A Method for Determining the Optimal Division of Express and Local Rail Transit Service

ALAN BLACK, Chicago Area Transportation Study, Chicago, Illinois

•MUCH RESEARCH has been devoted to the optimization of highway systems, and several papers have been published on such topics as the optimal spacing of expressways. However, relatively little study has been given to the optimization of mass transit systems. This paper describes an attempt to optimize one facet of a mass transit system. It is illustrative of the type of research that might profitably be devoted to studies of transit networks.

Certainly mass transit is deserving of research. In virtually all metropolitan areas, transit systems—including railroads, rapid transit, and busses—are experiencing a critical pinch between increasing costs and declining demand. Yet highway congestion has led many city planners and other observers of the urban scene to prescribe increased reliance on mass transit as the cure to traffic ills. Such programs would require substantial public investment, because the financial difficulties of transit systems have made them unattractive to private investors. It is a matter of public concern, therefore, to make every effort to provide the most efficient and economical transit systems.

The most important rail transit routes are radials carrying passengers to and from the central business district (CBD). The concept advanced here is that there should be a breakpoint on radial routes, with local trains serving the area between the breakpoint and the CBD, and express trains serving the area beyond the breakpoint. Express trains would carry CBD-bound riders non-stop from the breakpoint to the CBD. Express and local service could correspond to suburban railroad and rapid transit, respectively.

The question considered in this paper is the optimal location of the breakpoint between the two types of service. The objective is to find the breakpoint for which the total costs to the community will be the least.

This paper describes the general method for computing costs and determining the least cost breakpoint and demonstrates the application of the method to an actual situation in the Chicago region. It analyzes the behavior of the breakpoint and the factors that influence it. It also contains an evaluation of the method, with a discussion of its strengths and weaknesses.

METHOD

The problem is set up in an idealized form in which certain simplifying assumptions are made. The route to be considered is a radial route extending a fixed distance from the CBD and serving a given number of riders in a sector of the metropolitan area. Each of the stations along the route could serve either express or local trains.

The locations of the stations and the number of riders boarding or debarking at each station are taken as givens. Passenger volumes may vary from station to station, but the volume using any single station is assumed to be constant. It is assumed that all riders must be served by rail, and that they do not have the choice of an alternate mode of travel. Only those trips originating or terminating at the CBD are considered; these generally form the bulk of trips on a radial rail line.

For identification, the stations are numbered, starting at the CBD terminal, as 0, 1, 2, 3, etc., up to z, the station farthest from the CBD. The stations need not be uniformly spaced.

Two types of transit service are to be provided over this route: (a) local trains will run out from the CBD as far as a certain station, designated as station m-1, making all stops; and (b) express trains will run non-stop from the CBD to station m, and then will make all stops from station m to station z.

The problem is to find that breakpoint between local and express service for which all the riders in the corridor will be served at the least total cost, under a given set of conditions. The unknown is station m, the nearest stop for express trains. Each station is considered to be a potential breakpoint, and the total costs for each value of m are calculated.

Assumptions

The following are explicit statements of certain assumptions made in the problem:

1. All person-trips originate and terminate at the transit stations.
2. All person-trips have one end at the CBD terminal, station 0.
3. The number of train cars operated is proportional to the number of riders accumulated at the maximum load point on the line. This means the average car-loading at the maximum point is constant. (For local trains, the maximum point is between stations 0 and 1; for express trains, between stations 0 and m.)
4. The number of riders boarding or debarking at each station is constant.
5. Operating costs per car-mile are constant.
6. Delays at stops consume a constant amount of time per stop.
7. All trains run from terminal to terminal, with no equipment added or subtracted in midroute.
8. All trains run at maximum speed except when decelerating or accelerating for stops.
9. All express trains stop at all express stations, and all local trains stop at all local stations.

Objective

The objective is to minimize total costs, which consist of daily travel costs and capital costs for both express and local service.

Daily travel costs include three components: (a) operating costs for the railroad company or transit authority; (b) time required for passengers to travel at the maximum speed of the trains; and (c) time losses that passengers incur from delays at stops. Capital costs include the cost of constructing the lines plus the cost of equipment (rolling stock) required to serve the passenger volumes. To determine total costs, it is necessary to put daily travel costs and capital costs on a comparable basis, which can be done by converting both to annual costs.

Maximum speed travel time is important only to the extent that the two types of trains have different maximum speeds. If the same equipment is used on local and express service, or if the two types of trains have the same maximum speed, then the total maximum speed travel time for all riders becomes constant, and can be ignored.

Certain other costs have been omitted. One of these is the cost of waiting time for passengers at stations. It is assumed that the average waiting time for each passenger is constant, so the total waiting time for all passengers is also constant, and can be ignored. The cost of passengers traveling to and from transit stations is also omitted, because of the assumption that all trips originate and terminate at transit stations. Because it is assumed that all trips have one end at the CBD, no transfers are involved and the cost of transfer time is not considered. These costs are discussed later.

The cost of fares to passengers has not been included. It is the actual cost of providing the service which is of concern, and not what the passengers pay for the service. Fares are not a good indicator of actual operating costs, because they are often not sufficient to cover costs. Although fares represent a real cost to passengers, basically they are intended to reimburse the transit operator for its operating expenses. To include fares as well as the operating costs of the transit operator would be double counting.

Daily Travel Costs

Computation of the cost items included in daily travel costs may best be expressed in mathematical terms. The following factors are used:

p_i = number of passengers boarding or debarking at station i per day;

d_i = distance of station i from 0 (in miles);

k_1 = monetary value of one hour;

k_2 = express train car-mile operating cost,

k_3 = local train car-mile operating cost;

k_4 = average daily express train car-loading at maximum load point;

k_5 = average daily local train car-loading at maximum load point;

k_6 = maximum speed of express trains (in miles per hour);

k_7 = maximum speed of local trains (in miles per hour);

k_8 = time delay caused by one express train stop; and

k_9 = time delay caused by one local train stop.

Most of these items are self-explanatory. A time delay consists of the time lost by a train in decelerating for a stop, standing at the station to discharge and load passengers, and accelerating to maximum speed again.

All of the k's are assumed to be constant. The values of p_i and d_i are constant for any given station (any value of i), but vary from station to station. In the expressions that follow, these two factors appear only in summations, and the values of these summations depend on the value of m. The only independent variable is m.

The sum of the daily travel costs is made up of six cost items: operating costs, maximum speed time costs, and delay time costs, each for express and local trains. Computation of each of the six items is

$$\text{Local operating costs} = \frac{d_{m-1}\, k_3 \sum_{1}^{m-1} p_i}{k_5}$$

$$\text{Express operating costs} = \frac{d_z\, k_2 \sum_{m}^{z} p_i}{k_4}$$

The summations represent the number of passengers carried at the maximum load point on the line. Because z is the station farthest from the CBD, d_z is the distance of an express train run. Because m - 1 is the farthest station with local service, d_{m-1} is the distance of a local train run.

$$\text{Local maximum speed costs} = \frac{k_1 \sum_{1}^{m-1} p_i\, d_i}{k_7}$$

$$\text{Express maximum speed costs} = \frac{k_1 \sum_{m}^{z} p_i\, d_i}{k_6}$$

The summations represent the number of passenger-miles. Dividing by the speed converts this to the total passenger time.

$$\text{Local delay costs} = k_1 k_9 \sum_{1}^{m-1} i\, p_i$$

$$\text{Express delay costs} = k_1 k_8 \sum_{m}^{z} p_i \; (i - m + 1)$$

For local trains, i, the number of the station, also indicates the number of stops passengers boarding at i will sustain. Each passenger boarding an express train at a station, i, will sustain i - m + 1 stops. The summations represent the total number of delays that all passengers will experience.

Daily travel costs equal the sum of these six items. The use of summations in these expressions precludes the possibility of differentiating them. However, in any actual problem, the number of potential values of m would be limited, and so it would be neither especially difficult nor time-consuming to compute costs for all possible values of m, and to determine the minimum point by inspection.

Capital Costs

One component of capital costs is the amount of equipment—the number of train-cars—required to serve the passenger volume. The amount of equipment needed is that amount required to carry the passenger volume in the peak direction during the peak period of the day. During the rest of the day, the transit operator can get by with less equipment.

The basic relationship which determines the minimum number of passenger cars necessary is

Cars required = round trip time × car trips required

The reasoning behind this is simple: after a car passes any point on a line, it must make a round trip before it can pass that point again (going in the same direction). During this period (the round trip time) there must be one car in use for every car required to pass that point (as determined by the passenger volume). For a transit line, it is the number of car trips required at the maximum load point which determines the number of cars required.

Round trip time includes maximum speed time, plus delay time for all stops, plus layover time at the two terminals.

To determine the car trips required during the peak hour, it is necessary to determine the proportion of the daily, two-way volume that will occur in the peak direction during the peak hour. This volume must be converted to car trips. In doing this, a peak-hour car-loading should be used, rather than the daily average car-loading, since car-loading during the peak hour is much greater than the daily average.

This general rule holds: if the round trip time is x minutes, then the amount of equipment needed is determined by the greatest number of car trips required in the peak direction during any x consecutive minutes of the day. However, available data do not permit ready determination of the peak of x consecutive minutes, so it is necessary to approximate the distribution of travel over the day. In this problem, estimates of the proportion of daily travel during the peak hour and adjacent hours are used, and it is assumed that the demand is evenly distributed within each of these hours.

Three new factors must be introduced:

k_{10} = proportion of daily, two-way passenger volume occurring during peak hour, in peak direction;

k_{11} = ratio of peak-hour, peak-direction car-loading to daily average car-loading, at maximum load point; and

k_{12} = layover time for trains.

For simplicity, it is assumed that k_{11} and k_{12} are the same for both local and express trains.

The desired expressions for the peak hour and peak direction are

$$\text{Local car trips} = \frac{k_{10} \sum_{1}^{m-1} p_i}{k_{11}\, k_5}$$

$$\text{Express car trips} = \frac{k_{10} \sum_{m}^{z} p_i}{k_{11}\, k_4}$$

$$\text{Local round trip time} = 2 \left[\frac{d_{m-1}}{k_7} + k_9\, m + k_{12} \right]$$

$$\text{Express round trip time} = 2 \left[\frac{d_z}{k_6} + k_8\, (z - m + 2) + k_{12} \right]$$

The number of cars required equals the product of car trips per hour and round trip time, in hours.

This solution is fine, as long as round trip time does not exceed one hour. When it exceeds one hour, this method causes the number of cars required to exceed the number of car trips required during the peak hour. This would be all right if the passenger volume during the peak hour extended into the adjacent hour, but in fact it does not.

When the round trip time exceeds one hour, the number of cars required equals the number of car trips required during the peak hour plus the round trip time minus one hour times the number of car trips required during the next-to-peak hour. If the round trip time exceeds two hours, the same technique may be extended.

A transit operator requires some spare cars, so once the number of passenger cars required for the peak period of the day is determined, this figure is increased by 10 percent. If the cars are not self-powered and engines are also required, the number of engines needed can be based on a ratio of passenger cars per engine. Then the final figures are multiplied by unit costs to determine the total cost of equipment.

The other component of capital costs is construction cost. Express track must run the full length of the route, so its construction cost is constant. The number of stations is constant, regardless of location of the breakpoint, so this cost is constant. The cost of constructing local track is the most important variable construction cost. This can be crudely estimated on a per-mile basis, or fairly detailed cost estimates for each section of the route can be developed.

Figuring Total Costs

To add daily travel costs and capital costs together, both must be in the same terms, so both are converted to annual costs. Daily travel costs are figured for an average weekday. This can be converted to annual costs by multiplying by the ratio of annual passenger volume to an average weekday volume.

Capital costs can be converted to annual costs by multiplying by the capital recovery factor. The capital recovery factor converts an investment (first cost) into a uniform annual series of payments which reflect the time value of money. An interest rate, representing the minimum attractive rate of return, and a facility life span must be assumed. The resulting annual cost is the amount that would have to be paid every year if the first cost were borrowed at the specified interest rate for the specified facility life.

DEMONSTRATION OF THE METHOD

To illustrate the method, it was applied to an actual situation in the northwestern sector of the Chicago metropolitan region. This sector is served by a suburban railroad line, the Wisconsin Division of the Chicago and North Western Railway, and a rapid-transit line, the Logan Square elevated branch of the Chicago Transit Authority. The railroad line extends beyond the boundary of the metropolitan region into Wisconsin. The rapid transit line extends to Logan Square, about five miles from the Loop.

This situation is of some interest because the CTA proposes to extend its line in the median strip of the Northwest Expressway, which is adjacent to the railroad tracks, to a terminal at O'Hare Field. Space has been left in the expressway median strip for this.

Construction of the extension has not begun due to lack of funds. The CTA has been actively, but unsucessfully, seeking an appropriation from the State Legislature. The Chicago and North Western Railway is opposed to the extension, on the grounds that it would divert a large number of the railroad's passengers. Consequently, the question of whether the CTA line should be extended is a significant issue.

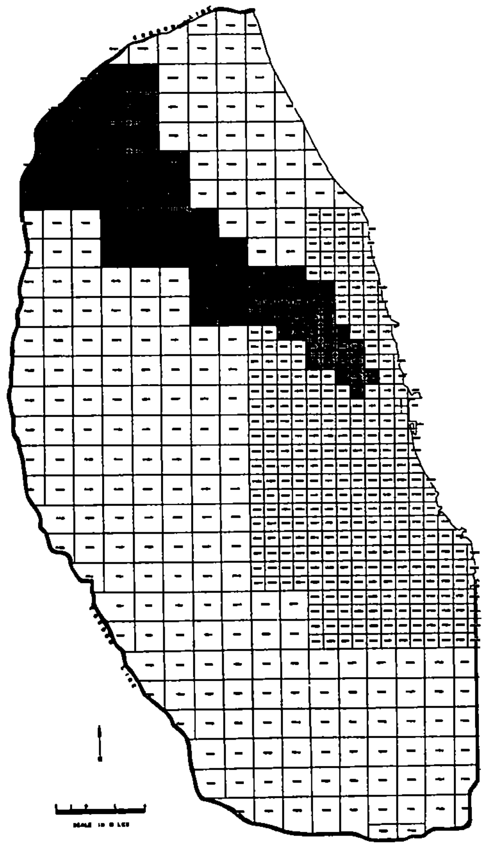

Figure 1. Sector served by rail transit route.

TABLE 1

VALUES OF d_i AND p_i USED IN DEMONSTRATION

Station (i)	Miles from CBD (d)	No. of Riders (p)
1	1	7,390
2	2	7,056
3	3	5,250
4	4	9,054
5	4.5	19,368
6	5	4,792
7	5.5	9,674
8	6	4,026
9	7	10,818
10	8	4,934
11	9	4,060
12	10	7,252
13	11	1,424
14	12	5,890
15	13	10,124
16	15	1,872
17	17	4,226
18	18	1,302
19	20	7,460
20	22	5,158
21	24	8,580
22	26	2,742
23	28	1,554
24	32	12,670
Total		156,676

The actual situation can be somewhat modified so that it fits the format of the idealized problem described, and a least cost breakpoint can be determined. The railroad line may be regarded as express service, and the CTA line as local service. It is necessary to assume that the railroad and rapid transit lines would be adjacent (which is not entirely true) and each station could be used by either one. It is assumed that the sector contains a constant number of transit riders who are indifferent as to which type of service they use. Only trips originating or terminating in the central area can be considered.

Accurate data were not readily available to use for inputs in the problem, so approximations were made. Because of this and the simplifying assumptions, there is no pretense that this exercise answers the question of whether the CTA line should be extended. This must be regarded purely as a demonstration of how the method works.

Values Used

The route is 32 miles long, ending at the boundary of the study area delineated by the Chicago Area Transportation Study (CATS). Besides the CBD terminal, the route has 24 stations. The distance of each station from the CBD and the number of riders boarding or debarking at each station are given in Table 1.

The sector used as the trade area for the route consists of 75 CATS analysis zones, and is shown in Figure 1. (External trips using the C & NW line were also included.) This consists primarily of 72 zones from which, according to a 1980 transit assignment made by CATS, (Assignment 88; CATS unpublished) transit riders would use either the C & NW line or the proposed CTA line to reach the CBD. Three other zones close to the CBD were added to create a more realistic situation, inasmuch as the assignment showed no passengers using some of the close-in CTA stations.

The locations of the stations were taken from the 1980 network used in the assignment. Three stations were omitted in the portion of the CTA route which would diverge from the railroad line and swing west to O'Hare Field. In two cases, stations were combined.

The volumes of riders using each station were based on 1980 estimates of central area transit trips generated by each zone in the sector. The trips from each zone were assigned to the station they would use in taking the shortest time path to a particular zone (Zone 001) in the heart of the CBD, according to the 1980 transit network coding. Certain modifications were made. Zones using the stations that were omitted were reassigned to other stations according to shortest time path calculations. The three zones added to the sector were assigned to stations which appeared reasonable.

Essentially, the processes described so far consisted of consolidating the two transit routes (railroad and rapid transit) of the 1980 assignment into a single route and allocating all the central area trips in the sector to stations on that route.

The value of time used was $0.85 per hour. This is the value used in CATS economic analyses of transit plans.

Operating costs per car-mile were set at $0.65 for rapid transit and $1.00 for railroad. A breakdown of CTA expenses made for 1954 showed that operating expenses, including injuries and damages but not including debt service or depreciation, for the rapid transit system were $0.566 per car-mile. From 1954 to 1960, operating expenses per car-mile for the entire CTA system rose from $0.630 to $0.736. No breakdown by type of equipment was made in 1960, but $0.65 seemed reasonable in view of the 1954 relationship.

In 1954, the Chicago and North Western Railway had directly assigned operating expenses of $0.770 per passenger car-mile for its entire operations. Expenses have risen since then, and it is likely that these particular commuter cars cost more to operate than the average for the entire railroad, so $1.00 seemed a reasonable estimate.

Average daily car-loading at the maximum load point was estimated at 40 passengers per car for rapid transit and 80 for railroad. Chicago cordon count data for 1961 showed the rapid transit system had an average car-loading of 41.8 inbound and 40.3 outbound for the period from 7 AM to 7 PM. No such data were available for the railroad, but

80 seemed reasonable. Although the C & NW's cars have 160 seats, they run many trains at off-peak hours.

The maximum speed of rapid transit trains was set at 45 mph, a figure verified by clocking. The maximum speed of railroad trains was set at 55 mph, based on a study of schedules for long commuter runs.

Delay time for rapid transit was set at 42.5 sec, based on assumed rates of acceleration and deceleration of 2 mph per sec and a standing time of 20 sec. For railroad, delay time was set at 90 sec, based on rates of acceleration and deceleration of 1 mph per sec and a standing time of 35 sec.

The cost of a rapid transit car was estimated at $80,000, which is representative of prices paid in recent years. The cost of a railroad car was estimated at $200,000. This is intended to include the cost of locomotives, and is based on $150,000 for each passenger car and $200,000 per locomotive, with one locomotive required for every four passenger cars. The C & NW, in 1956, paid $145,700 for each passenger car used on its line.

It was difficult to ascertain the construction cost of extending the CTA line, because a number of varying estimates have been made. An average cost of $2,500,000 per mile was adopted as a reasonable approximation. This is not supposed to include the cost of constructing stations, because in the problem it is assumed that the stations will be in existence, regardless of whether they are used by railroad or rapid transit.

In calculating equipment costs, it was estimated that the proportion of the daily, two-way volume occurring in the peak hour and peak direction was 14 percent. The proportions occurring in the two highest adjacent hours were estimated at 11.2 and 3.2 percent. These estimates were based on a combination of data from the 1961 cordon count and a subsample of CATS home interviews.

The ratio of peak-hour, peak-direction car-loading to daily average car-loading was taken as 175 percent, meaning a car-loading of 70 for rapid transit and 140 for railroad. According to the 1961 cordon count, average car-loading for the rapid transit system was 68.9 inbound from 8 to 9 AM, and 76.0 outbound from 5 to 6 PM. Ratios to the 12-hr average were 165 percent for inbound and 188 percent for outbound.

Layover time was estimated at 3 min at each terminal.

In converting daily costs to annual costs, it was assumed that there are 300 average weekday equivalents in a year. In 1960, the CTA rapid transit system had 112,924,491 revenue passengers, which was 297.0 times the 380,182 counted in a spot check on a November weekday.

In converting capital costs to annual costs, the minimum attractive rate of return was put at 10 percent, and facility life was assumed to be 25 years. This produces a capital recovery factor of 0.11017. It was assumed that all equipment required would have to be purchased new. It was assumed that the only construction required would be to extend the rapid transit line outward from its present terminal.

Results

Total costs for each possible value of m were computed, and the minimum point was found to occur when m = 13, the station 11 mi from the CBD. This would mean that the rapid transit line would be extended to the station 10 mi out. Total annual costs for this value of m are given in Table 2. Figure 2 shows the curve of annual costs for all values of m, and Figure 3 shows daily travel costs.

When m = 13, the rapid transit line would carry 93,674 riders, and the railroad line, 63,002 riders. There would be 486,205 passenger-miles on rapid transit, and 1,330,154 passenger-miles

TABLE 2

ANNUAL COSTS FOR m = 13

Cost	Amount (dollars)
Travel:	
Operating	12,127,000
Maximum speed time	8,926,000
Delay time	4,589,000
Total	25,642,000
Equipment	4,098,000
Construction	1,653,000
Total	31,393,000

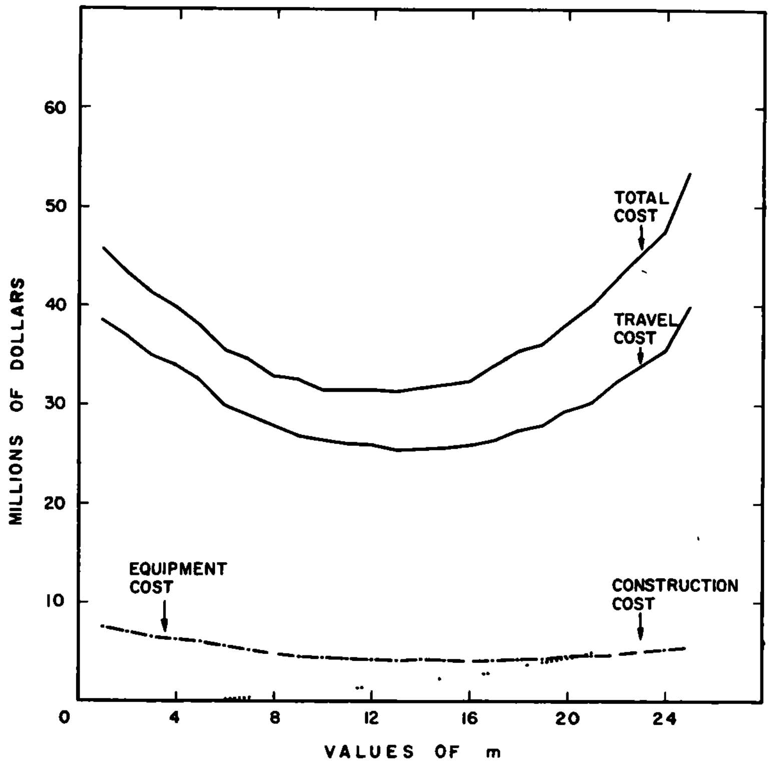

Figure 2. Annual costs for all values of m.

on railroad. Mean trip length would be 5.2 mi for rapid transit riders, and 21.1 mi for railroad riders.

For this breakpoint, daily operating cost per passenger would be \$0.1625 for rapid transit and \$0.4000 for railroad. Operating cost per passenger-mile came out at \$0.0313 for rapid transit and \$0.0189 for railroad.

This value of m would require 170 rapid transit cars and 118 railroad passenger cars. Six additional miles of rapid transit track would have to be built.

These calculations were intended primarily to illustrate how the method would be applied to an actual situation. The results appear to be reasonable, but because of uncertainty about some of the approximations and assumptions, they should not be taken too literally. Changes in some of the inputs might well produce a different minimum point.

ANALYSIS

Flatness of Total Cost Curve

The curve of total costs, plotted against values of m, is fairly flat at the bottom (Fig. 2). Total annual costs for m = 12 are only \$55,000 greater than the minimum, and for m = 11, only \$65,000 greater. Because total costs in this area exceed \$31,000,000, these differences are very slight.

This suggests that the precise location of the breakpoint does not matter greatly, and that it could be moved a few stations either way without seriously increasing costs. If total costs were allowed to exceed the minimum by \$1,000,000, m could have any value from 10 to 16 (corresponding to distances of 8 to 15 mi from the CBD).

However, beyond these limits, total costs begin to rise sharply. Extension of the rapid transit line from its present terminal to the optimal breakpoint would produce annual savings of \$6,851,000.

The flatness of the bottom of the curve lends some importance to securing good estimates of the factors used in solving a problem. Moderate variations in the inputs could well shift the location of the minimum point. However, even if the inputs are changed, the curve will still be flat in approximately the same area.

It would be more reassuring, perhaps, if there were a single breakpoint that was far more advantageous than any other. This does not appear to be in the nature of things. On the other hand, the existence of a range of breakpoints that are approximately equal in cost gives latitude of choice which may be desirable because of the context in which a decision must be made. Other factors that cannot be readily quantified (political, social, economic, or aesthetic) can be taken into consideration in selecting the precise location for the breakpoint.

Influence of Factors on Breakpoint

The least cost breakpoint is a result of a series of pulls in opposite directions, with each pull having a certain weight. Changing the weights of the pulls will alter the location of the breakpoint.

Some of the component cost items take the form of U-shaped curves and may have minimum points that differ from the over-all minimum for total costs. The minimum for travel costs occurs when m = 13, and for equipment costs, when m = 16. Construction costs rise steadily as the value of m increases (Fig. 2).

The minimum for operating costs occurs when m = 13, and delay time costs are minimized when m = 19. Maximum speed time costs increase steadily as the breakpoint moves out, because express trains have a faster speed (Fig. 3).

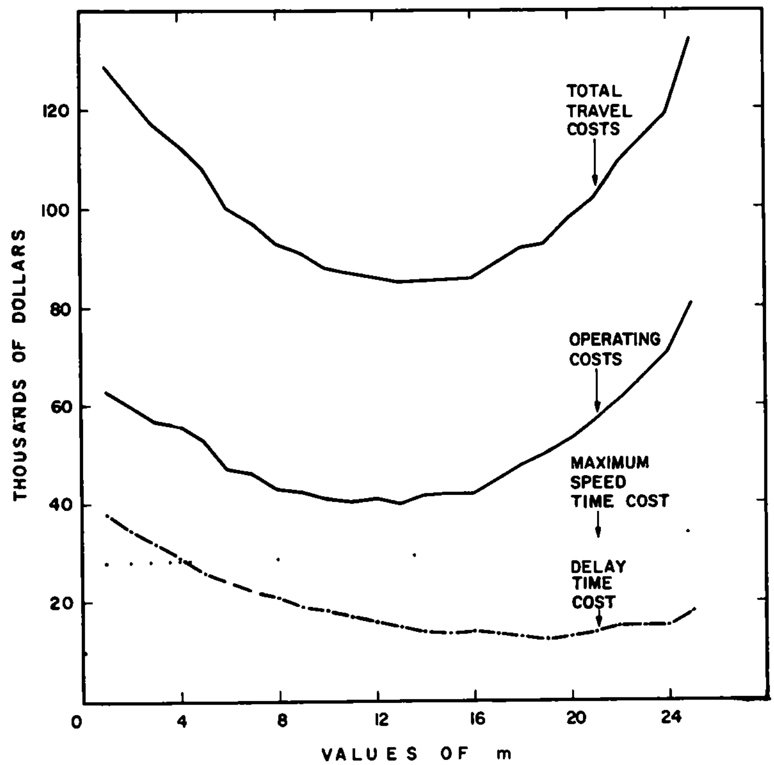

Figure 3. Daily travel costs for all values of m.

What happens to the breakpoint as the magnitude of the various factors changes? In general, any change that favors express service, or makes it comparatively more economical, will move the breakpoint in, whereas any change that favors local service will move the breakpoint out. For example, if express car-loading increases, the breakpoint will move in, but if express operating costs increase, the breakpoint will move out. If local maximum speed increases, the breakpoint will move out, but if local operating costs increase, the breakpoint will move in.

The distribution of riders over the stations on the line is an important factor. As the proportion of riders on the outer parts of the line increases, the breakpoint will move out.

A change in the value of time might move the breakpoint either way. This would depend on the relationship of the minimum point of total time costs (maximum speed time plus delay time) to the over-all minimum point. An increase in the value of time will pull the over-all minimum closer to the minimum point of time costs.

Superiority of Two Types of Service

The provision of both local and express service, instead of only one type of service, produces savings in operating costs, delay time costs, and equipment costs.

Operating costs are proportional to the number of car-miles. Under the assumptions of the problem, cars must travel from terminal to terminal, even if they only carry passengers for a portion of their run and are empty the rest of the way. With one type of service, every train must run the full length of the route. With two types of service, trains in local service will not run the full length of the route, and some of the unneeded car-miles will be eliminated.

Delay time costs are proportional to the number of stops that passengers on trains must sustain. With one type of service, all passengers must sustain delays for all stations between their station and the CBD. With two types of service, express passengers do not sustain delays between the breakpoint and the CBD.

Equipment costs increase directly with round trip time. With one type of service, all trains must make a round trip over the full route length. With two types of service, local trains cover a shorter route and their round trip time is less. Express trains also have shorter round trip times because some of their stops are eliminated.

Cause of Minimum Point

Although the total number of passengers and passenger-miles remains constant, regardless of the value of m, there will be minimum points for operating costs, delay time costs, and equipment costs. This will be true even if the same type of equipment is used on both local and express service (i.e., car-mile operating cost, car-loading, maximum speed, and delay time each have a single value).

As the breakpoint moves out from the CBD, local costs increase slowly at first, but gradually the rate of increase becomes faster and faster. When the breakpoint nears the outer extremity of the route, increments in local costs are very large. The same thing happens to express costs as the breakpoint moves in from the outer terminal. The combination of these two functions produces a U-shaped curve with a minimum point somewhere in the middle of the route.

There is no minimum point for maximum speed time costs, in the sense that there is no point with zero slope. If maximum speed is the same for both types of service, the sum of maximum speed time costs will be constant. If there is a difference, then these costs will be least when there is only one type of service utilizing only the faster equipment.

Likewise, there is no minimum point for construction costs. They will be least if there is no construction. If there is no existing transit route, they will be least if only one type of service is provided. However, the influence of the three factors that have minima is great enough to give the total cost curve a U-shape.

EVALUATION

Strengths

It has been shown that the combination of express and local service can be more economical than a single type of service in accommodating a constant volume and pattern of daily travel movements. Total costs will vary according to the location of the breakpoint. The method described in this paper offers a rational basis for determining the optimal breakpoint. It should provide a sounder foundation for decision than intuitive judgment.

The method is essentially a shortcut procedure for calculating costs. A series of full transit assignments, each using a different breakpoint, could provide more detailed results, but this would require considerable time and expense to carry out. It is unlikely that the results supplied by assignments would be much different.

The method may seem unduly complex to some. A simpler formulation could be achieved, but only at the cost of making the method more academic and less applicable to real situations. The method is intended to utilize actual data.

A mathematically neater solution could be obtained by replacing the summations in the mathematical expressions for computing costs by integrations. It would be possible to develop a continuous curve which described the number of riders at each station (p_i) as a function of distance from the CBD (d_i). This would permit determination of the minimum point by calculus, eliminating the bother of computing costs for all values of m. However, this would necessitate assuming some regular shape to the density of riders according to distance from the CBD—a simplification that would only roughly correspond to reality.

A key component of the method is the separation of travel time into time required for trains to travel at maximum speed and time consumed by making stops. For each passenger, any stop other than those at which he gets on and off is superfluous, and the more of these that can be eliminated, the better. This separation is essential to breakpoint considerations.

The method of calculating equipment needs is reliable. As a check, the method was used to calculate how many rapid transit cars the entire CTA system needs to handle peak-hour loads. It was estimated that 1,012 cars are needed. The CTA has reported that it uses about 1,000 cars on an average weekday.

The exclusion of non-CBD trips is unfortunate, but trips to and from the CBD do make up the bulk of radial transit trips. Tables 3 and 4 derived from 1956 and 1980 CATS transit assignments, give evidence of this.

Weaknesses

As presented in this paper, the method has a number of weaknesses. The simplified, idealized picture on which it is based does not conform to reality quite well enough. It would be desirable to eliminate some of the assumptions and limitations. The principal weaknesses are discussed here.

TABLE 3

TRANSIT TRIPS IN NORTHWEST SECTOR[a]

Transit System	Total	Trips: Entering CBD No	Entering CBD %	Short of CBD No	Short of CBD %
C & NW RR	28,688	27,784	96 8	904	3 2
Logan Sq El.[b]	57,408	57,408	100 0	0	0

[a]According to Transit Assignment 86 (1956 trips to 1956 network)
[b]Between Logan Square and CBD boundary

TABLE 4

TRANSIT TRIPS IN NORTHWEST SECTOR[a]

Transit System	Total	Trips: Entering CBD No	Entering CBD %	Short of CBD No.	Short of CBD %
C & NW RR	46,664	45,496	97.5	1,168	2.5
NW Rapid Transit[b]	143,146	119,504	83.5	23,642	16.5

[a] According to Transit Assignment 88 (1980 trips to 1980 network).
[b] Between O'Hare Field and CBD boundary

Use of Empirical Data. —Some of the factors used as inputs will always be difficult to pin down. Accurate estimates of car-mile operating costs would require a thorough cost accounting analysis. The monetary value of time will always be a subject for debate. Estimates of passenger volumes, especially when forecast for some future date, are susceptible to the usual dangers of predicting how people will behave.

One problem in using data stems from the fact that railroad and rapid transit usage tend to behave differently. Train riding is more concentrated in the rush hours than is usage of rapid transit. Probably the variation in car-loading during the day also differs considerably. In the method, it is assumed that the two types of service are interchangeable, and riders at a station will use whichever type stops there. In calculating the hourly distribution of riding, railroad and rapid transit data were pooled and over-all figures were used.

Probably the ratio of annual passenger volume to an average weekday volume also differs for rapid transit and railroad. In the demonstation, the ratio for rapid transit was used for both.

Assumption of Constant Operating Cost. —Operating costs are pegged to car-miles, and it is assumed that the operating cost per car-mile is constant. This is probably not entirely true. Operating costs actually include both variable and fixed components. In solving such a problem, only the variable costs should be considered. However, it would be very difficult to separate variable and fixed costs, because some would shade into each other.

Operating costs undoubtedly vary in response to factors other than the number of car-miles operated. Train length is probably an important factor; the longer the train, the less the cost per car. Speed may also be significant—certain costs, such as power, would probably rise with increasing speed, but labor costs per car-mile would probably fall. However, working rules give some fixity to labor costs. A train crew may have to be paid whether it operates a train, does some other type of work, or is off duty during the midday.

The complexity of the factors involved makes it appear infeasible to use anything other than over-all operating cost per car-mile.

Use of Average Car-Loading as a Criterion for Car Trips. —It was assumed that the number of cars operated is proportional to the number of riders at the maximum load point on the line, which makes the average car-loading at this point constant. What is needed is some quantitative criterion for determining how many cars will be operated to serve given numbers of riders. This assumption was adopted as such a criterion.

Of course, average car-loading varies greatly according to hour and direction. It also varies considerably between different lines.

As for the criterion a transit operator uses to decide how many cars to schedule, reportedly the CTA schedules enough cars so that the number of standees will be 50 percent of the number of seats at the maximum load point. However, this standard is exceeded on some lines during the rush hour and is not attained during midday hours. The CTA operates an unvarying minimum schedule during midday, regardless of passenger volumes. It is doubtful that transit operators use any systematic quantitative method for determining how many cars to schedule.

Consideration of Waiting Time. —The cost of the time during which riders wait for trains at stations was omitted from the total costs. If some average waiting time were assumed, as is done in transit assignments, then the total waiting time for all riders would be constant.

The major determinant of waiting time is the frequency of trains, which is directly related to the frequency of cars and inversely related to train length. If a constant train length were assumed, then the total waiting time for all riders again would be constant.

Train length was not considered in the method. No satisfactory method for calculating variable train lengths, or train frequencies, was discovered. Again, some systematic quantitative criterion is needed.

Consideration of Time to Reach Stations. —Another cost ignored was the time required for passengers to reach the stations. It was assumed that all trips originated

TABLE 5

RAIL TRANSIT TRIPS IN NORTHWEST SECTOR, C & NW RAILROAD PLUS LOGAN SQUARE ELEVATED[a]

Trips	Number	Percent
To CBD	66,884	77.7
Short of CBD	904	1.0
Through CBD	18,308	21.3
Total	86,096	100.0

[a] According to Transit Assignment 86 (1956 trips to 1956 network).

TABLE 6

RAIL TRANSIT TRIPS IN NORTHWEST SECTOR, C & NW RAILROAD PLUS NORTHWEST RAPID TRANSIT[a]

Trips	Number	Percent
To CBD	131,942	69.5
Short of CBD	24,810	13.1
Through CBD	33,058	17.4
Total	189,810	100.0

[a] According to Transit Assignment 88 (1980 trips to 1980 network).

or terminated at the stations. In the idealized situation in which the two rail lines are adjacent and could use the same stations, the costs of reaching the stations would be constant. However, if the two lines were some distance apart, with stations at different places, then the total time required to reach stations might vary according to location of the breakpoint.

Of course, if the two lines are not adjacent, it becomes questionable whether they both have the same potential trade area.

Consideration of CBD Trips.—It is assumed that all trips have one end at the CBD. Actually, there are two other types of trips: (a) those that never reach the CBD, but have both ends on the transit route outside of the CBD, and (b) those that go through the CBD to some other sector, and only use the CBD as a transfer point.

It would be possible to include these two kinds of trips in the method, provided that the number of riders using each station on the route could be broken down into these three categories of trips (CBD, short of CBD, and through CBD). At present, transit assignments do not put out this information.

If this information were known, then the through-CBD trips would be lumped with the CBD trips and treated in the same way. However, the short-of-CBD trips would require different treatment, because some of these trips would have to transfer between express and local trains. The local line would be extended to m to permit transfer there. For each potential value of m, it would be necessary to know how many passengers would transfer there. Otherwise, it would be impossible to know how many riders would use local or express trains within the breakpoint.

Evidence from the assignments indicates that the number of short-of-CBD trips on the C & NW line is negligible, but that on the rapid transit line these might account for as much as 16.5 percent of all trips (Tables 3 and 4). Of course, these totals include trips occurring anywhere on the line; the number accumulated at any one point would probably be much less.

In one transit assignment (Assignment 97; CATS unpublished), a transfer point between the C & NW and CTA lines was established, but only 520 passengers transferred there.

Although it is not possible to break down the trips at each station into the three categories, it is possible to estimate the breakdown for all rail transit trips in the sector (Tables 5 and 6).

Though it would be desirable to include non-CBD trips, it is not possible to do so at present. In any case, CBD trips make up the vast majority of all trips.

Consideration of Transfer Time.—If short-of-CBD trips were included, then the time required for passengers to transfer between the two types of service should be incorporated in the total costs. The number of passengers transferring for each value of m would be known, and a constant average transfer time could be assumed.

Assumption of Constant Number of Riders at Each Station. —It is assumed that the number of riders boarding or debarking at each station is constant. No account is taken of the effects of fares, frequency of service, comfort, or over-all speed on riding habits. In reality, these factors affect people's choice between railroad and rapid transit, and between rail transit and other modes of travel.

It is also assumed that riders always use the same station, which may not be true. Near the breakpoint, riders might switch to a station farther from the CBD in order to use railroad instead of rapid transit.

It would be possible to develop quantitative expressions for the influence of fares, frequency of service, and over-all speed on passenger volumes. Choices between stations, and between rail transit and other modes of travel, might be approximated through full assignments.

The only point in introducing these complexities would be to increase the realism of the method. Unless it could be shown that the mathematical expressions and assignments were reliable descriptions of how people behave, there would be no advantage in substituting complicated approximations for simple ones.

The assumption that the number of riders using a transit station is constant probably holds for the majority of transit riders. Evidence indicates that only a minority have any real choice between travel alternatives (1). (CATS' findings for the Chicago region are similar.)

CONCLUSIONS

The author believes this general method is the right approach to the search for an optimal breakpoint. Some refinements are needed to make the idealized problem more applicable to actual situations. However, it is seldom possible to represent all the complexities and nuances of a real situation in mathematical abstractions; nor is it really necessary. Beyond a certain point, further gains in precision are superfluous, especially when the results may be used rather grossly, anyhow.

Two types of transit service can be more economical than one, despite the additional capital investment that may be required. Savings in delay time costs, operating costs, and equipment costs have been demonstrated. The location of the breakpoint affects costs. There is a range of possible breakpoints within which costs are at or very near the minimum.

The two most important improvements needed are to include non-CBD trips and to find a better criterion for the number of car trips operated. A better method of estimating operating costs and inclusion of waiting time costs and transfer time costs would also be useful refinements. Other weaknesses are relatively minor.

Although the problem considered in this paper is a specialized one, the method should have considerable applicability. There are several radial routes in Chicago where the question of a breakpoint is pertinent. Undoubtedly there are similar situations in other cities that have rail transit systems. The method can be applied to railroads alone, inasmuch as their routes often have a large number of tracks. It is a common practice for railroads to run both local and express trains over the same routes.

Just as two types of service are more economical than one, it may be that three types would be more economical than two (with three sets of tracks on the same route). Perhaps the optimal arrangement of transit service would be to have non-stop service between the CBD and every station. These possibilities have not yet been investigated.

These are examples of the types of further study of transit systems which might be rewarding. Other topics deserving research include the optimal spacing of transit stations (2), the optimal spacing of radial transit routes, and optimal scheduling of trains.

REFERENCES

1. Keefer, L.E., " 'Choice' Transit Trips." Pittsburgh Area Transportation Study Research Letter, 3: No. 1 (Jan.-Feb. 1961).
2. Schneider, M., "Travel Time Minimization, Transit." CATS Research News, 4: No. 2 (June 2, 1961).