A Capacity-Restraint Algorithm for Assigning Flow to a Transport Network

WALTER W. MOSHER, JR., Assistant Research Engineer, Institute of Transportation and Traffic Engineering, University of California, Los Angeles

The task of assigning units of movement to a complex transport network has, in the past, been accomplished by techniques that are limited by their requiring constant network link costs. In general, these existing transport flow assignment techniques ignore the fact that as the number of units of movement assigned to a network is increased, localized congestion will occur and total network cost will rise disproportionately. Thus, the effect of link capacity on transport flow is not considered, and as a result many links in the transport network become unrealistically overloaded and others receive very little or no assigned flow.

This paper presents a "capacity-restraint" algorithm which permits the evaluation of network performance based on arbitrarily selected network figures of merit. The values of these figures of merit depend on network loading, and are governed by individually determined link performance functions. Each link performance function may be any linear or nonlinear relation of link flow to cost (e.g., distance) for that flow. Once a performance function is established for each network link, the network can be loaded in an optimum manner either by minimizing the figure of merit for the entire system or by equalizing the path figures of merit over appropriate sets of paths. For road networks, the latter is the one most likely to occur.

• THE QUESTION of how a complex transport system should be designed must be considered in the context of what the ultimate goals are of the affected socio-economic group. Further, because these ultimate goals are not static but dynamic functions of the progress of society, it is necessary to have planning techniques that allow new technology to influence the design and ultimate realization of the system.

The problem reduces to that of supplying to the populace a system designed to achieve optimum living conditions. This implies maximum interaction and freedom of movement at a minimal cost, and therefore requires that desired living conditions be compromised if the minimum total cost for a particular system cannot be borne by the populace.

To design such a system it is necessary to determine what factors comprise optimum living conditions. A recent survey of 400 recognized leaders of the planning profession established that there are at least 6 broad planning criteria (8). These are resources, land use, economic, transportation, socio-humanistic, and catastrophe. Of course, the tabulation of these broad planning criteria does not imply that the relative importance of them and their associated subcriteria is known. The nature of these criteria is such that they interact with each other to a great extent, thereby making it extremely difficult to determine absolute relations between them and the desired, social-oriented, living conditions.

It is evident from the preceding that the first step in designing a transport system

Paper sponsored by Committee on Origin and Destination.

is to establish, empirically or theoretically, these relations between the various planning criteria. Inasmuch as the determination of the relative importance of the factors constituting "optimum" or "desired" living conditions is subject to individual interpretation, and because they are also subject to change as society advances, it is imperative that these relations provide for dynamic system goals.

After this has been accomplished, an analytical model representing the proposed system as well as the present system must be devised. This entails the determination of mathematical relations for and between the various planning criteria as functions of time and cost (in the general sense; e.g., safety, travel time, distance, and psychological stress). Boundary conditions in consonance with the proposed society goals must also be determined. Whenever possible, these boundary conditions should not be expressed as discrete values, but instead as regions within which system realization would be acceptable. For example, in the case of land allocation to families, rather than state a particular square footage per person (e.g., 15,000 sq ft), the constraint should be given as not less than a fixed amount or as between two fixed limits (e.g., 7,500 to 20,000 sq ft).

When this has been accomplished, analytical techniques may be used to determine the compatibility of the various subsystems comprising the planned system, and to indicate necessary revisions where incompatible situations exist. Further, as the planned system is effectuated, information concerning its operation at any given time may be used to check the correctness of the system model and to indicate changes that will improve the accuracy with which future system configurations can be forecast.

THE TRANSPORT PLANNING PROBLEM

Once a set of defining factors for a proposed system has been established, including its associated equations and parameters, it is necessary to determine the type and extent of facilities required to accommodate it. In the transportation sense, this task amounts to the design of roads, streets, freeways, and mass transit facilities such that the desired system interaction and freedom of movement is provided. For the support functions of the system (e.g., communication, power, sewage, and flood control), the design task is that of providing adequate capacity to handle the expected loads. For both transportation and support, it is desirable to build toward the ultimate required capacity in parallel with total system realization and in such a way that the partially realized system is operating in its optimum manner at each moment of time.

It is apparent that many alternate designs for accommodating the various transport fluxes are possible. Route selection, capacity, configuration, accessibility, etc., of each network link will affect the overall system realization, because the cost of constructing each link as well as the cost to society for using that link will vary as these factors are changed. It is, therefore, necessary to consider many alternate plans when attempting to find the best configuration for a given system, and to balance the cost of constructing each configuration with the benefits derived therefrom.

Because there is no assurance that any particular design will attain a solution in consonance with the defined system goals, each proposed design must be tested and, if necessary, modified until an allowable solution is achieved. These considerations indicate that an iterative procedure is required that will allow the evaluation of how well the system goals are satisfied after each iterative pass, and which will indicate changes that should or must be effectuated to achieve these goals better.

The heart of such a procedure will be a technique for the determination of patterns of movement within a proposed system over its proposed transport links. This entails the determination of the number of units of movement, for each transport flux, between every pair of origins and destinations, and then the assignment of these units of movement to appropriate transport links. Because the number of units of movement between zones is a function of the cost to accomplish these movements, which is, in turn, dependent on the total network loading, an iterative procedure is again suggested. In this instance, the first iteration cycle begins by determining the number of units that will move between each zone pair for an estimated set of link use costs. These costs might be estimated from the link cost function by arbitrarily setting all link flow values to zero. Next, the network is loaded, and the resulting link costs determined. From these new costs, the interzonal movements are redetermined, thus beginning the second iteration cycle. This procedure is repeated until the interzonal movements do not change significantly from one iteration cycle to the next.

From the preceding discussion, three distinct phases to the transport planning problem become evident. They are (a) design of the transport networks, (b) determination of interzonal movements for each transport flux, and (c) assignment of the interzonal movements to the transport networks.

Figure 1 shows the sequence of these phases of the transport planning problem within the framework of the total socio-economic system planning task. As shown, a mathematical representation of the "desired" or "optimum" living conditions is established by relating the various planning criteria to each other and to the socio-economic group. These relationships form the framework for a mathematical model of the transport plans, and are used to determine the extent to which any proposed transport plan is compatible with the "desired" living conditions. From the mathematical model, the transport networks are designed and loaded. Network loading is done in consonance with the mathematical relationships of the model that govern interzonal movement. Using the loaded networks as a basis, the cost (e.g., freedom of movement, esthetics, safety, as well as dollars) to the socio-economic group for using the proposed system is determined and compared to their desired goals.

The next section of this paper is a brief review of existing techniques for the forecasting phase (phase 2) and the assignment phase (phase 3) of the transport planning task.

EXISTING NETWORK TRANSPORT ASSIGNMENT AND PERFORMANCE EVALUATION TECHNIQUES

During recent years, the transport forecasting and transport assignment phases of the transport planning task have received considerable attention from the planning profession. The end result is a multiplicity of techniques for studying transport network performance. Transport forecasting involves the prediction of future travel patterns

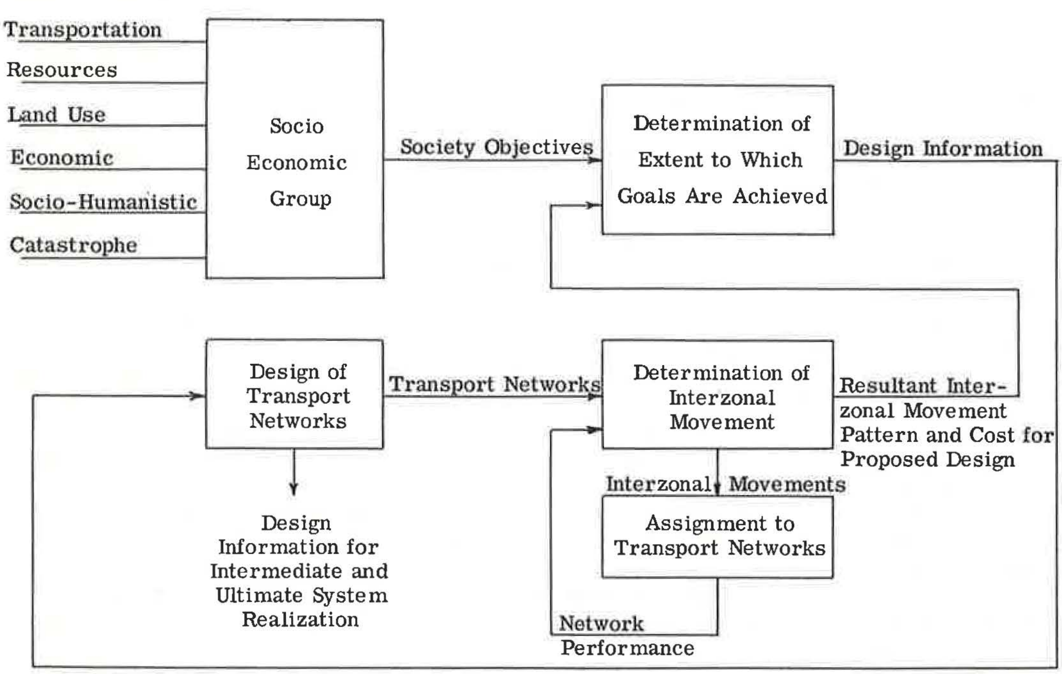


Figure 1. Socio-economic system planning.

and requirements for a transport system. Transport assignment, on the other hand, entails the determination of routes, within the proposed or existing transportation network, over which the forecast transport flows will move.

Transport Forecasting

Numerous theories and procedures have been developed to assist the transport planner in the task of predicting future transport requirements. The common denominator of these techniques is a recognition that future transport requirements depend on the ultimate configuration of the land use plan. In this respect, the type, intensity, and rate of development of the land use for each zone of a planned system greatly influences the evolution of every transport network within the system. If a region is to expand and grow according to some master plan, it is imperative that techniques be made available with which the system planner can determine how to provide transport facilities at a rate commensurate with total region expansion. To this end, transport forecasting and transport assignment work hand in hand; however, the best transport assignment technique cannot reliably determine network performance if the transport forecast is inaccurate.

The multitude of transport forecasting techniques fall within four broad classifications. These are growth factor models, gravity models, opportunity models, and multiple regression models.

Until recently, the most popular techniques were based on the growth factor model. These techniques accomplish the forecasting by multiplying the existing transport interzonal flows by predicted growth ratios. These ratios were determined by dividing the expected future transport demand by present transport demand. The original techniques did this, based on a system-wide ratio of expected growth, and thus the resulting forecasts were not accurate in zones where extreme differential growth occurred. This problem was very serious, particularly in metropolitan areas where urban development was replacing agricultural land use. In these situations, the demand for transport service at the time of a study could be less than one-thousandth of the expected demand required at the forecast time, yet the average growth factors would predict only small increases in transport demand.

To overcome these difficulties, methods were designed that provide for differential growth in the various study zones. The first such technique was developed by T. J. Fratar in 1952 in conjunction with a traffic survey in Cleveland, Ohio. Subsequently, modified versions of this technique have been proposed and used. Of these, the more well known are the Detroit method (5) and the average factor method.

Growth factor techniques have several severe handicaps. The most significant is that they ignore the distance between zones as well as the cost, congestion, speed, etc., of the routes over which the interzonal flows must travel. The end result is that it is as attractive for units of movement to move between two zones that are adjacent and directly connected as it is for them to move between zones at opposite extremities of the system which can be reached only by slow, costly, and lengthy paths. For some simple and geographically small systems, this characteristic will not be of major significance, but most transport systems do not fall within this category.

A second major handicap is that no provision has been made to consider bidirectional flow. Interzonal flows are estimated as a total flow without regard to direction. This makes it impossible to study the effect of peak system loads on the network because, for most transport systems, the directed peak loads are not in balance. Quite frequently, in complex transport systems, the relative magnitudes of the directed interzonal flows will differ by several orders.

Finally, the computer requirement for processing a moderate-size system (500 zones) is in the microsecond clock time, 32,000-word storage class. Even with machines such as the IBM 7090, the processing time will be approximately 10 minutes for one such forecast. To forecast large networks, such as those that would be required for metropolitan areas, approximately 10 to 20 times as many zones must be considered. Because the amount of high-speed storage required and the computing time used will increase as the square of the number of zones in the system, it is apparent that large systems cannot be processed by these techniques.

At present, the most widely used forecasting techniques are extensions of the basic gravity model. Gravity model techniques overcome the major objections of the growth factor model techniques first, by including friction factors between zone centroids to discourage undesirable trips, and second, by providing a mathematical structure that treats that interzonal flow as two separate directed flows.

Initially, the gravity model postulated that the transport flow desire between any two zones in a transport system is directly proportional to the product of the population in each zone and inversely proportional to some power of the distance between the centroids of the zones. In use, however, modifications were made on this basic model so that it would more closely approximate observed, real world, network performance. Friction factors were incorporated into the model whereby the flow between zones is governed, in part, by empirically determined functions of time, distance, etc.

The primary weakness of gravity model techniques, which is incidently a weakness of all existing forecasting techniques, is that the effect of network capacity on the forecasted flows is not considered. This problem has been partially overcome in several instances by the inclusion of an assignment procedure in the forecasting model. These techniques begin by assigning the forecasted interzonal flows to the transport network. Then, based on the resulting network congestion in the loaded network, new interzonal friction factors are determined. These friction factors are fed back to the forecasting model and a new forecast of interzonal flow is obtained. This cycle is repeated until stable interzonal forecasts are achieved.

A second weakness is that the effect of socio-economic factors on the system's transport requirements is not included in the gravity equations. This is not an inherent fault of the model, but stems from lack of information as to how these factors influence the transport requirement for each zone and socio-economic group. If these relationships could be determined, they can be incorporated into the friction factors of the gravity model equation, and will thereby greatly increase the reliability and accuracy of the forecasted interzonal movements.

A final fault is that the forecasted interzonal movements do not necessarily yield the required number of trip ends at each destination zone. This indicates that the interzonal friction factors were not in consonance with the actual impedance to flow between those zones. To overcome this problem, an iterative approach is used wherein the friction factors are modified between successive iteration passes. Convergence is generally achieved after two or three passes.

The opportunity model, sometimes called the Chicago method, was developed by the Chicago Area Transportation Study group (CATS) and is based on the following premise: Total travel time from each origin zone to all destination zones is minimized, subject to the condition that every destination zone has a stated probability of being accepted as a trip end if it is considered. To simplify the model, these probabilities are assumed to be constant.

The premise has been restated as follows: A trip prefers to remain as short as possible, but its behavior at any destination node in the network is determined by the probability of it ending at that node. If the nearest destination node is an unacceptable trip end, the trip must consider the next nearest destination node, and if that is unacceptable, consider the next nearest, and so on.

This technique has essentially the same limitations and disadvantages as gravity model techniques have. In use by CATS, however, it has achieved good success. When compared with seven different gravity model forecasts in a test forecast study, the Chicago method did a better job of forecasting known interzonal movements.

The multiple regression model, developed by the California Division of Highways, approaches the forecasting task in a manner considerably different than the approaches used by the previously mentioned models. This model consists of a system of equations, each of the same form, with coefficients determined from empirical transport performance data. After the coefficients have been determined, a forecast of current interzonal flows is made and compared to actual observed flows. If they do not agree, the form of the equations is changed, either by adding more or higher order terms, until the current forecast agrees with the observed current interzonal flows.

The most serious limitations of the multiple regression forecasting model are computation time and reliability of equation form. Because of the complexity of the regression technique, its computation time will be between 10 and 20 times as great as that required for any of the previously mentioned techniques. Because the time required for computing the regression coefficients increases as the square of the number of zones in the network, the upper limit to network size with existing computer techniques is in the neighborhood of 1,000 zones.

The question of validity of the equation form results from the nature of the technique. Equations that fit the survey year data may predict lower or in some cases negative interzonal flows for the forecast year. This can be caused by negative regression coefficients for the "growth" characteristics, and may occur, even when all characteristics are positive, for any of the following reasons: The variables associated with the various characteristics are highly correlated, the zone variables are not descriptive of the zone, or an incorrect equation form has been used.

On the positive side, in several different network studies, comparison forecasts with the gravity and growth factor techniques were made, and the multiple regression technique yielded the best forecasts in each instance.

In summary, of the existing techniques, the gravity and the opportunity models offer the most promise for future research in the transport forecasting field. These techniques entail relatively simple computational procedures and provide for considerable flexibility in choice of equation form. Where research could do the most good, is the area of developing relations between the many variables affecting the desirability or probability of a trip transfer occurring between pairs of zones.

Transport Assignment

The transport assignment phase of the transport planning task has received considerably more attention than has the transport forecasting phase. In general, transport assignment consists of loading a proposed network and then determining how the loaded network performs. Most assignment techniques begin with the allocation of forecasted interzonal flows to the interconnecting links of a proposed transport network. When this is complete, system performance is checked with the planned goals of the system. Measures such as average travel time per unit of distance, cost per unit of movement, and density of flow per network link can be evaluated and used to locate regions within the system where network design improvements are required if the proposed system goals are to be achieved.

In accomplishing the preceding, the usual assignment criterion is to assign interzonal flow such that total network operating cost will be minimized. Unfortunately, most existing techniques fall far short of accomplishing this goal. The underlying problem is that all techniques employ some form of "shortest route" algorithm wherein the interzonal flows are assigned to a supposed shortest or best route interconnecting each zone pair.

It is evident that the determination of which routes are the "shortest" or "best" within a complex network is not a simple task. The best path between two zones when the network is moderately loaded does not necessarily remain the best path as network loading increases. Assignments based on network link costs for any predetermined value of link flow will, in general, be incorrect for other values of link flow.

To date, the mathematical models proposed for the assignment task are very limited in number. Most techniques are digital in nature and can be grouped into two groups: shortest route tree techniques and matrix techniques. All are based on the doubtful assumption that link loading does not affect the selection of route for a given interzonal flow. Only three investigators have attempted to modify these basic techniques to circumvent this problem.

Within the shortest route tree techniques, there are several basic digital algorithms as well as a couple of analog methods. The analog methods however, are not applicable to transport planning and are not discussed. In general, shortest route tree techniques determine the "shortest" or "best" route tree (distance, time, cost, or other arbitrary parameter or combinations thereof may be used as the "shortest" criterion) from any single node to all other nodes in a network. The shortest route tree is defined as the set of links and nodes, connected such that no loops occur, that constitute the least cost paths from the origin or starting node to all of the other nodes in the network. In networks where more than one path between a pair of nodes has the same least cost, only one of these paths is included in the shortest route tree.

The first algorithm for determination of the shortest route tree, suggested by Dantzig (6), is based on the simplex method, and is very slow and cumbersome to apply. It is generally used only for simple hand solutions. The techniques of Minty (24) and Moore (22) are considerably more efficient, and are suitable for computer solution. Both Minty and Moore have formulated their algorithms so that the links of the networks are directional. Because very few practical transport networks are symmetrical, and because each link in these networks can usually be traversed only in one direction, directional links are an essential part of any transport network. Of the two, Moore's algorithm is less time consuming and easier to implement, a fact that makes it attractive for large complex networks.

The problem of simultaneously finding the shortest route between every pair of nodes in a network was first solved by Shimbel (29). Subsequently, Bellman (2) formulated a matrix method algorithm for obtaining solutions according to Shimbel's technique. Although Bellman's method is applicable to the solution of large transport network assignment problems, it is not feasible to use, because the amount of computer memory and computation time required for the determination of all minimum paths in a network is too great. To date, those who have developed computer programs for the transport assignment task have used techniques based on the repeated application of Moore's algorithm.

To present a complete discussion of each transport assignment program presently in use would be too voluminous for this paper. Instead, Table 1 gives the relative capabilities and limitations of the various techniques. Because some of the programs listed include transport forecasting techniques as a part of the transport assignment package, they are so identified. Detailed discussions of each listed program are found in a report by Mertz (7).

Most of the existing techniques are limited as to size of network that can be studied (Table 1). Even more disconcerting is the fact that only three of the programs are published, and of these only two are programed for computers available in the United States. Of the two, only the Washington, D. C., program is of adequate size to study complex networks.

The most serious limitation of these existing programs is treatment of the capacity-restraint problem. Only three of the systems consider the problem at all, and then only in a limited sense.

The Detroit Arterial system, after the completion of the assignment cycle, determines all paths that are loaded in excess of their maximum capacity. It then redistributes the loads for these paths to several of the next best paths in proportion to their ability to receive them.

The Chicago system treats capacity restraint on a zone-by-zone basis. After the trips from one zone are distributed to their destinations in accordance with the appropriate minimum path tree, new link impedances are computed from the partially loaded network. From these new link impedances, the minimum path tree for another zone is determined, and the trips from this zone are assigned accordingly. This cycle is repeated until all trips have been assigned.

Traffic Research Corporation, Ltd., uses a relaxation technique whereby trips are assigned to as many as four different paths between each pair of zones, and are assigned in a manner such that the loaded path costs are equal for all paths of each zone pair. To accomplish this, the first step is to load the minimum path trees based on their zero flow link impedances. Next, new link impedances are computed from the link flow volumes of the loaded network, and new minimum path trees are computed therefrom. Finally, the network is reloaded by assigning each interzonal flow to its original and new minimum path in inverse proportion to its respective path costs. This process is repeated until the system stabilizes with up to four paths allowed for each pair of zones.

TABLE 1

TRANSPORT ASSIGNMENT PROGRAMS

System or Organization	Computer	Number of Zone Centroids	Number of Nodes	Number of Links	Number of Links from Any Zone	Capacity Restraint	Forecasting Option	Published	Computer Time for Typical Job (hr)	Corresponding Typical Job
Chicago system	32K IBM 704	700	4,095	14,000	Unlimited	Optional	Opportunity	No	5	650 centroids, 4,000 nodes, 13,000 links, no capacity restraint
Washington, D. C., and Minnesota system	IBM 704[a]	Any node a centroid	4,000		4	No	Growth factor	Yes	5	500 centroids, 3,500 nodes
Detroit Expressway	IBM 650 with 2 tape units and 1 Ramac	400	75	425		No		No	20	30 nodes, 20,000 interzonal movements
Detroit Arterial	IBM 704	Any node a centroid	999			Yes		No	4	Maximum network size
Service Bureau Corp.	8K IBM 704	Up to 300 nodes a centroid	1,350		7	No	Gravity	No	6	250 centroids, 1,300 nodes
Traffic Research Corp., Ltd.	IBM 650[b] with tapes	100	1,800			Yes	Gravity	No	60	100 centroids, 250 nodes
State of Connecticut	Remington Rand File Model 1	Up to 145 nodes a centroid	10,000			No	Gravity	No	39	120 centroids, 1,400 nodes, 2,500 links
California	IBM 650[c]	Any node a centroid	699	1,000		No		No	125	699 nodes, 210 of which centroids
Missouri	IBM 650	200	791			No		Yes	120	270 centroids, 1,300 nodes
Road Research Lab.	Ferranti Pegasus	44		255		No		Yes	1	100 links

[a] Reprogramed to run on IBM 7090 with use of IBM 704-7090 compatibility program.
[b] Being reprogramed for IBM 7070.
[c] Being reprogramed for IBM 8K 704.

Of these three systems, only the Traffic Research Corporation system approaches a realistic solution. Granted that the other systems are better than ignoring capacity altogether, they leave much to be desired. Even the TRC system has several drawbacks, the most significant of which is the number of times the minimum path trees must be constructed before system stability is achieved. Tests of convergence made by Irwin et al. (14), indicate that approximately ten assignment cycles are required before the system stabilizes. For large networks, this could be very time consuming. Another questionable aspect of the technique is the limitation to four paths for any interzonal flow. In many transport networks it is entirely possible that ten, twenty or even a hundred alternate routes of near equal cost might exist. This is particularly true for flows between network zones that are geographically far apart. That all feasible alternate routes should be considered is self evident, yet they cannot be, for their inclusion would certainly increase, to an impractical extent, the number of assignment cycles required to achieve system stability.

In spite of these limitations, the TRC system is far more sophisticated than any of the others in Table 1. Some other interesting features of this system are forecasting feedback, land use feedback, and consideration of transport mode capacity restraints. The forecasting feedback option allows the forecast interzonal flows to be changed as a consequence of network congestion. In this manner, the forecasted flows will be consistent with the ultimate path costs of the loaded network. Similarly, units of movement may switch from one transport mode to another as a function of the loaded network cost for each transport mode. The land use feedback option allows the planner to determine the compatibility of the proposed transport plans and the land use plan. If ultimate land use realization is known as a function of freedom of movement within the system, changes in the land use plan, due to network performance, can be estimated and new assignments made.

In the next section of this paper a new approach to the transport assignment task is offered. It is an analytical matrix technique, and is designed to overcome the network assignment and analysis difficulties, inherent in existing assignment techniques, that are due to inadequate consideration of network link capacity-flow relations.

THE CAPACITY-RESTRAINT ALGORITHM

The primary feature of the capacity-restraint algorithm is its provision for arbitrary linear or nonlinear network link cost functions. In addition, it allows an entire transport network to be treated at once, with any number of permissible paths for each interzonal movement. Finally, because the technique is analytic, it will yield the solution for all paths of each interzonal movement, simultaneously. Provision has been made for limiting the permissible paths for any origin-destination pair to those meeting various desirability criteria, such as complexity, cost, and distance. Also, if network capacity is inadequate to accommodate any given interzonal movement, the saturated network links, responsible for each flow constriction, are determined and identified.

There are two basic options available with this assignment algorithm: (a) loading of the network to obtain optimum total system figure of merit, and (b) loading of the network so that for each pair of centroids the corresponding path figures of merit are equalized. These figures of merit can be any desired measure of network and path performance. The first option is appropriate for networks in which the route assignment for each unit of movement can be enforced. In networks in which each unit of movement is a free agent, the ultimate network assignment will be more realistically given by the second option. This latter option is applicable to such networks as those for highway transportation.

The following are several phases applicable to both assignment options:

1. Construction of the "connection matrix."
2. Expansion of the matrix as a set of modified determinants.
3. Determination of the "link cost" functions.
4. Determination of the "path cost" functions, elimination of undesirable paths, and making of preliminary checks for inadequate link capacity.

5. Establishment of the system of equations describing the network.
6. Elimination of path flow variables that must be zero to satisfy the boundary conditions; determination and identification of any links that are supersaturated.
7. If no links are supersaturated, solution of the system of equations.

All seven phases of this algorithm are extensions of existing mathematical techniques. Before the detailed procedures for each phase are presented, a brief description of the underlying theory and purpose of each phase is given. Perhaps the most interesting aspects of the algorithm are those of phases 1 and 2. The procedure used here was first described by Yoeli (30). To the author's knowledge, no concise, and simple explanation of this procedure has been published. Briefly, the technique determines all possible paths between each pair of nodes in a complex network in a manner such that all redundant paths are eliminated, and only paths that do not double back on themselves remain.

Link cost functions are established during phase 3. For each link, these are generally nonlinear functions of link loading, and represent the cost, time, safety, distance, etc., or combinations thereof, of using that link. Linear functions and constants, however, should be used whenever possible because their use reduces the overall complexity of the system of equations to be solved during phase 7.

Phase 4 of the algorithm serves three purposes: first, to determine the path cost functions from the link cost functions; then, to find links with inadequate link capacity; and last, to provide to the investigator the opportunity to eliminate paths from the network that are not satisfactory to him. The elimination of paths at this stage in the algorithm will greatly reduce the complexity of the equations to be solved, and thereby substantially reduce the required computer time.

The fifth phase consists of setting up a system of network describing nonliner equations. This set of equations will have more unknowns that equations, and will be solved in phase 7 either by the Lagrange method of undetermined multipliers (if total network cost is to be minimized), or, after writing equal path cost equations, as simultaneous nonlinear equations (if origin-destination flow path costs are to be equalized).

Phase 6 produces a preliminary interzonal flow assignment to determine the existence of any regions of the network wherein the links are saturated so that some of the network's interzonal flows are not accommodated. In such instances, the saturated links are tabulated and the analysis is terminated until the network is modified to provide the required additional capacity. In addition, if the first assignment option is selected, the zero flow and saturation flow boundary conditions are introduced during this phase. It is necessary to eliminate any variable from the system of equations for which the first partial derivative of the total network cost equation with respect to that variable does not vanish for some value within that variable's allowed range of variation.

The final phase of the algorithm consists of the solution of the system of equations. Because of the complexity of the equation, this cannot be done by analytical techniques. The method of solution proposed is that of successive evaluation and minimization of error, a technique that has been programed for the IBM 7090 and has been used successfully for problems of this type. A user's manual is available (33).

Following is a detailed description of the procedure for each phase of the capacity-restraint algorithm. These descriptions are intended to provide a base from which computer programs can be written.

Phase 1—Construction of Connection Matrix

To construct the connection matrix, the network must be represented as a lattice-weighted, directed, linear graph. This is accomplished by splitting the area served by the network into subareas of arbitrary size and configuration. Each subarea, referred to as a zone, is assigned a mnemonic which is associated with its mathematical centroid. These mathematical centroids, each of which is located at its corresponding zone's center of gravity are determined from the zone characteristics. After all zone centroids have been located, the network graph is constructed by interconnecting these centroids with the transport links of the proposed network. Finally, each link is assigned a mnemonic for each direction of allowed traversal. During phase 3, these mnemonics are replaced by mathematical functions of link flow.

Figure 2 is an example of a lattice-weighted, directed, linear graph for a transport network of 10 zones, interconnected by 21 directed links. The arrows on the graph indicate the allowed directions of movement over each link, and the associated mnemonics identify each directed link's use function. The network nodes are identified by mnemonics A through J.

The connection matrix for a network of N nodes will be a square matrix of order N, and will be constructed with the assistance of the lattice-weighted linear graph in the following manner. First, the rows and columns of the matrix are labeled with the network's node mnemonics. The rows of the matrix are arbitrarily chosen to represent origin nodes, and the columns represent destination nodes. Because each network zone can be both an origin and a destination zone, it is evident that the network node mnemonics of each zone centroid must appear as the label of one row and one column in the matrix. Next, the values for each cell of the matrix are entered directly from the lattice graph. This is accomplished by recording the link mnemonic for each origin-destination pair in a cell of the matrix located at the intersection of the link's origin row and destination column. If a particular destination cannot be reached directly from a given origin (because no direct link exists), the matrix cell corresponding to that origin-destination pair receives a zero entry. Matrix cells for which the origin and destination zones are the same receive an entry of one.

Using this procedure, the connection matrix for Figure 2 is shown in Figure 3. In the situation where each link is bidirectional and has the same use function in each direction, the entries $f_{i,j}$ and $f_{j,i}$ will be identical, and the matrix will be symmetrical. For most complex networks, however, this matrix will be nonsymmetrical.

Phase 2—Expansion of Matrix as a Set of Modified Determinants

To determine the routes of all paths from every origin to every destination in a network, the connection matrix is split into N^2 submatrices. When each submatrix is expanded as a determinant, an expression results which, after changing all arithmetic signs to a plus and simplifying by Boolean techniques, indicates all possible paths between a given origin and a given destination.[1] The submatrix for a particular origin-destination pair is obtained from the connection matrix by eliminating the matrix column corresponding to the given origin and the matrix row corresponding to the given destination. The elimination of the column corresponding to the origin centroid pre-

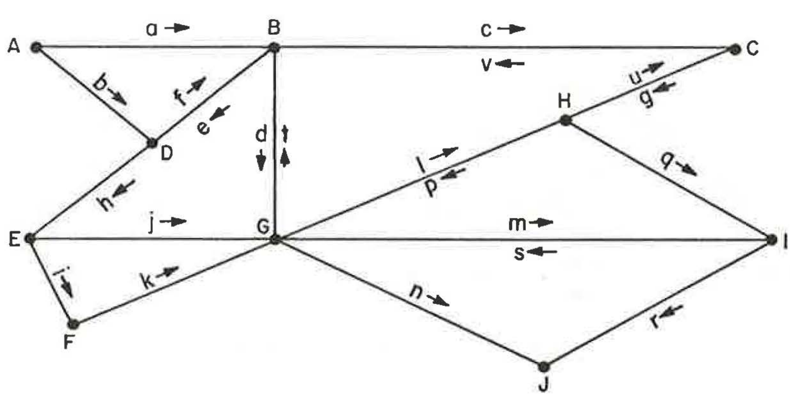


Figure 2. Lattice-weighted, directed, linear graph.

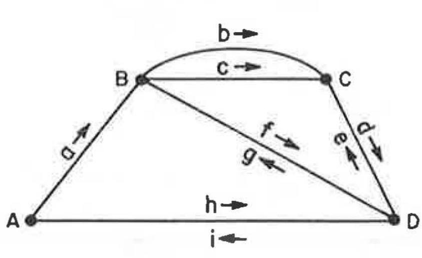
[1] In Boolean algebra techniques, the expression a + b is read a "or" b; thus if two points in a network are connected by two paths, for example a and b, the admissible paths between these two points are given by the expression a + b.

	A	B	C	D	E	F	G	H	I	J
A	1	a	0	b	0	0	0	0	0	0
B	0	1	c	e	0	0	d	0	0	0
C	0	v	1	0	0	0	0	g	0	0
D	0	f	0	1	h	0	0	0	0	0
E	0	0	0	0	1	i	j	0	0	0
F	0	0	0	0	0	1	k	0	0	0
G	0	t	0	0	0	0	1	ℓ	m	n
H	0	0	u	0	0	0	p	1	q	0
I	0	0	0	0	0	0	s	0	1	r
J	0	0	0	0	0	0	0	0	0	1

Figure 3. Connection matrix.

cludes that centroid from being used in any path as an intermediate destination. Likewide, the elimination of the row corresponding to the destination centroid prevents that centroid from occurring in any path as an intermediate origin.

After the Boolean expression for a submatrix has been simplified, it is rewritten as a first-order expression.[2] In this expression, each concatenation of mnemonics represents one possible path for the given origin-destination zone pair. Each such path can be traced on the network lattice graph by marking each link whose mnemonic appears in that concatenation. Because of the techniques used in expanding the submatrix and in the subsequent Boolean simplification, the sequence of the mnemonics in each concatenation does not necessarily correspond to the sequence in which the corresponding links are encountered when moving between the given origin and destination. The terms are, however, all-inclusive and nonredundant. That is, all links of a path are represented once and only once in that path's concatenation. Likewise, the simplified Boolean submatrix expansion is all-inclusive and nonredundant. Any redundant concatenations occurring as a result of the determinant expansion (i.e., concatenations that do not include the mnemonics for enough links to form a contiguous path from the



	A	B	C	D
A	1	a	0	h
B	0	1	b+c	f
C	0	0	1	d
D	i	g	e	1

Figure 4. Simplified network.

[2] A first-order expression is one that contains only concatenated link mnemonic terms separated by the Boolean "or." Each such term represents a path between the origin and destination zones of the submatrix from which it was obtained. For example, if the path between two zones consists of hypothetical links c, d, and e, the term representing this path could be written as (c)(d)(e), c·d·e, or simply cde. In any case, the concatenation indicates that the path consists of links c "and" d "and" e. It should be noted that the sequence of the link mnemonics within the path concatenation is not important.

origin zone to the destination zone, and concatenations that include some mnemonics more than once) are eliminated by the Boolean simplification procedure.

To illustrate this procedure, the simplified network shown with its connection matrix in Figure 4 is given. For example, to determine all paths from origin B to destination C, column B and row C are eliminated from the connection matrix. The resulting submatrix is expanded and simplified as shown in Figure 5.

As Figure 5 shows three paths exist from origin B to destination C: (a) the path consisting of link b; (b) the path consisting of link c; and (c) the path consisting of links f and e. Terms hbi and hci, both of which represent noncontiguous and therefore improper paths, were eliminated by the Boolean simplification of the expression. The Boolean algebra rule used to eliminate these terms was $(1 + X) = 1$. In the preceding case, the hbi term is eliminated by the b term, whereas the hci term is eliminated by the c term; for example, $b + hbi = b(1 + hi) = b$.

In a like manner, all paths for each remaining origin-destination zone pair can be found. Table 2 gives all paths between all origins and all destinations for the network shown in Figure 4.

Phase 3—Determination of Link Cost Functions

Because the link cost functions should include all factors influencing the performance of each network, the form of these functions may vary considerably for different transport fluxes. For some, the functions may be linear relations, whereas for others, highly complex mathematical relations might be required. Considerable research by economists, sociologists, engineers, and land use planners is necessary before accurate link cost functions can be formulated. For certain transport fluxes, however, these functions can be approximated to a degree sufficient for useful network performance studies. Because it is beyond the scope of this paper to determine or argue

	A	C	D
A	1	0	h
B	0	b+c	f
D	i	e	1

$$P_{BC} = b + c + fe + h(b+c)i$$
$$= b + c + fe + hbi + hci$$
$$= b + c + fe$$

Figure 5. Submatrix and resulting paths from B to C.

TABLE 2

ALL PATHS BETWEEN ALL ORIGIN-DESTINATION PAIRS FOR NETWORK OF FIGURE 4

Origin	Destination	Path
A	B	$P_{AB} = a + hg$
A	C	$P_{AC} = ab + ac + afc + he + hgb + hgc$
A	D	$P_{AD} = abd + acd + af + h$
B	A	$P_{BA} = bdi + cdi + fi$
B	C	$P_{BC} = b + c + fe$
B	D	$P_{BD} = bd + cd + f$
C	A	$P_{CA} = di$
C	B	$P_{CB} = dg + adi$
C	D	$P_{CD} = d$
D	A	$P_{DA} = i$
D	B	$P_{DB} = g + ai$
D	C	$P_{DC} = e + bg + cg + aib + aic$

the merits of various cost function formulations, the following discussion is limited to illustrating two or three proposed functions and their relationship to the capacity-restraint algorithm.

The most widely used link cost function is the constant. As mentioned previously, existing network analysis techniques with few exceptions, use constant link cost functions for all links. The few exceptions use "step" link cost functions, whereby the cost functions change, either at fixed points in the assignment process, or between successive iterative assignments to the network. Step link cost functions are suitable, for some links, in networks for certain of the "service" transport fluxes such as communications, water, and power where cost increases occur at discrete levels of flow.

Another function, more suitable for general use, is the "linear" link cost function. This function allows cost per unit of flow to increase linearly with flow, but it does not prevent inordinately high link loadings from occurring; therefore, link loadings in excess of link capacity can occur in densely loaded networks.

Because neither of the preceding functions is able to prevent flows in excess of link capacity from being assigned to the links, they are not suitable for the general transport link cost function. They can, however, be used to great advantage in the capacity restraint model whenever they closely approximate a particular network link cost function throughout its range of conceivable assigned usage. The advantage in these instances is that the network equations, to be solved at later stages in the algorithm, become simpler as the number of constant or linear link cost functions used is increased.

Two functions that have been suggested as reasonable candidates for the general transport link cost function are the hyperbolic and logarithmic functions. The characteristics of these functions are such that the change in "cost per unit of flow" as flow increases is small for low flow values, but large as saturation flow is approached. For both functions, both incremental "cost per unit of flow" and "total link cost" approach infinity as link saturation flow is approached.

The logarithmic link cost function is given by

$$f_i(m_i) = \log (M_i) - \log (M_i - m_i) + T_i \tag{1}$$

in which

$f_i(m_i)$ = cost per unit of flow for m_i units of flow;
m_i = number of units of flow using link i;
M_i = saturation flow for link i; and
T_i = zero flow cost for link i.

From Eq. 1, the cost as the limits of possible flow are approached is described by the following:

$$\text{As } m_i \to M_i, \qquad f_i(m_i) \to \infty$$

$$\text{As } m_i \to 0, \qquad f_i(m_i) \to T_i.$$

The logarithmic link cost function is shown in Figure 6.

The hyperbolic link cost function is

$$f_i(m_i) = \tau_i - \frac{M_i(T_i - \tau_i)}{m_i - M_i} \text{ and } \tau_i < T_i \tag{2}$$

in which

$f_i(m_i)$ = cost per unit flow for m_i units of flow;
m_i = number of units of flow using link i;
M_i = saturation flow for link i;
τ_i = function's cost asymptote; and
T_i = zero flow cost for link i.

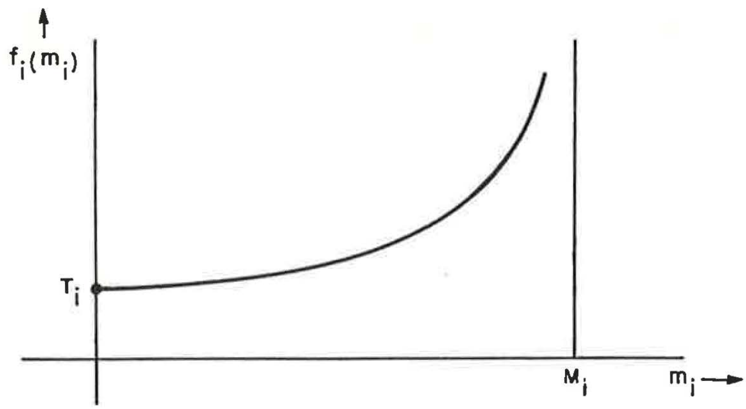


Figure 6. Logarithmic link cost function.

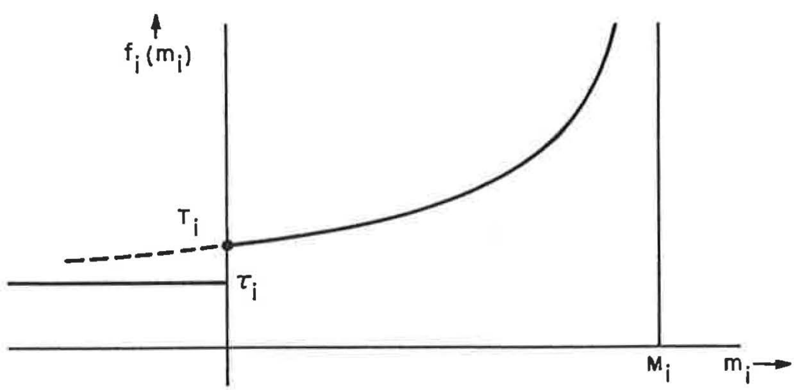


Figure 7. Hyperbolic link cost function.

The function cost asymptote, τ_i, controls the flatness of the function at low values of m_i. As τ_i decreases, the rate of change of $f_i(m_i)$ for low values of m_i increases, and for increasing values of τ_i, the rate of change of $f_i(m_i)$ decreases. The smallest possible rate of change for $f_i(m_i)$ occurs at the upper limit of τ_i, namely T_i. From the hyperbolic link cost function, the link costs as the limits of possible link flow are approached are given by the following:

$$\text{As } m_i \to M_i, \qquad f_i(m_i) \to \infty$$

$$\text{As } m_i \to 0, \qquad f_i(m_i) \to T_i$$

Figure 7 shows the hyperbolic link cost function. This function is used as the link cost function in the discussion of the remaining phases of the capacity-restraint algorithm. Zero flow cost for any link is assumed to be analogous to the zero flow travel time required to traverse that link and, for simplicity, it is assumed that the cost asymptote for all link cost functions is zero.

In Figure 4, an assignment of hypothetical forecast flows from centroid A to centroids B and D, and from centroid D to centroid B will be accomplished. These three flows will be assumed to be

$$M_{AB} = 25 \quad (3)$$

$$M_{AD} = 100 \quad (4)$$

$$M_{DB} = 450 \quad (5)$$

For the purpose of this example all other interzonal flows will be assumed zero. From Table 2, the paths for the three non-zero flows are

$$P_{AB} = a + hg \quad (6)$$

$$P_{AD} = abd + acd + af + h \quad (7)$$

$$P_{BD} = g + ai \quad (8)$$

From Eqs. 6 through 8, it is observed that link cost functions are required for links a, b, c, d, f, g, h, and i.

Table 3 gives, for each of these links, the assumed values of their zero flow travel time (T_i) and their saturation link flow (M_i). Also given for each link are the hyperbolic link cost function [$f_i(m_i)$], the unit flow cost [$f_i(1)$], and the maximum allowable number of units (max_i) that may be assigned to that link. The maximum flow limit, max_i, is one unit of flow less than M_i.

Phase 4—Determination of Path Cost Functions and Test Links

Phase 4 begins by assigning to each path a mnemonic that will identify that path and its subsequent flow. Once this is accomplished, each path cost function is obtained by adding the link cost functions for every link included in that path. Next, these path cost functions are evaluated for unit flow loading, thereby providing a means for comparing the various paths of any particular origin-destination pair at the zero flow boundary. The flow capacity of each path is then obtained by determining the smallest "maximum link flow" of any link in that path.

At this point in the algorithm, tests and conditions are applied to the paths which, if they are not satisfied, will eliminate those paths from further consideration. The number and type of such tests change for each type of network under consideration, and may vary in complexity to whatever extent the investigator is willing to apply logic to the system under study. The elimination of paths at this time will greatly simplify the balance of the analysis, and thereby decrease the required computational time.

The tests and conditions discussed next are not intended to be exhaustive. Many other considerations will become obvious as particular networks are studied.

In communication networks, a reasonable limitation to the number of links in a path is a realistic restriction. The number of links in any path is easily determined from the path equations obtained during phase 2. All paths consisting of more than a predetermined number of links might be eliminated from the list of allowable paths.

TABLE 3

LINK COST FUNCTIONS

Link	M_i	T_i (sec)	$f_i(m_i)$	$f_i(1)$	max_i
a	100	10	$\frac{1,000}{100 - m_a}$	10.10	99
b	200	5	$\frac{1,000}{200 - m_b}$	5.03	199
c	25	5	$\frac{125}{25 - m_c}$	5.21	24
d	200	10	$\frac{2,000}{200 - m_d}$	10.05	199
f	300	15	$\frac{4,500}{300 - m_f}$	15.05	299
g	500	10	$\frac{5,000}{500 - m_g}$	10.20	499
h	50	10	$\frac{500}{50 - m_h}$	10.20	49
i	500	4	$\frac{2,000}{500 - m_i}$	4.01	499

From the path use function evaluation, it is possible to limit the assignment to the best N paths between each pair of zones. These best paths are determined by comparing the path unit flow costs. Caution must be exercised in applying this type of restriction because it is entirely possible that when the network is loaded with all interzonal flows, the individual costs for some or all of these best paths will greatly exceed that of the best path that was discarded.

Paths may also be eliminated for reasons such as total cost is too great, distance is too long, or travel time is too great. Such restrictions require detailed knowledge of the link characteristics, but considerable information of this type is already available, because it is required in phase 3 to determine the link use functions. Excessive cost and travel time paths can be rejected at this point in the algorithm only on the basis of unit flow loading, or on approximate equal path cost loading.

As undesirable paths are eliminated from the list of allowable paths for any origin-destination pair, the total capacity of the balance of the paths for that origin-destination pair must be compared to the required interzonal flow over them. If these remaining paths do not have adequate excess capacity, the analysis must stop, and the network must be redesigned. The question of how much excess capacity there should be for a given interzonal flow cannot be determined analytically; after some experience has been gained with the use of this algorithm, however, it should be possible to estimate appropriate values.

Continuing the example of Figure 4, each path was assigned a mnemonic. The path cost functions were determined and evaluated for unit flow, and maximum path flows were obtained (Table 4).

Phase 5—Establishment of System of Equations

The system of equations to be solved during phase 7 is obtained from the network link use functions, the interzonal flow requirements, the network total cost equation, and a set of relations between path and link flow. Many of these equations are linear relations, and can be used to eliminate variables from the system of equations through simple substitution techniques. For solution by conventional manual techniques, much

TABLE 4

PATH COST FUNCTIONS

O-D Pair	Path	Path Mnemonic	Path Cost Function	Unit Flow Cost	Path Flow Capacity
A-D	abd	1	$\frac{1,000}{100 - m_a} + \frac{1,000}{200 - m_b} + \frac{2,000}{200 - m_d}$	25.18	99
	acd	2	$\frac{1,000}{100 - m_a} + \frac{125}{25 - m_c} + \frac{2,000}{200 - m_d}$	25.36	24
	af	3	$\frac{1,000}{100 - m_a} + \frac{4,500}{200 - m_f}$	25.15	99
	h	4	$\frac{500}{50 - m_h}$	10.20	49
A-B	a	5	$\frac{1,000}{100 - m_a}$	10.10	99
	hg	6	$\frac{500}{50 - m_h} + \frac{5,000}{500 - m_g}$	20.40	49
D-B	g	7	$\frac{5,000}{500 - m_g}$	10.20	499
	ai	8	$\frac{1,000}{100 - m_a} + \frac{2,000}{500 - m_i}$	14.11	99

simplification is achieved if this is done. Substitution is not advisable if computer techniques are to be employed for the algorithm, because the required computer program logic for accomplishing the substitution would be very complex, and the elimination of these variables would not greatly reduce the computer solution time.

The system of equations consists of three basic equation types. First, the relationship between the link cost functions and total network use cost yields

$$C_T = \sum_i m_i f_i(m_i) \tag{9}$$

in which

C_T = total network use cost;
i = index of summation (taking on mnemonic for each link in analysis);
m_i = number of units of movement assigned to link i; and
$f_i(m_i)$ = link cost function for link i evaluated for m_i units of movement.

An alternate form of this equation, found by summing the individual origin-destination cost equations, is

$$C_T = \sum_{p,q} C_{p,q} = \sum_{p,q} \sum_j N_j g_j \tag{10}$$

in which

C_T = total network use cost;
p, q = index of summation (taking on mnemonics for each origin-destination pair in analysis);
$C_{p,q}$ = total cost of flow from origin p to destination q by all paths interconnecting them;
j = index of summation (taking on mnemonics for each path of a given origin-destination pair);
N_j = number of units of movement using path j; and
g_j = path cost function for path j found by summing link cost functions of all links comprising that path.

Because in most network studies it is important to individually evaluate the total cost ($C_{p,q}$) for each interzonal flow of the network, this second form of the total cost equation is more convenient.

The next group of equations are restraining equations, and are formulated by setting each required interzonal flow equal to the sum of all of the path flows which could accommodate it. There will be as many equations of this type as there are interzonal flows in the network. These equations are of the form:

$$M_{p,q} = \sum_j N_j \tag{11}$$

in which

$M_{p,q}$ = required interzonal flow from origin p to destination q;
j = index of summation (taking on mnemonic for each path of a given origin-destination pair); and
N_j = number of units of movement using path j.

The last group of equations are also constraining equations, and are formulated by setting the link flow variable of each link equal to the sum of all path flows that traverse that link. There will be one equation of this type for every network link. These equations are of the form:

$$m_i = \sum_k N_k \tag{12}$$

in which

m_i = number of units of movement assigned to link i;
k = index of summation (taking on mnemonic for each path using link i); and
N_k = number of units of movement using path k.

Returning to Figure 4, Eq. 9 becomes

$$C_T = m_a f_a(m_a) + m_b f_b(m_b) + m_c f_c(m_c) + m_d f_d(m_d) +$$

$$m_f f_f(m_f) + m_g f_g(m_g) + m_h f_h(m_h) + m_i f_i(m_i) \tag{13}$$

Eq. 10 and its individual total interzonal cost functions are written, with the help of Table 4, as follows:

$$C_T = C_{AB} + C_{AD} + C_{DB} \tag{14}$$

$$C_{AB} = N_5\left[f_a(m_a)\right] + N_6\left[f_h(m_h) + f_g(m_g)\right] \tag{15}$$

$$C_{AD} = N_1\left[f_a(m_a) + f_b(m_b) + f_d(m_d)\right] + N_2\left[f_a(m_a) + f_c(m_c) + f_d(m_d)\right] + N_3\left[f_a(m_a) + f_f(m_f)\right] + N_4\left[f_h(m_h)\right] \tag{16}$$

$$C_{DB} = N_7\left[f_g(m_g)\right] + N_8\left[f_i(m_i) + f_a(m_a)\right] \tag{17}$$

Next, from Eq. 11, using Table 4 and Eqs. 3, 4, and 5, three constraining equations are

$$M_{AB} = N_5 + N_6 \tag{18}$$

$$M_{AD} = N_1 + N_2 + N_3 + N_4 \tag{19}$$

$$M_{DB} = N_7 + N_8 \tag{20}$$

Last, from Eq. 12, using Table 4, eight more constraining equations are

$$m_a = N_1 + N_2 + N_3 + N_5 + N_8 \tag{21}$$

$$m_b = N_1 \tag{22}$$

$$m_c = N_2 \tag{23}$$

$$m_d = N_1 + N_2 \tag{24}$$

$$m_f = N_3 \tag{25}$$

$$m_g = N_6 + N_7 \tag{26}$$

$$m_h = N_4 + N_6 \tag{27}$$

$$m_i = N_8 \tag{28}$$

Eqs. 13 through 28 now define the entire network, but inasmuch as there are more unknowns than equations, an explicit solution of them is not possible. If the first assignment option is chosen, this situation will be handled by the Lagrange method of undetermined multipliers, and a solution will be found that minimizes the total network cost function, C_T. If the second assignment option is chosen, additional equations will be written from the path cost functions (g_j) of Eqs. 15 through 17. In this case, each path cost function in each of these equations (the bracketed expressions in each equation are the path cost functions) will be set equal to an undetermined constant. These constants will be the same for all path cost functions of a given interzonal cost equation, but different for those of different interzonal cost equations. When this is accomplished, there will be as many unknowns as equations and an explicit solution for them will be possible. This solution will load the network so that, for each pair of network zones, the path costs per unit of flow of each of its paths will be equalized.

Before these equations can be solved, boundary conditions must be applied and the network must be tested for supersaturation.

Phase 6—Testing for Supersaturated Links and Applying Boundary Conditions

At this point in the algorithm, the network is loaded by assigning all flow between each pair of origin and destination zones to the best paths (least cost for unit flow paths) between them. From the required interzonal flows, the list of permissible paths (determined in phase 3), and the list of unit flow path costs (determined in phase 4), the network assignment is accomplished as follows:

1. All interzonal flows for origin-destination pairs having only one admissible path are assigned.
2. Interzonal flows for which there is more than one admissible path, are assigned in sequence to increasing number of admissible paths.
3. The interzonal flow for each multi-path origin-destination pair is assigned to its best zero flow path until that path is saturated. Any remaining flow is assigned to the next best paths, in order of their desirability, until either the entire flow is assigned or all allowable paths for that flow are saturated.
4. After each interzonal flow has been assigned, the cumulative link loadings must be determined. Whenever a link becomes saturated, all paths, for interzonal flows not yet assigned, which use that link are temporarily eliminated from the list of admissible paths. If any interzonal flow cannot be accommodated because all of its paths contain at least one saturated link, these saturated links and that interzonal flow are noted, and the additional capacity required to accommodate the flow is determined.

When the network capacity is inadequate to satisfy all interzonal flows, the analysis terminates at the completion of the preceding assignment. The tabulation of saturated links and unsatisfied interzonal flows can be used either to redesign the transport network, or to indicate where changes must be made in the land use plan.

If the assignment is successful, and the second assignment option has been selected, phase 6 is complete. The next step is to obtain the solution of the system of equations by the procedure in phase 7.

When the first assignment option is chosen, phase 6 continues. The next step is to modify the system of equations for the network such that the conditions imposed by the Lagrange technique are satisfied. Because it is desired to find variable values that will minimize total network cost, and because the necessary condition for the existence of such an extremum for a differentiable function is the vanishing of the first partial derivatives of the function with respect to its independent variables (29), it is necessary to determine the behavior of each first partial derivative of the cost function throughout the admissible range for its corresponding variable. Each first partial derivative of the cost function must vanish for an admissible value of its corresponding variable.

The differentiation and solution of complex functions is a time-consuming and difficult task to accomplish with a digital computer, especially when the form of the equation can change from application to application. Because the nature of this algorithm is such that a digital computer must be used, and because each network studied will undoubtedly have equations of a form peculiar to itself, it is not feasible to use classical methods to obtain the first partial derivatives of the network cost equation. Instead, a procedure that requires only the repeated evaluation of the cost equation and a comparison of the resulting values has been devised.

As background, the total network cost equation consists of the arithmetic sum of the products of link flow and the corresponding link cost functions. The link cost functions are monotonic increasing functions of the link flow variables. Each link flow variable is a dependent variable, and is dependent on the path flow variables of all paths using that link. All path flow variables that received assigned values during the preliminary assignment are also dependent variables, and are dependent on the network's interzonal flows and link saturation values. The path flow variables that received no assigned flow during the preliminary assignment are the independent variables of the system of equations, and are, therefore, the variables for which the first partial derivatives of the cost equation must vanish. Any such variable, for which the corresponding partial derivative does not vanish for an admissible value of that variable, must be set equal to zero and eliminated from the system of equations. This implies that the minimum total network cost is achieved when these variables are zero, and therefore when no flow is assigned to their corresponding paths. All link flow variables that are functions only of eliminated path flow variables must also be dropped from the equations. Finally, because the link cost functions are monotonic increasing functions of flow, their evaluation as saturation link flow is approached yields successively higher values of cost per unit of flow.

Keeping the preceding facts in mind, the path flow variable test procedure is as follows:

1. The total network cost function is evaluated for the assigned flows that were determined by the preliminary assignment of this phase of the algorithm.
2. These assigned path flows are decremented one at a time, by one unit of flow, and a new total cost is computed for each case.
3. These new total costs are grouped by origin-destination pairs.
4. For each zone pair group, the path flow variable which when decremented yields the lowest total network cost is determined. These path flow variables are used as reference variables for the balance of this test procedure, and are used at their decremented value whenever they are used as reference.
5. By zone pair group, the path flow variables that received no initial assignment are incremented one at a time by one unit of flow. Using the appropriate reference variable, the new total network cost is computed. (At this point in the procedure, situations may occur where some link loadings are incremented to their saturation boundary limit. Whenever this happens, one of the other path flow variables using that link must be decremented by one unit, and a corresponding path flow variable not using that link must be incremented by one unit of flow. This must be done in a manner such that the decremented variable produces the maximum decrease in total network cost, and the incremented variable produces the minimum increase in total network cost.)
6. The total network costs computed in step 5 are now compared to the total network cost computed in step 1. If the new cost is higher than the original cost, the corresponding path flow variable will be zero and it is eliminated from the system of equations. If the new cost is less than the old cost, the corresponding path flow variable remains in the system of equations. Variables passing this test, and thereby remaining in the analysis, will not necessarily be assigned non-zero values in the final assignment. Because units of movement are integral entities, the final path flows will be rounded off to their nearest integral values, thus it is entirely possible that some paths will be assigned zero flow.

The following is offered as proof that the preceding procedure eliminates all variables for which the corresponding first partial derivatives of total network cost do not vanish for an admissible value of those variables, and furthermore that it eliminates no other variables. As outlined, the procedure determines the differential of the total cost function at each independent variable's zero flow boundary. If the increment used in determining this differential cost is allowed to approach zero, the true derivative at the boundary is obtained. Now, if the partial derivative of the total cost function is negative at a variable's zero flow boundary, it is only necessary to show that this derivative becomes positive somewhere in the admissible range of that variable in order to establish the existence of a minimum extremal value of the function for that variable. Because any partial derivative of the function is always positive at the saturation flow boundary of its corresponding variable (because total cost approaches infinity as a variable approaches its saturation value), it follows that any variable that has a negative derivative at the zero flow boundary has an admissible value for which the corresponding first partial derivative will vanish.

To complete this proof, it is necessary to show that if, at any variable's zero flow boundary, the partial derivative of the total cost equation is positive, it will remain positive throughout that variable's admissible range. Now, because the path use functions are monotonic increasing functions of link loading, and because the derivatives of the individual link use functions are also monotonic increasing, it follows that the derivatives of the path's use functions are monotonic increasing. Because the total cost equation is given by the sum, for all network paths, of the products of path flow and the corresponding path use function, it is apparent that the partial derivative of this equation with respect to any path flow variable will be monotonic and will be arithmetically smallest at the zero flow boundary of that variable. Therefore, if the differential of total cost with respect to a particular variable is positive at that variable's zero flow boundary, it will remain positive throughout that variable's admissible range.

To illustrate this phase of the algorithm, the network of Figure 4 is considered. From the required interzonal flows of Eqs. 3, 4, and 5, and the unit flow path costs of Table 4, a preliminary assignment is made in accordance with the rules for part 1 of this phase. The resulting flows are as follows:

$N_1 = 0$ $\qquad$ $N_5 = 25$

$N_2 = 0$ $\qquad$ $N_6 = 0$

$N_3 = 51$ $\qquad$ $N_7 = 450$

$N_4 = 49$ $\qquad$ $N_8 = 0$

From these path flows, and Eqs. 21 through 28, the individual link flows are as follows:

$m_a = 76$ $\qquad$ $m_f = 51$

$m_b = 0$ $\qquad$ $m_g = 450$

$m_c = 0$ $\qquad$ $m_h = 49$

$m_d = 0$ $\qquad$ $m_i = 0$

Because all interzonal flow requirements are satisfied, the analysis may advance to the test procedure of part 2 of this phase. For step 1 of the test procedure, the total network cost function is evaluated with the help of the path cost functions given in Table 4, using Eq. 9 as follows:

$$C_T = \sum_i m_i f_i(m_i)$$

$$= m_a\left[\frac{1,000}{100-m_a}\right] + m_b\left[\frac{1,000}{200-m_b}\right] + m_c\left[\frac{125}{25-m_c}\right] +$$

$$m_d\left[\frac{2,000}{200-m_d}\right] + m_f\left[\frac{4,500}{300-m_f}\right] + m_g\left[\frac{5,000}{500-m_g}\right] +$$

$$m_h\left[\frac{500}{50-m_h}\right] + m_i\left[\frac{2,000}{500-m_i}\right]$$

$$= 3,167 + 0 + 0 + 0 + 922 + 45,000 + 24,500 + 0$$

$$= 73,589 \qquad (29)$$

TABLE 5

UNIT DECREMENTED VARIABLE VS RESULTING NETWORK COST

Zone Pair	Decremented Variable	Resulting Network Cost
A-D	N_3	73,400
	N_4	61,089
A-B	N_5	73,422
D-B	N_7	72,609

The results of the next three steps of the test procedure are given in Table 5.

The results of step 5 of the test procedure are given in Table 6. In the case of path 6, the special situation where a link flow cannot be incremented occurred. From Eq. 27 it was noted that $m_h = N_4 + N_6$, and because N_4 already used all available capacity for link h, it was necessary to decrement N_4 by one unit before N_6 could be incremented. As a result of this, it was necessary to increase one of the alternate path flows for zone pair A-D by one unit in order to maintain the required A to D interzonal flow. Because there are three such alternate paths to choose from (namely, paths 1, 2, and 3), all were individually incremented and the respective total network costs were computed. These costs are 61,104, 61,104, and 61,111. Because the minimum of these total costs occurred for both path 1 and path 2, the path 1 flow was arbitrarily chosen to be incremented. The total network cost shown in Table 6 for incremented N_6 was therefore computed with N_4 decremented by 1 and N_1 incremented by 1.

Comparing the total network costs of Table 6 to the original total network cost of Eq. 29 shows that variable N_6 must be set equal to zero and eliminated from the system of equations. Variables N_1, N_2, and N_8 must remain in the system of equations.

Phase 7—Solution of System of Equations

The equation solution procedure of this phase of the algorithm is entirely different for the two assignment options, and is therefore presented as two distinct processes. For the first assignment option, which requires total network cost to be minimized, the Lagrange technique of undetermined multipliers is used. The second assignment option, which requires the path costs for each origin-destination group of paths to be equalized, employs an iterative technique to solve a consistent set of equations for the flow variables.

The procedure for the first option begins by eliminating from the system of equations of phase 5 all variables so destined by phase 6. When this is complete, the Lagrange technique is employed to create a set of consistent equations. Finally, these equations are solved by the method of successive evaluation and minimization of error. The detailed steps are as follows:

1. Appropriate variables are eliminated from the equations of phase 5, and the constraining equations rewritten setting them equal to zero.
2. The differentials of the total network cost equation and all of the constraining equations of step 1 are determined, and each differential is set equal to zero.

TABLE 6

UNIT INCREMENTED VARIABLE VS RESULTING NETWORK COST

Zone Pair	Incremented Variable	Reference Decremented Variable	Resulting Network Cost
A-D	N_1	N_4	61,284
	N_2	N_4	61,284
A-B	N_6	N_5	74,624
D-B	N_8	N_7	72,829

3. Each differentiated constraining equation is multiplied by an undetermined function, λ, where λ is a different function for each equation. These functions are the Lagrangian multipliers.

4. All equations of step 3 are added to the differentiated total cost equation of step 2, and terms collected according to derivative.

5. Each such collection of terms is set equal to zero. These equations, together with the constraining equations of step 1, constitute a consistent set of equations, and may be solved by conventional techniques.

6. The system of equations is simplified and solved by the method of successive evaluation and minimization of error.

Concluding the example problem of Figure 4, the preceding procedure is implemented as follows: During phase 6 it was determined that N_6 must be eliminated from the system of equations; therefore, from Eqs. 18 through 28 the new constraining equations are

$$0 = M_{AB} - N_5 \quad (30)$$

$$0 = M_{AD} - N_1 - N_2 - N_3 - N_4 \quad (31)$$

$$0 = M_{DB} - N_7 - N_8 \quad (32)$$

$$0 = m_a - N_1 - N_2 - N_3 - N_5 - N_8 \quad (33)$$

$$0 = m_b - N_1 \quad (34)$$

$$0 = m_c - N_2 \quad (35)$$

$$0 = m_d - N_1 - N_2 \quad (36)$$

$$0 = m_f - N_3 \quad (37)$$

$$0 = m_g - N_7 \quad (38)$$

$$0 = m_h - N_4 \quad (39)$$

$$0 = m_i - N_8 \quad (40)$$

Next, the total network cost equations (Eq. 9) and the constraining equations (Eqs. 30 through 40) are differentiated. Because each term of Eq. 9 is of the same form, given in general by Eq. 2, the differential of the equation is computed for the general case:

$$dC_T = 0 = \sum_i \frac{\partial\left[m_i f_i(m_i)\right]}{\partial m_i} dm_i$$

$$= \sum_i \left\{ m_i \left[\frac{M_i(T_i - \tau_i)}{(m_i - M_i)^2} \right] + \left[\tau_i - \frac{M_i(T_i - \tau_i)}{m_i - M_i} \right] \right\} dm_i \quad (41)$$

in which i = a, b, c, d, f, g, h, i.

If, as for this example, $\tau_i = 0$, Eq. 41 becomes

$$dC_T = \sum_i \frac{M_i^2 T_i}{(m_i - M_i)^2} dm_i \quad (42)$$

Now as a result of eliminating path 6, N_5 becomes a constant and can be eliminated from the system of equations by substituting its value into each equation in which it occurs. This is done because it will reduce the number of equations and thereby simplify their solution. From Eqs. 30 through 40, using $N_5 = M_{AB}$, the results of step 2 and step 3 are

$$0 = \lambda_1(-dN_1 - dN_2 - dN_3 - dN_4) \quad (43)$$

$$0 = \lambda_2(dm_a - dN_1 - dN_2 - dN_3 - dN_8) \quad (44)$$

$$0 = \lambda_3(dm_c - dN_1) \quad (45)$$

$$0 = \lambda_4(dm_d - dN_2) \quad (46)$$

$$0 = \lambda_5(dm_d - dN_1 - dN_2) \quad (47)$$

$$0 = \lambda_6(dm_f - dN_3) \quad (48)$$

$$0 = \lambda_7(dm_g - dN_7) \quad (49)$$

$$0 = \lambda_8(dm_h - dN_4) \quad (50)$$

$$0 = \lambda_9(dm_i - dN_8) \quad (51)$$

Adding Eqs. 43 through 51 to Eq. 42, then collecting terms by corresponding differentials, and finally setting each collection of terms equal to zero (steps 4 and 5) yields the following system of equations. The values T_i and M_i used in Eq. 42 are given in Table 3.

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_5 = 0 \quad (52)$$

$$\lambda_1 + \lambda_2 + \lambda_4 + \lambda_5 = 0 \quad (53)$$

$$\lambda_1 + \lambda_2 + \lambda_6 = 0 \quad (54)$$

$$\lambda_1 + \lambda_8 = 0 \quad (55)$$

$$\lambda_7 + \lambda_{11} = 0 \quad (56)$$

$$\lambda_2 + \lambda_9 + \lambda_{11} = 0 \quad (57)$$

$$\lambda_2 + \frac{100,000}{(m_a - 100)^2} = 0 \quad (58)$$

$$\lambda_3 + \frac{20,000}{(m_b - 200)^2} = 0 \quad (59)$$

$$\lambda_4 + \frac{3,125}{(m_c - 25)^2} = 0 \quad (60)$$

$$\lambda_5 + \frac{400,000}{(m_d - 200)^2} = 0 \quad (61)$$

$$\lambda_6 + \frac{1,350,000}{(m_f - 300)^2} = 0 \quad (62)$$

$$\lambda_7 + \frac{2,500,000}{(m_g - 500)^2} = 0 \quad (63)$$

$$\lambda_8 + \frac{25,000}{(m_h - 50)^2} = 0 \quad (64)$$

$$\lambda_9 + \frac{1,000,000}{(m_i - 500)^2} = 0 \quad (65)$$

These equations, together with the constraining equations (Eqs. 30 through 40), can now be solved. Because the preceding equations are relatively simple, it is not difficult to combine them so that all the variables are eliminated. This simplification will always be easy to accomplish, regardless of network or link cost function complexity.

In this example, Eqs. 55, 56, and 58 through 65 are substituted into Eqs. 52 through 54 and 57. The resulting equations, together with the equations obtained by substituting Eqs. 3, 4 and 5 into Eqs. 30 through 40 are the system of equations to be solved.

$$\frac{25,000}{(m_h - 50)^2} - \frac{100,000}{(m_a - 100)^2} - \frac{200,000}{(m_b - 200)^2} - \frac{400,000}{(m_d - 200)^2} = 0 \quad (66)$$

$$\frac{25,000}{(m_h - 50)^2} - \frac{100,000}{(m_a - 100)^2} - \frac{3,125}{(m_c - 25)^2} - \frac{400,000}{(m_d - 200)^2} = 0 \quad (67)$$

$$\frac{25,000}{(m_h - 50)^2} - \frac{100,000}{(m_a - 100)^2} - \frac{1,000,000}{(m_i - 500)^2} = 0 \quad (68)$$

$$\frac{2,500,000}{(m_g - 500)^2} - \frac{100,000}{(m_a - 100)^2} - \frac{1,000,000}{(m_i - 500)^2} = 0 \quad (69)$$

$$25 - N_5 = 0 \quad (70)$$

$$100 - N_1 - N_2 - N_3 - N_4 = 0 \quad (71)$$

$$450 - N_7 - N_8 = 0 \quad (72)$$

$$m_a - N_1 - N_2 - N_3 - N_5 - N_8 = 0 \quad (73)$$

$$m_b - N_1 = 0 \quad (74)$$

$$m_c - N_2 = 0 \quad (75)$$

$$m_d - N_1 - N_2 = 0 \quad (76)$$

$$m_f - N_3 = 0 \quad (77)$$

$$m_g - N_7 = 0 \quad (78)$$

$$m_h - N_4 = 0 \quad (79)$$

$$m_i - N_8 = 0 \quad (80)$$

The solution of these equations yielded the values given in Table 7. Because fractional units of flow are not admissible, the nearest integral value for each flow is also given on the table.

From Eq. 10, the total cost for each interzonal flow and the total network cost for the loaded network are computed. Table 8 shows these results for both the computed path flow values and the nearest integral path flow values.

If the second assignment option is chosen, the solution procedure begins by writing the equal path cost equations described in the discussion of phase 5. These equations, together with the constraining equations (Eqs. 18 through 28) are then solved for the path flow variables. The following are the steps in detail:

1. The equal path cost per unit of flow equations are written. These are obtained from the path composition information of Eqs. 6, 7, and 8 and the link use functions given in Table 3, or from the path use functions given in Table 4, by setting each path use function equal to an appropriate unknown path cost constant.

2. A system of equations is set up consisting of the network constraining equations and the equations of step 1.

3. This system of equations is solved by the method of successive evaluation and minimization of error. Because this method of solution adjusts the values of the path flow variables between each equation evaluation cycle, it is possible that, as convergence to the solution is approached, some path flow variables will be reduced to zero. If this occurs, and the corresponding equation's zero flow path cost is higher than the loaded cost of every other path of the same zone pair, then that equation is temporarily eliminated from the system of equations. Equations so eliminated are reinserted at later equation evaluation cycles, only if the path costs for the loaded paths of that equation's zone pair become larger than the eliminated path's zero flow value.

TABLE 7

VALUES OF PATH FLOW VARIABLES FOR FIRST ASSIGNMENT OPTION

Path		Nearest Integral Flow
N_1	20.21	20
N_2	2.55	3
N_3	32.95	33
N_4	44.29	44
N_5	25.00	25
N_6	0	0
N_7	442.27	442
N_8	7.73	8

TABLE 8

TOTAL INTERZONAL FLOW COST AND TOTAL NETWORK COST FOR FIRST ASSIGNMENT OPTION

Zone Pair	Computed Flow Total Cost	Nearest Integral Flow Total Cost
A-B	2,163	2,273
A-D	9,637	9,702
D-B	39,005	38,865
Total	50,805	50,840

For this assignment option, the example problem is concluded, first, by writing the equal path cost per unit of flow equations:

Path 1

$$\frac{1,000}{100 - m_a} + \frac{1,000}{200 - m_b} + \frac{2,000}{200 - m_d} = K_{AD} \tag{81}$$

Path 2

$$\frac{1,000}{100 - m_a} + \frac{125}{25 - m_c} + \frac{2,000}{200 - m_d} = K_{AD} \tag{82}$$

Path 3

$$\frac{1,000}{100 - m_a} + \frac{4,500}{300 - m_f} = K_{AD} \tag{83}$$

Path 4

$$\frac{500}{50 - m_h} = K_{AD} \tag{84}$$

Path 5

$$\frac{1,000}{100 - m_a} = K_{AB} \tag{85}$$

Path 6

$$\frac{500}{50 - m_h} + \frac{5,000}{500 - m_g} = K_{AB} \tag{86}$$

Path 7

$$\frac{5,000}{500 - m_g} = K_{DB} \tag{87}$$

Path 8

$$\frac{1,000}{100 - m_a} + \frac{2,000}{500 - m_i} = K_{DB} \tag{88}$$

The network's constraining equations were determined during phase 5 and are given by Eqs. 18 through 28. The required network interzonal flows, given by Eqs. 2, 4, and 5, are substituted into these equations to complete the system of equations.

The solution of this system of equations yielded the path flow assignment given in Table 9. Because the values of the path cost constants for the loaded network are determined in the process of solving these equations, each path's total cost may be computed by multiplying its assigned flow by its corresponding path cost constant. These total path costs are also given in the table.

Also from Table 9, the path cost constants sometimes differ considerably for paths between the same zone pair. This can occur for either of two reasons. First, because only integral flow assignments are permissible, and because cost of a path can change by a considerable amount as saturation loading is approached, it follows that identical path costs are impossible. Second, the unit flow cost for some paths will be greater than the loaded cost of other paths, even when these other paths carry the entire interzonal flow. Because such paths will receive no assignment, their costs are unimportant.

TABLE 9

PATH FLOW ASSIGNMENT AND COST FOR SECOND ASSIGNMENT OPTION

Zone Pair	Path	Flow	Cost Per Unit Flow	Total Path Flow Cost
A-D	N_1	18	100.0	1,800
	N_2	3	100.1	300
	N_3	34	100.2	3,407
	N_4	45	100.0	4,500
A-B	N_5	25	83.3	2,083
	N_6	0	186.2	0
D-B	N_7	442	86.2	38,101
	N_8	8	87.4	699

TABLE 10

TOTAL INTERZONAL FLOW COST AND TOTAL NETWORK COST FOR SECOND ASSIGNMENT OPTION

Zone Pair	Cost
A-B	2,083
A-D	10,007
D-B	38,800
Total	50,890

Table 10 gives the total interzonal flow costs and the total network cost for the second assignment option. For this example assignment, the network assignment and total network costs were not substantially different for the two assignment options.

ACKNOWLEDGMENTS

Robert Brenner was of considerable assistance in formulating the techniques for applying the boundary conditions to the system of equations of the capacity-restraint algorithm. B. Bussell provided valuable information concerning the use and construction of connection matrices. Frank Haight was instrumental in furnishing insight into the applicability of various types of mathematical functions to the representation of transport flow costs.

REFERENCES

1. Bock, F. C., and Cameron, S. H., "Investigation of Methods for the Assignment of Trip Demand to a Road Network. Phase 2, A Family of Special Purpose Computing Machines for High Speed Solution of Assignment Problem." Armour Research Foundation of Illinois, Institute of Technology ARF Project E091 (Oct. 1957).
2. Bellman, R., "On a Routing Problem." Quart. Appl. Math., 16:87-90 (April 1958).
3. Calland, W. B., "Traffic Forecasting and Origin and Destination Trip Assignment Techniques." 11th Annual Convention, Western Section, Institute of Traffic Engineers, Proc., paper B (June 1958).
4. Carroll, J. D., Jr., "A Method of Traffic Assignment to an Urban Network." HRB Bull. 224, 64-71 (1959).
5. "Detroit Metropolitan Area Traffic Study. Part II, Future Traffic and a Long Range Expressway Plan, March 1956." Speaker - Hines and Thomas, Lansing, Mich. (1956).
6. Dantzig, G. B., "Discrete-Variable Extremum Problems." Operations Res. Soc. of America, Jour. 5:266-277 (1957).
7. Edwards and Kelcey. "Electronic Computer Program for Traffic Assignment (BPR Program No. T-4)." U. S. Bureau of Public Roads, Office of Operations, Division of Development.
8. Case, H. W., Brenner, R., and Campbell, B., "An Examination of Criteria for Judging Regional Master Plans." UCLA, Dept. of Eng., Special Report (1962).
9. Ford, L. E., Jr., and Fulkenson, D. R., "Maximal Flow Through a Network." Canad. Jour. Math., 8:399-404 (1956).
10. Fratar, T. J., "Forecasting Distribution of Interzonal Vehicular Trips by Approximations." HRB Proc., 33:376-384 (1954).
11. Fratar, T. J., "Vehicular Trip Distribution by Successive Approximations." Traffic Quart., 8:53-65 (Jan. 1954).

12. Garrison, W. L., and Marble, D. F., "Analysis of Highway Networks: A Linear Programming Formulation." HRB Proc., 37:1-17 (1958).
13. Hoffman, W., and Pavley, R., "Applications of Digital Computers to Problems in the Study of Vehicular Traffic." Western Joint Computer Conf., Proc., Los Angeles, pp. 159-161 (May 1958).
14. Irwin, N. A., Dodd, N., and von Cube, H. G., "Capacity Restraint in Assignment Programs." HRB Bull. 297, 109-127 (Oct. 1961).
15. Kalaba, R. E., and Juncosa, M. L., "Communications Networks: I. Optimal Design and Utilization." RAND Corporation, Res. Memo. RM-1687 (April 1956).
16. Lynch, J. T., Brokke, G. E., Voorhees, A. M., and Schneider, M., "Panel Discussion on Inter-Area Travel Formulas." HRB Bull. 253, 128-138 (Jan. 1960).
17. Mertz, W. L., "Review and Evaluation of Electronic Computer Traffic Assignment Programs." HRB Bull. 297, 94-105 (Oct. 1961).
18. Mertz, W. L., "The Use of Electronic Computers." Traffic Eng., 30:23-27+ (May 1960).
19. Mertz, W. L., "Traffic Assignment to Street and Freeway Systems." Traffic Eng., 29:27-33+ (July 1959).
20. Michle, W., "Link Length Minimization in Networks." Operations Res. Soc. of America, Jour., 6:234-243 (1958).
21. Minty, G. J., "A Comment on the Shortest-Route Problem." Operations Res. Soc. of America, Jour., 5:724 (1957).
22. Moore, E. F., "The Shortest Path Through a Maze." International Symposium on the Theory of Switching, Proc., Part II, pp. 285-292 (April 2-5, 1957). Annals, Computation Lab., Harvard Univ., Vol. 30 (1959).
23. Osofsky, S., "The Multiple Regression Method of Forecasting Traffic Volumes." Traffic Quart., 13:423-445 (July 1959).
24. Pollack, M., and Weibenson, W., "Solutions of the Shortest-Route Problem—A Review." Operations Res., 8:224-230 (March-April 1960).
25. Prager, W., "Problems of Traffic and Transportation." Proc. Symp. on Operations Res. in Business and Industry, Kansas City, Mo., pp. 105-113 (1954).
26. Prager, W., "On the Role of Congestion in Transportation Problems." Zeitschr. augewande Math. u. Mech., 35:264-268 (June-July 1955).
27. Rapaport, H., and Abramson, P., "An Analog Computer for Finding an Optimum Route Through a Communication Network." IRE Trans., Professional Group on Communications Systems, CS-7:37-42 (May 1959).
28. Shimbel, A., "Structure in Communication Nets." Proc. Symp. on Information Networks, Polytechnic Institute of Brooklyn, 3:199-203 (April 12-14, 1954).
29. Sokolnikoff, I. S., and Sokolnikoff, E. S., "Higher Mathematics for Engineers and Physicists." McGraw-Hill pp. 158-167 (1941).
30. Yoeli, M., "The Theory of Switching Nets." IRE Trans., Professional Group on Circuit Theory, CT-6:152-157 (May 1959).
31. "Traffic Assignment by Mechanical Methods." HRB Bull. 130 (Jan. 1956).
32. "Traffic Assignment." HRB Bull. 61 (Jan. 1952).
33. Reichenbach, H. G., "Routine TBU." University of California, Los Angeles, Institute of Transportation and Traffic Engineering, Report 62-21 (May 1960).

Appendix

GLOSSARY

Capacity Restraint. —The process whereby assigned volumes are related to the capacity of the highway facilities in such a manner that overloaded routes become less attractive as minimum path candidates.

Centroid. —The center of gravity of a zone, located at the mathematical center of the zone as determined by the cost functions of the zone.

Connection Matrix. —A matrix representation of all links interconnecting the nodes of a network.

Converted Flow. —A component of the normal flow pattern which has made a change in its usual mode of transport.

Diverted Flow. —A component of flow which has changed from its previous path of travel to another route without a change in origin, destination, or mode of transport.

Generated Flow. —Transport flow that exists because of a particular land use.

Generator. —An area that, due to its particular kind of land use, creates transport demand.

Induced Flow. —The added component of flow which did not previously exist in any form but which results when new or improved transport facilities are provided.

Land Use Feedback. —When the location, size, costs, etc., of a proposed transport network have been fixed by virtue of transport assignment and other considerations, it is recognized that the building of the proposed system will in itself alter land use patterns which in turn affect the forecasts of interzonal movements. This definition implies that the whole transport planning process is a continuing iterative system.

Link. —The one-way portion of the transport network connecting two nodes.

Link Flow. —The total interzonal flow assigned to a link in the network. These are sometimes referred to as leg volumes.

Link Impedance. —A value assigned to each link in the network. This impedance may be some average value of travel time, it may be distance if minimum distance routes are desired, it may be cost for use of the link, or it may be any other parameter or combination of parameters so desired.

Link Cost Function. —The mathematical relation describing the cost per unit of flow expended by using a given link.

Loading the Network. —The process of assigning the interzonal flows to the network.

Network Description. —The transport network under consideration described in tabular form as nodes, directional links, link impedances, link distances, turn restrictions, etc.

Node. —A point of intersection in a transport network.

Path. —The aggregate of all links in a route between any two nodes in a transport network.

Path Cost Function. —The mathematical relation describing the cost per unit of flow expended in using a given path.

Potential Flow. —The total flow that would in all probability move between two zones (on a given route), assuming ideal transmission facilities.

Routing or Trace. —A part of a tree. In the Moore algorithms, it is the minimum path through the network from one node to another.

Transit. —The movement of people on mass public media, such as buses, streetcars, and subways, and thus a subclassification of "transportation."

Transport. —Used in the most general sense to include the movement not only of people and goods, but also of other fluxes such as energy, information, water, and sewage.

Transportation. —The movement of people and goods, and therefore, a subclassification of the more general "transport."

Transport Network. —A network of links and nodes describing the possible flow paths for interzonal or intercentroidal movements for any single transport flux.

Tree. —The aggregate of all the minimum path routings from a node to all other nodes in the network.

Tree Building. —The use of an algorithm for computing minimum paths.

Zone. —An area of a system throughout which its describing parameters can be considered constant.