lambda)$. The delay in arrival at the hub or arrival at a spoke city from the hub is the sum of two exponential distributions. If X' is the sum of 2 exponential random variables, one with a mean of $(1/\lambda)$, and the other with a mean of (β/λ) , the planned delay time required at a spoke city for a flight arriving from the hub is C_α , where Probability $[X' \geq C_\alpha] = \alpha$. Similarly, if Y' is the maximum of m independent random variables, each of which is the sum of two exponential random variables with means $(1/\lambda)$ and (β/λ) , the planned delay time required at the hub is D_α , where Probability $[Y' \geq D_\alpha] = \alpha$. Computations in A.5, compute expressions for C_α , D_α , the planned delay time and the other cost components that depend upon hub delay.

Next, an example shows that congestion at the hub generally does not change the optimal network except when congestion at the hub is unrealistically large. This is because delay related cost due to hub congestion is a small fraction of total cost. Figure 16 shows the total cost as a function of β for optimal values of f and α for an example

using data of Figure 9. When β is less than 3.25, the optimal network is the hub, for larger values $k = 2$ is best. The lesson is that congestion at the hub has little effect on optimal networks, except when hub congestion is very large.

7. CONCLUSIONS AND EXTENSIONS

This paper demonstrated how structural parameters such as the distance between cities, the demand rates and the number of cities served affects the optimal airline network. Results show that hub, direct, subtour, and tour networks can each be optimal for selected parameters. The effects of network choice on the optimal service frequency and schedule reliability are studied. Conclusions include that direct service has the lowest schedule frequency and highest schedule reliability of all network designs. On the other hand, hub and spoke networks have high optimal schedule frequency and low schedule reliability. Finally, the paper shows that congestion at the hub has relatively small effect on the optimal network design. This implies that even with increasingly congested hub airports, hub networks will continue to operate.

The paper focused on four type of networks, but we have also studied other variations, such as systems where close together cities are served *directly*, and other cities are served by a *hub and spoke system*. Analysis of this *fifth* network type yields results similar to those of the hub system.

Although this model considers airline economics, the approach can be used to consider other transportation networks such as common carrier, less than truck load and cargo airline. Alternate transportation networks require estimates of carrier fixed and variable cost, such as were derived in Figures 1 and 2. Some intuition about networks in other transportation modes is provided. An insight is that if fixed and variable costs exhibit constant returns to scale, then direct service is optimal.

Many factors are ignored to develop a tractable analytic model. Asymmetries in distances between cities and demand rates are not considered. Several cost components are omitted such as landing fees, and control costs to manage the network. Control costs include airline's costs to reschedule delayed flights. These costs are low compared to the components explored here but ought to be considered in a complete model. The flexibility of networks to expansion also has been ignored. A hub and spoke network can easily add additional cities without major changes to the rest of the schedule. A major unexplored issue is the effect of the cost structure on optimal pricing decisions. As demand increases, when and how an airline should switch from direct or tour-based structures to hub and spoke systems is another important research avenue. Another important issue is detailed analysis of the effects on the results of traffic originating/destined at the hub.

APPENDIX

In the following, refer to the i th city on the p th subtour as city (i, p) . The same notation can be used for hub and spoke networks (in that case, i will always be 1), and for tour networks (p will always be 1).

A.1. Passenger Costs: In-Flight Time

Let the in-flight time from city (i, p) to city (i', p') be given by $ft((i, p), (i', p'))$. Therefore, for all $i, i' = 1, \dots, k$:

$$ft((i, p), (i', p')) = \begin{cases} \frac{2\pi R}{nv} |i' - i| & \text{if } p = p' = 1, \dots, m, \\ \frac{2\pi R}{nv} (i - 1) + \frac{2R}{v} + \frac{2\pi R}{nv} (i' - 1) & \text{if } p, p' = 1, \dots, m, p \neq p'. \end{cases}$$

Summing in-flight time for all traffic, for $k > 0$,

$$\begin{aligned} F(k) &= \rho \sum_{(i,p)} \sum_{(i',p')} ft((i, p), (i', p')) \\ &= \rho 2m \sum_{i=1}^{k-1} \sum_{i'=i+1}^k ft((i, 1), (i', 1)) \\ &\quad + \rho m(m-1) \sum_{i=1}^k \sum_{i'=1}^k ft((i, 1), (i', 2)). \end{aligned}$$

For direct routing, the in-flight time between any two cities is the time required to travel along the chord that connects the two cities and is given by

$$ft(i, i') = \frac{2R}{v} \sin\left(\frac{|i' - i|\pi}{n}\right).$$

Summing up, we have for $k = 0$,

$$F(0) = \rho \sum_i \sum_{i'} ft(i, i') = \rho n \sum_{i'=2}^n ft(1, i').$$

The cumulative in-flight time for all passengers is,

$$F(k) = \begin{cases} \rho n \frac{2R}{v} \cot\left(\frac{\pi}{2n}\right) & \text{if } k = 0, \\ \frac{2\pi R \rho}{v} \left\{ \frac{k^2 - 1}{3} + (k-1)(n-k) \right\} \\ + \frac{2R}{v} \rho n(n-k) & \text{if } 0 < k \leq n. \end{cases}$$

The cotangent expression is the sum of cord lengths from one city to each of the other $n-1$ equispaced cities. (This result is found in Jolly 1961, as the of summation of series #436.)

A.2. Passenger Cost: Ground Time

For subtours the cumulative ground time for trips between cities (i, p) and (i', p') is,

$$gt((i, p), (i', p')) = \begin{cases} |i - i'|g & \text{if } p = p' = 1, \dots, m, \\ ig + i'g & \text{if } p, p' = 1, \dots, m, p \neq p'. \end{cases}$$

Summing up over all passengers,

$$\begin{aligned} G(k) &= \rho \sum_{(i,p)} \sum_{(i',p')} gt((i, p), (i', p')) \\ &= \rho 2m \sum_{i=1}^{k-1} \sum_{i'=i+1}^k gt((i, 1), (i', 1)) \\ &\quad + \rho m(m-1) \sum_{i=1}^k \sum_{i'=1}^k gt((i, 1), (i', 2)). \end{aligned}$$

For direct routings, the total ground time between any two cities is clearly g . Thus,

$$G(k) = \begin{cases} \rho n(n-1)g & \text{if } k = 0, \\ \rho \frac{ng}{3} (k^2 - 1) + \rho ng(n-k)(k+1) & \text{if } 0 < k \leq n. \end{cases}$$

A.3. Passenger Cost: Planned Delay Time

Let A_α be a number such that, Probability $[X \geq A_\alpha] = \alpha$, where X is an Erlang(2, λ) random variable. Also let, B_α be a number satisfying Probability $[Y \geq B_\alpha] = \alpha$, where Y is the maximum of m independent Erlang(2, λ) random variables. Thus, whenever an aircraft, lands at a spoke city, the planned delay time is A_α ; a landing at the hub requires planned delay time of B_α . Then for subtour, the planned delay time in traveling from city (i, p) to city (i', p') is denoted by the function dt :

$$dt((i, p), (i', p')) = \begin{cases} (i' - i)A_\alpha & \text{if } p = p' = 1, \dots, m, \\ (i - 1)A_\alpha + B_\alpha + i'A_\alpha & \text{if } p, p' = 1, \dots, m, p \neq p'. \end{cases}$$

Summing up over all passengers,

$$\begin{aligned} D(k) &= \sum_{(i,p)} \sum_{(i',p')} dt((i, p), (i', p')) \\ &= 2\rho m \sum_{i=1}^{k-1} \sum_{i'=i+1}^k (i' - i)A_\alpha \\ &\quad + \rho m(m-1) \sum_{i=1}^k \sum_{i'=1}^k (i - 1)A_\alpha + B_\alpha + i'A_\alpha. \end{aligned}$$

For direct routing, the planned delay time in traveling between any two cities is A_α . Therefore the cumulative planned delay time is,

$$D(k) = \begin{cases} \rho n(n-1)A_\alpha & \text{if } k = 0, \\ \rho A_\alpha \left\{ \frac{n(k^2 - 1)}{3} + n(n-k)k \right\} \\ + \rho B_\alpha n(n-k) & \text{if } 0 < k \leq n. \end{cases}$$

By the definition of A_α ,

$$\int_0^{A_\alpha} \lambda e^{-\lambda y} (1 - e^{-\lambda(A_\alpha - y)}) dy = 1 - \alpha,$$

Therefore, integrating and solving yields,

$$A_\alpha = \left(\frac{1}{\lambda}\right) \ln\left(\frac{1 + \lambda A_\alpha}{\alpha}\right)$$

and if

$$x = \ln\left(\frac{1 + x}{\alpha}\right),$$

then, $A_\alpha = \frac{1}{\lambda}x$. Parameter x can be computed by iterative substitution: assume an initial value for $x^0 > 0$ and set the value at the $j + 1$ th iteration at $x^{j+1} = f(x^j) = \ln((1 + x^j)/(\alpha))$. The sequence converges monotonically to x because $df(x)/dx < 1$ for any $x > 0$; see Fletcher (1987).

By the definition of B_α , $\text{Prob}[\text{max. of } m \text{ i.i.d. Erlang}(2, \lambda) \text{ random variables} \leq B_\alpha] = 1 - \alpha$. Therefore $(\text{Prob}[\text{Erlang}(2, \lambda) \text{ random variable} \leq B_\alpha])^m = 1 - \alpha$, or

$$B_\alpha = \frac{1}{\lambda} \ln\left(\frac{1 + \lambda B_\alpha}{1 - (1 - \alpha)^{(1/m)}}\right).$$

If

$$y = \ln\left(\frac{1 + y}{1 - (1 - \alpha)^{(1/m)}}\right),$$

$B_\alpha = \frac{1}{\lambda}y$. The value of y can be found by iterative substitution. Note that x is a function of α alone, while y is a function of α and m . Therefore,

$$D(k) = \begin{cases} \rho n(n-1) \frac{x}{\lambda} & \text{if } k = 0, \\ \rho \frac{x}{\lambda} \left\{ \frac{n(k^2 - 1)}{3} + n(n-k)k \right\} & \text{if } 0 < k < n, \\ + \rho \frac{y}{\lambda} n(n-k) & \text{if } k = n. \end{cases}$$

A.4. Airline Variable Cost: Seat-Miles Flown

The number of passengers traveling to each spoke city is ρ , spread over f dispatches. For direct routings, each plane requires ρ/f seats. In the hub and spoke network design, each dispatch serves $n - 1$ cities. Therefore the seats needed per plane is $(n - 1)\rho/f$. In the subtour and tour networks, the number of passengers on board the plane varies as it visits various cities on the route. In general, when the plane leaves the i th city, there are $i(n - 1)\rho/f$ passengers on board. Thus, for the subtour network, the maximum number of passengers (and the capacity of the plane) is equal to $k(n - k)\rho/f$ seats. For a tour, the maximum number of seats needed is when the plane departs from the $\frac{n}{2}$ th city, which is equal to $n^2\rho/4f$ seats.

Summarizing, the size of each plane is

$$\text{SIZE}(k) = \begin{cases} \frac{\rho}{f}, & k = 0, \\ \frac{k(n-k)\rho}{f}, & 0 < k < n, \\ \frac{n^2\rho}{4f}, & k = n. \end{cases}$$

The total seat-miles flown by aircraft is the product of the total distance traveled per dispatch, the frequency, f , and the capacity. For direct routing, the distance traveled per dispatch is equal to the sum of the chord lengths connecting any two cities and the distance involved in the shifts from city i to city $i + 1$. The first term is equal to $2Rn \cot(\pi/2n)$, and the second is equal to $2\pi R$. Thus the total seat-miles flown per day is $\rho(2Rn \cot(\pi/2n) + 2\pi R)$. For the subtour, the total distance traveled per dispatch is equal to the product of the distance traveled by a single plane and the number of planes. The latter is equal to the number of subtours, m . If $m > 1$, the planes travel to the hub: the distance traveled by each plane is equal to $(2\pi R/n)2(k - 1) + 2R$. Hence the seat-miles flown per day per subtour is,

$$\left(\frac{2\pi R}{n} 2(k - 1) + 2R\right) m \rho k(n - k).$$

If $k = n$, the distance traveled per dispatch is equal to $(2\pi R/n)2(n - 1)$; the seat-miles flown per day is equal to

$$\frac{2\pi R}{n} 2(n - 1) \frac{n^2\rho}{4}.$$

Summarizing, total seat-miles are

$$\text{SM}(k) = \begin{cases} 2R\rho\left(n \cot\left(\frac{\pi}{2n}\right) + \pi\right) & \text{if } k = 0, \\ \left(\frac{4\pi R}{n} (k - 1) + 2R\right) m \rho k(n - k) & \text{if } 0 < k < n, \\ \pi R(n - 1)n\rho & \text{if } k = n. \end{cases}$$

A.5. Congestion at the Hub

For a flight leaving the hub, the planned delay time is C_α ; for a flight arriving at the hub the planned delay time is D_α . For k such that $24 > k > 0$, the planned delay time for a passenger traveling from city (i, p) to city (i', p') is

$$dt((i, p), (i', p')) = \begin{cases} (i' - i)A_\alpha & \text{if } p = p' = 1, \dots, m, \\ (i - 1)A_\alpha + D_\alpha & \text{if } 0 < k < n, \\ + C_\alpha + (i' - 1)A_\alpha & \text{if } p, p' = 1, \dots, m, p \neq p'. \end{cases}$$

Summing up over all passengers,

$$\begin{aligned}
D &= \sum_{(i,p)} \sum_{(i',p')} st((i,p), (i',p')) \\
&= 2\rho m \sum_{i=1}^{k-1} \sum_{i'=i+1}^k (i' - i) A_\alpha \\
&\quad + \rho m(m-1) \sum_{i=1}^k \sum_{i'=1}^k (i - 1) A_\alpha + D_\alpha \\
&\quad + C_\alpha + (i' - 1) A_\alpha.
\end{aligned}$$

Therefore,

$$D(k) = \begin{cases} \rho n(n-1) A_\alpha & \text{if } k = 0, \\ \rho A_\alpha \left\{ \frac{n(k^2 - 1)}{3} + n(n-k)(k-1) \right\} \\ \quad + \rho(C_\alpha + D_\alpha)n(n-k) & \text{if } 0 < k \leq n. \end{cases}$$

By the definition of C_α ,

$$\int_0^{C_\alpha} \lambda e^{-\lambda x} (1 - e^{-(\lambda/\beta)(C_\alpha - x)}) dx = 1 - \alpha.$$

Therefore,

$$C_\alpha = -\frac{\beta}{\lambda} \ln[(e^{-\lambda C_\alpha} + (\beta - 1))/\beta].$$

Note that if

$$z = -\beta \ln[(e^{-z} + (1 - \alpha)(\beta - 1))/\beta],$$

$C_\alpha = \frac{1}{\lambda} z$. Parameter z can be found by iterative substitution in a manner similar to that used to find x and y . Therefore planned delay time at a spoke city for a plane arriving from the hub is inversely proportional to λ when β is fixed.

If the Prob[max. of m i.i.d. random variables $\leq D_\alpha$] = α , then {Prob[Each random variable $\leq D_\alpha$]} m = α , and Prob[Each random variable $\leq D_\alpha$] = $\alpha^{(1/m)}$. Therefore,

$$D_\alpha = -\frac{\beta}{\lambda} \ln[(e^{-\lambda C_\alpha} + (1 - \alpha^{(1/m)})(\beta - 1))/\beta].$$

Note that if

$$w = -\beta \ln[(e^{-w} + (1 - \alpha^{(1/m)})(\beta - 1))/\beta],$$

$$D_\alpha = \frac{1}{\lambda} w.$$

The distance traveled by the aircraft is the same as analyzed in Section 2.4. The capacity required rises compared with that found in Section 2.5 because planned delay time increases. The travel time for one full trip is

$$T(k) = \begin{cases} \frac{4nR}{v} \cot\left(\frac{\pi}{2(n+1)}\right) + \frac{2\pi R}{v} + n(n-1)(g + A_\alpha) & \text{if } k = 0, \\ \frac{4\pi R}{nv} (k-1) + \frac{2R}{v} + 2kg \\ \quad + 2(k-1)A_\alpha + D_\alpha + C_\alpha & \text{if } 1 \leq k < n, \\ \frac{4\pi R}{nv} (k-1) + (2n-1)g \\ \quad + 2(n-1)A_\alpha + D_\alpha + C_\alpha & \text{if } k = n. \end{cases}$$

Therefore, the total aircraft seats needed per day is $TS(k) = n/k \text{ SIZE}(k) \lceil fT(k) \rceil$.

ACKNOWLEDGMENT

Many thanks for helpful comments of participants of seminars at 1994 MIT Summer Camp, TIMS/ORSA-Phoenix, University of Edinburgh, and WHU (Koblenz, Germany). Special thanks to Don Rosenfield and Joyce Mehning for their suggestions. This paper benefited from many helpful comments and suggestions from the associate editor and two referees.

REFERENCES

- BAILEY, E. E., D. R. GRAHAM, AND D. P. KAPLAN. 1985. *De-regulating the Airlines*. MIT Press, Cambridge, MA.
- BEN-AKIVA, M. AND S. LERMAN. 1985. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. MIT Press, Cambridge, MA.
- BRUECKNER, A. K. AND P. T. SPILLER. Competition and Mergers in Airline Networks. *Intl. J. Industrial Org.* **9**, 323–342.
- CAVES, D., L. R. CHRISTENSEN, AND M. W. TRETHEWAY. 1984. Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ. *Rand J. Economics* **15**, 471–489.
- FLETCHER, R. 1987. *Practical Methods of Optimization*. John Wiley, New York.
- JENG, C.-Y. 1988. An Idealized Model for Understanding Impacts of Key Network Parameters on Airline Routing. *Transp. Res. Record* **1158**, 5–13.
- JOLLY, L. B. W. 1961. *Summation of Series*. Dover, New York.
- KANAFANI, A. AND A. A. GHOBRIAL. 1985. Airline Hubbing—Some Implications For Airport Economics. *Transp. Res.* **19A**, 15–27.
- KUBY, M. AND R. G. GRAY. 1993. The Hub Network Design Problem With Stopovers and Feeders: The Case of Federal Express. *Transp. Res.* **27A**, 1–12.
- LEDERER, P. J. 1993. A Competitive Network Design Problem with Pricing. *Transp. Sci.* **27**, 25–38.
- MARSTEN, R. E. AND M. R. MULLER. 1980. A Mixed-Integer Programming Approach to Air Cargo Fleet Planning. *Mgmt. Sci.* **26**, 1096–1107.
- MC SHAN, S. AND R. WINDLE. 1989. The Implications of Hub-and-Spoke Routing for Airline Costs and Competitiveness. *Logist. Transp. Review* **25**, 209–230.
- MORRISON, S. A. AND C. WINSTON. 1985. An Econometric Analysis of the Demand for Intercity Passenger Transportation. *Res. Transp. Econ.* **2**, 213–237.
- OUM, T.-H. AND M. W. TRETHEWAY. 1990. Airline Hub and Spoke Systems. *J. Transp. Res. Forum* **30**, 380–393.
- OUM, T.-H., A. ZHANG, AND Y. ZHANG. 1993. Airline Network Rivalry. Faculty of Commerce and Business Administration Working Paper, University of British Columbia.
- TREITSCH, D. 1993. Scheduling Flights at Hub Airports. *Transp. Res.* **27B**, 133–150.