A Method of Traffic Assignment to an Urban Network

J. DOUGLAS CARROLL, JR., Study Director,
Chicago Area Transportation Study, Chicago, Illinois

● THE CHICAGO Area Transportation Study was given, as a prime objective, the task of developing a plan for new transportation facilities for the Chicago urban region. Since an unlimited number of plans or variations of plans is possible, the search clearly must be toward the one that is best or near-best.

Judgment of a plan involves many considerations but foremost is the measure of service the proposed new network of facilities will provide for future travelers. One of the tools for making such measures is traffic assignment. This involves the allocation of travelers moving between specified zones of an urban region onto particular travel routes. Having assigned every trip to a route, the level of service of the proposed network can be evaluated.

This allocation of trips is often a very complicated and time consuming task. In the Detroit Study it was possible to do this for a proposed expressway network in a period of from three to four weeks (1). This method gave useful information and did so within a reasonable time period, considering the magnitude of the task, but there were shortcomings. Only express routes received assigned volumes, whereas, in order to provide the greatest utility of the results, the loads on the surface arterial streets should have been known. Obviously, it would be questionable strategy to build an express route in a location where the surface streets could provide adequate service. A second shortcoming in the Detroit method was that it did not allow for assignment of travel to public transit routes.

In the planning of the Chicago Study, much effort was devoted to the development of a method which would rapidly assign travelers to the surface arterial streets as well as to express routes and also to rapid transit and surface transit routes. This was a particularly difficult problem because of sheer magnitude. The following measures will give some scale to the size of the problem:

(a) Six hundred thirty zones in a region of about 2,400 square miles and 396,000 interzonal traffic movements.

(b) A surface street network of 2,500 miles of route with 2,500 intersections.

(c) About 350 miles of proposed new express facilities.

(d) A rapid transit, suburban rail, and bus service network made up of nearly 2,500 route miles.

(e) A 1956 weekday total of over 6,500,000 vehicle trips and nearly 2,500,000 transit trips.

(f) If every possible route between every possible pair of zones were considered, the number of combinations would be greater than the number of atoms in the world.

The method of allocation had to be such that judgment was not involved in the process (that is, anyone using the same assumptions would obtain identical answers) and that the results would be unique for a specified network and a specified population of trips. Above all, it had to be fast for clearly it is desirable to be able to test many different possible plans.

The method selected, after much research, represented a consolidation of ideas from many sources. The beginnings came in a proposal made by Armour Research Foundation staff, working under contract to the State of Illinois on the development of network assignment methods. Their report pointed up the work of Moore (2) who had proposed a systematic and economical method for finding the shortest path through a maze. They also developed a computational program which would accomplish this work on a digital computer (3).

It was obvious that the computer program, as they developed it, could not handle a network of the size specified because the internal storage limits of any known machine were too small and the computational time would have been prohibitive.

Mr. Morton Schneider, Chief Computer Program Planner of the Chicago Area Transportation Study staff, developed an ingenious modification of this minimum path program to bring the traffic assignment within the range of computational feasibility. The elements of Mr. Schneider's method are outlined below, using the highway network as an example.[1]

First, he considers how to represent the street network to the computer. To do this, the key properties of the network were identified:

(a) The network of streets and highways consists of intersections and route sections between intersections.

(b) This network can be mapped so that each intersection is placed at a point in an oblong matrix of points measuring 55 x 95. This allows a maximum of 5,225 intersections.

The following examples illustrate the method of representing the actual street network. Figure 1, shows a portion of the existing street system from the basic network map.

In Figure 2, one can see the drafted form of the same map, converted to a form for

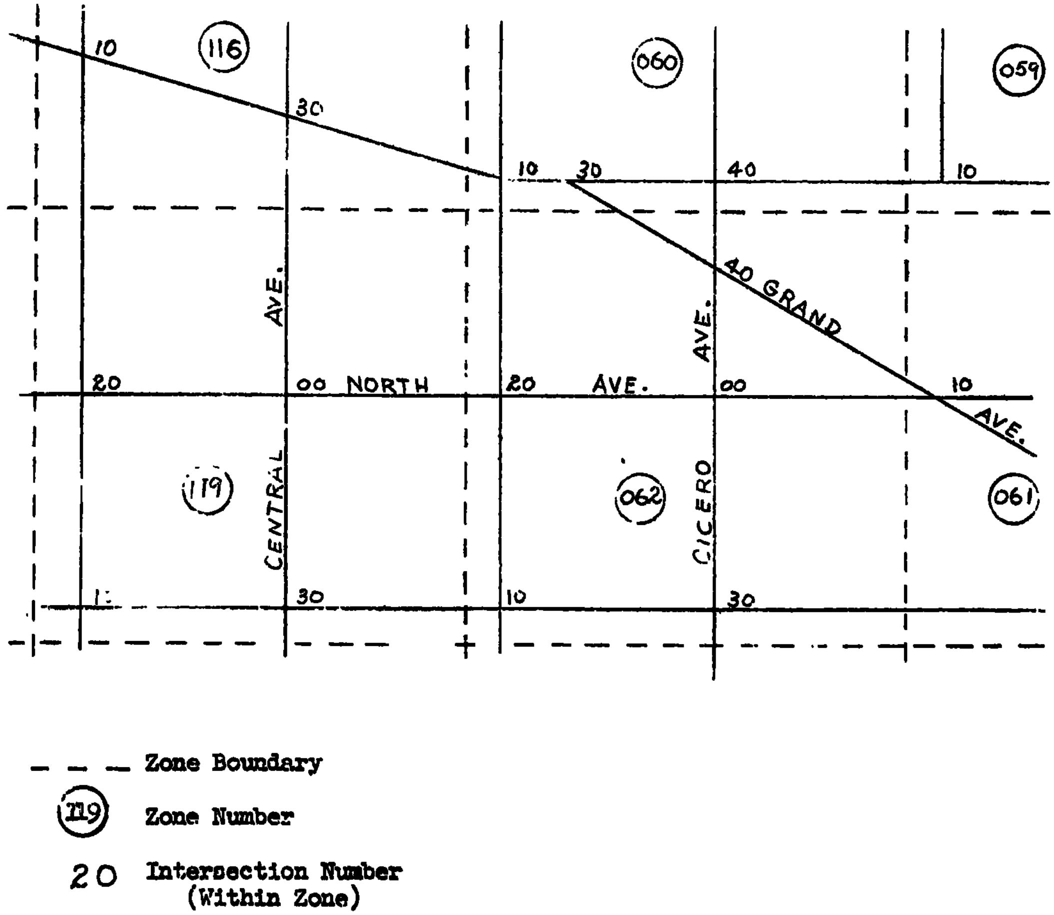


Figure 1. Section of Chicago area arterial street network.

---

[1] Credit should be given to CATS staff members, E. W. Campbell, E. L. Gardner, L. E. Keefer and P. J. Caswell, who helped to solve technical problems.

coding for representation in the memory of the computer.

Each intersection is given a number representing its location in machine memory and for each such location or intersection it is possible to code to four route sections. This is shown in Figure 3. In this fashion, all intersections have the possibility of interconnecting with their neighbors.

Figure 4 illustrates an eight legged intersection case and shows that with four route sections coded to each intersection, all possible connections in the network are allowed for, that is, any intersection may have as many as eight connecting route legs if it is possible to go from that intersection to each next adjacent one, although obviously the typical intersection will have only four connecting route sections. Now the simplicity of the system appears—the intersection requires no code number. It is simply at a location in the machine, that is, an address, so no use of numbers is required.

The same is true of a route section. At each memory location there is a word. Depending upon the computer used, this "word" consists of a number of possible digits. Assuming ten digits, the first two digits of each word may be used to represent the route

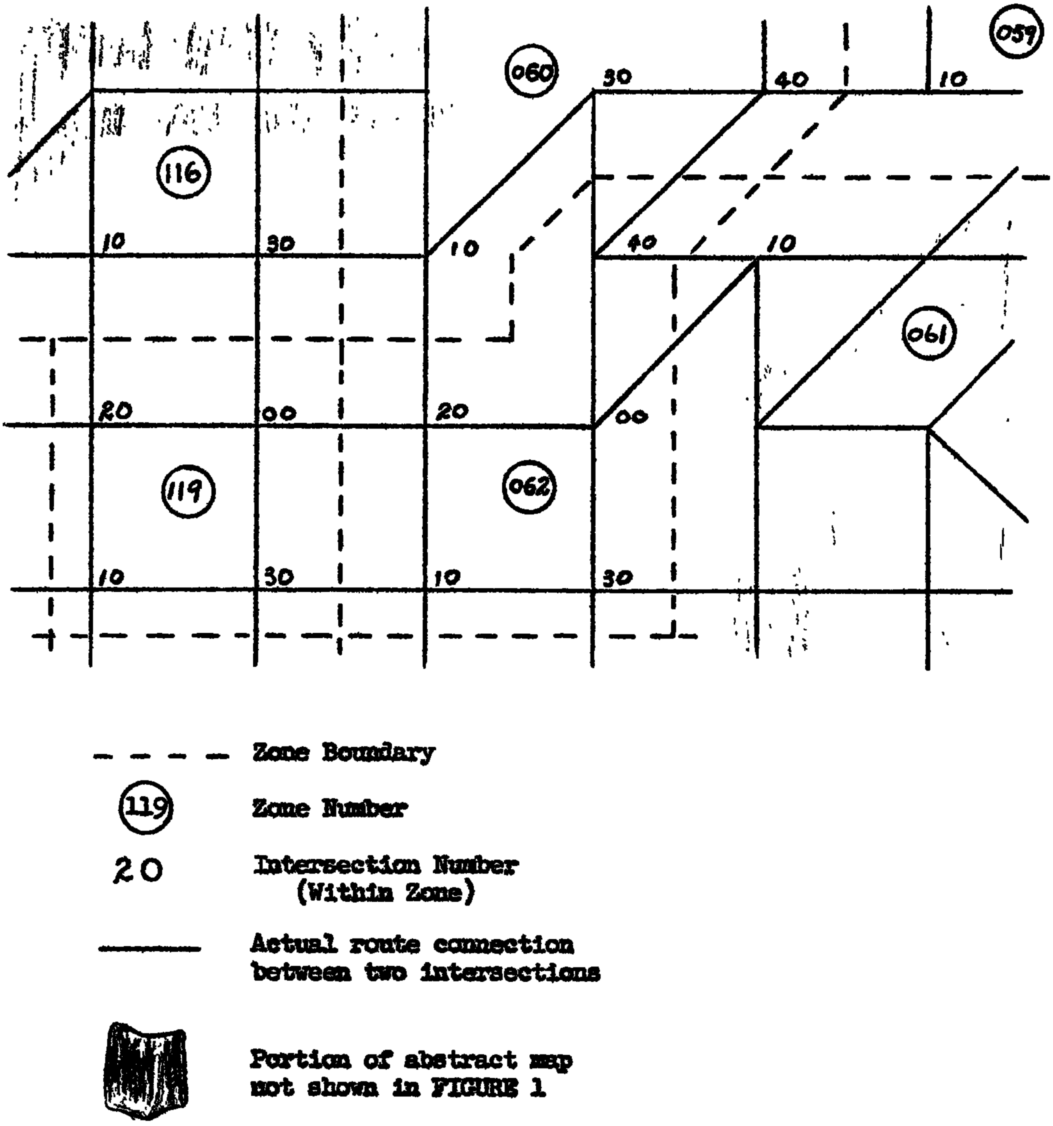


Figure 2. Abstract street network of Figure 1.

section in Direction 1, the next two a connection in Direction 2 and so on. In this fashion a space in a word, in memory, identifies a particular route section of the network.

In the appropriate section of each word, it is possible to code in the unit resistance which represents that piece of the network. In this case, one is concerned with minimizing journey time. (Clearly, cost, distance, or any other measurable factor could be used.) Thus time of travel to the nearest $\frac{1}{10}$ of a minute required to traverse each route section can be read from the coded map and stored in computer "memory." As an example, consider the true intersection, A, which appears in Figure 5. In the word, significant digits are had in the first two locations which indicate to the machine that there is a road leaving the "home" intersection and going to the one stored immediately above. And it shows that 0.4 minutes are required to traverse this route section and reach the next intersection. No road is indicated in Direction 2, a 1.5 minute route section lies in Direction 3 and none in Direction 4. Note that the other two connecting routes are referenced at other intersection addresses. The last two digits of the "word" may be used to mark which route section is part of the "minimum path tree" being computed.

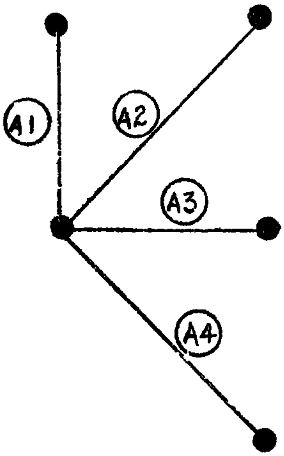


Figure 3. System of route leg identification.

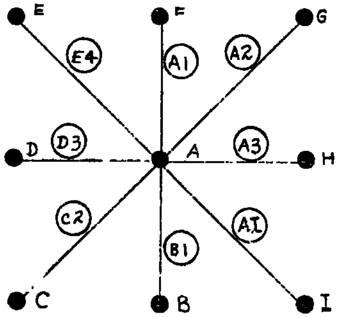
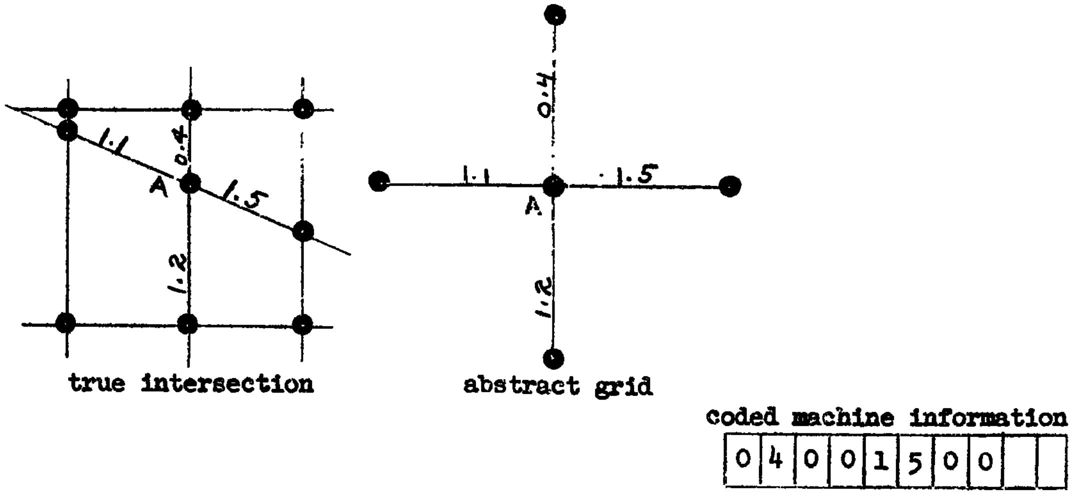


Figure 4. Diagrammatic representation of possible interconnection of one intersection with eight adjacent ones.

What the machine program does, then is to begin at an intersection which corresponds to a zone center. From this beginning point (Zone 1) the program, using a system of address modification, carries



true intersection

abstract grid

coded machine information

Figure 5. Example of network coding.

out a search of the network proceeding systematically outward from Zone 1 to every other intersection and identifies that route which requires the least travel time. One of the very great time saving devices is the economy of search as Mr. Schneider has designed it. This program insures that the shortest travel route in terms of the measure used—in this case travel time—is shown. Basically, the machine is programmed to search outward from an origin zone identifying minimum path routes successively until the most distant zones have been reached. This is known as building a "minimum path tree." The descriptive term "tree" is used because, starting from one center intersection of a zone, there is one and only one path from that beginning point to each next zone reached. As successively more distant intersections are found, more branches are developed, so that the entire trace of minimum paths identified would seem like a tree with the trunk at the origin zone and with ever increasing numbers of branches leading to the outermost zones.

After the minimum path from one origin zone to all others (that is, a "copy") is complete, the trips from the origin zone to all destination zones can be loaded onto the network routes that have been identified. Each zone thus is taken in sequence and the process is repeated until all zones have been treated and all trips have been loaded onto the network.

This is an over-simplified description of techniques. Detailed coding and machine procedures will be written up in a separate publication. This degree of detail has been used to illustrate the principal elements of coding and logic in using a digital computer for traffic assignment.

It is possible, under this system, to add new routes or to subtract others with a minimum of re-coding. Other features allow the storing of turning movements where these are of significance, or of restricting turns where this is a characteristic of the network (that is, limited access highways or toll roads).

All of the properties of the highway network can be applied to a transit network excepting that many bus routes run on the same street. It seems reasonable, however, to generalize bus service so that it is analogous to arterial streets and to code rapid transit and railroad routes like separate expressways. In this fashion, special assignments can be made to a large and complex network of transit routes.

Criticism of Method of Highway Assignment

A number of significant questions have been raised about this system of traffic assignment. The two principal ones are: (a) that this is simply the old "all or nothing" method which was displaced by the diversion curve and (b) that this will give highly artificial or unrealistic results.

Critics of this approach as being only the old "all or nothing" method point out that empirical research has shown that travelers between a pair of zones will split in their selections of route so that assigning an entire zone-to-zone movement to a single route is false.

There are a number of facets in the reply to this criticism. First, there is the simple matter of scale. In the Chicago area there are 630 zones and an average weekday travel volume of about 6,500,000 vehicle trips in 1956. Since there are 396,000 directional pairs of unique zones possible, the average movement will be about 16 trips and since sample data are used, this is less than one sample per zonal pair. It, therefore, seems ridiculous to split each transfer onto alternate routes.

Historically, the proportional allocation of zonal transfers came about because large zones—both geographically and in terms of trip volumes—have been used. Clearly, for example, if one is assigning a single driver to a network for one trip, he must take one, and only one, path. It can easily be shown that the more detailed geographic dispersion of zones will produce a substantial distribution of loads between the various competing routes.

On the other hand, tests have shown that some routes, because they do not connect directly to zone centers and/or because they have been given somewhat lower levels of speed than alternate streets, do actually come up with no trips allocated. This may indicate a more complex network than is justified in terms of the number of zones and more work is needed to evaluate the appropriate balance between network complexity and zone size. But it is also likely that streets with no assigned volumes represent a significant result and should be there to illustrate particular weaknesses in the network or in the value of speed given to those sections of route.

Those who have advocated diversion curves for use in traffic assignment have done so to meet the requirements of designers who ask the traffic engineer to tell them how many vehicles will use a ramp or a facility in the peak hour. The use of these empirical curves may be useful for such assignments to an already determined line.

But it is also possible that diversion curves are wrong. They have been established by making observation of an existing system's usage. Naturally, such systems tend to be in traffic equilibrium —that is, "traffic seeks its own level." It follows, therefore, that diversion rates are a function of the capacities and traffic pressures in the region being examined. If the expressway being measured for diversion had either fewer or more lanes, it seems quite clear that correspondingly greater or less traffic would use it and, therefore, diversion rates would change. For example, if the Shirley Highway (5, 6) had been built with six instead of four lanes, is it not possible that a different diversion rate would occur? The same comment would hold true for the competing surface route which is used to calibrate expressway diversion rates. Therefore, it is quite likely that route choices are very heavily influenced by the precise situation being observed. If this is so—and it must be true in some degree— then where is the value of using a diversion curve for assignment and which of many different curves should be used.

If it is argued simply, that travelers are non-rational in route choice and, therefore, it is better to use a diversion curve to estimate these non-rationalities, that again makes little sense. It suggests that there is a feeling that accuracy is increased simply by using a device to disperse groups of events about a mean or average answer. This certainly is not calculated to increase accuracy in such cases. Referral, incidentally, to most diversion curves shows that where journey times on two competing routes approach equality, the diversion rates approach 50 percent—that is, half of the trips use each facility. This appears to lend support to the "all or nothing" approach which merely says to use whichever route is superior. Final results from this "all or nothing" method would be very similar to those obtained by using a diversion curve which indicated a 50 percent split of traffic when the two competing routes used were equal in respect to the determining criterion.

In sum, there are quite apparent and inherent weaknesses in diversion curves as they are developed and used today. One of the greatest is that such curves provide an answer which is unrealistic by definition and yet is assumed to be correct. Therefore, the analyst does not truly know what he has obtained because the elaborate construction of the diversion curve merely produces an answer.

The second major criticism of this proposed method is that the results are "unrealistic." In other words, critics say that certain sections of the network will have no traffic while others can achieve loads well above their "capacity."

This raises the significant question of the true purpose of traffic assignment. To answer this criticism, it is necessary to distinguish between the simulation of traffic flow and traffic assignment.

A simulation of traffic flow would, if properly done, produce a result identical with the current usage of the urban network. In other words, the result of accurate flow simulation would be a current traffic flow map. Since drivers will seek to find the best travel path to their destination, congestion will be avoided and no single route in the urban system is likely to be severely congested while adjacent routes are free. There is very strong evidence that journey time is inversely related to rates of flow and travel times vary with changes in volume. Thus most urban systems appear to have a kind of equilibrium of flow as of any one time. It is likely that the destination and route of travel are both determined by the network and its level of usage.

To the planner, then, the extent of driver diversion is not visible on a traffic flow map and it is quite difficult to read the need for improved routes where the overloads are evenly distributed throughout the system. In short, needs cannot be determined by reading a traffic flow map. It is precisely for this reason that origin-destination data were collected.

It is possible to confuse a request for realism with traffic assignment as it can be of use to the planner. The planner needs to know where an improvement in capacity will do the most good. This method of traffic assignment by being "unrealistic" is able to magnify the points of great system stress and thus insure most judicious placement of improvements.

Also, when a plan is finalized, this method can be modified so that capacity restraints are introduced to the network and trips are diverted from congested to alternate routes, thus more realistically simulating predicted usage. In this fashion, capacities can be dealt with explicitly and the extend of diversion caused by capacity constraints can be measured. It is of substantial interest to note that this cannot be achieved by diversion curves. They are not sensitive to capacity constraints excepting those which were in effect when these curves were empirically established.

There are some other unique features possible with this system. Since the machine program computes the journey times from a single zone to every other one (that is, a "copy"), the elapsed travel times between that zone and the others are on hand in the computer memory, available for instant use. Knowing that trips are made from any zone to possible destination zones, in ascending amounts as the number of trip destinations at the receiving zone increases and in descending amounts as the time or distance of travel increases, (6) it is possible by stating such relationships mathematically to compute zonal interchange volumes as they are needed to allocate trips to the network. This eliminates the great problem of storing data for some 396,000 possible zonal pairs and greatly speeds up the entire process of assignment. This is especially useful for future traffic assignments because zonel interchanges are calculated as needed and reflect and space time system created by the new network. While much work can be done on the question of the appropriate method of computing zonal interchanges, this system has the advantage of allowing the changes in the network of streets to be used in calculating new travel interchange. Thus, wherever street improvements change the relative nearness of any pair of zones, it is possible to adjust travel frequencies in the direction indicated by present behavior patterns. This could be called "generated" traffic, although in this system it would be re-directed or diverted. It is believed that this is, in the long run, a more sensible way of computing future zonal interchanges than the current iterative methods in use.

Results of Tests Using This Method.

It was found that 6,500,000 auto trips could be assigned to the complete express and arterial network of the greater Chicago region in about 11 hours, using an IBM 704 computer with 32,000 words of core memory. The input to the machine consisted of 5,225 data cards to represent the existing network, 630 cards representing zone totals, and the program deck.

One copy (that is, the computation and identification of the shortest path from one zone to all others) requires about 30 seconds, the calculation of the interzonal transfers requires about four seconds, and the loading of the interzonal trips onto all route sections used requires from 10-80 seconds. Altogether, one complete copy which calculates and assigns all trips from one zone to all others is accomplished in an average of about one minute. By hand, this would require about five man-weeks, require extensive mapping and be subject to human error. Thus, a conservative cost for a man to do one copy by hand would be about $450. One copy on the machine costs about $10.

The output of the machine is a fully loaded network punched out on 5,225 cards or less. In addition, the total vehicle miles of travel are known and the total trips generated for each copy reported. All turning movements into, out of, and through the expressway network, are separately tallied. The entire assignment for the Chicago region was accomplished in one day. Future assignments can be completed in somewhat less time.

There are still many possible refinements in technique, but this particular system seems to go in the direction wanted—that is, rapid assignments at reasonable cost so that continuing tests can be made of route modifications.

One last point is of great significance. It has been customary to make traffic assignments to a single route. This is known to be inaccurate because the change of one route will create echoing changes in travel volumes throughout the entire network. It follows, then, that any traffic assignment to a single route would best be made if the entire network were treated and there were some adjustment in zonal interchange volumes. The speed and the methods outlined do go in this direction.

REFERENCES

1. Campbell, E. Wilson, "A Mechanical Method for Assigning Traffic to Expressways." HRB Bul. 130.
2. Moore, Edward F., "The Shortest Path Through a Maze." Paper presented at the International Symposium on the Theory of Switching, Harvard University, (1957).
3. "Algorithm for Finding Shortest Paths Between Pairs of Points in a Network." Armour Research Foundation Phase Report No. 1 to the Chicago Area Transportation Study, June (1957).
4. Trueblood, D. L., "Effect of Travel Time and Distance on Freeway Usage." HRB Bul. 61, January (1952).
5. Voorhees, Alan W., "A General Theory of Traffic Movement." Detroit Metropolitan Area Traffic Study, Michigan State Highway Department, Vol. 1, Chapter 5.
6. Ikle, F. C., "Sociological Relationships of Traffic to Population and Distance." Traffic Quarterly, Vol. 8, No. 2.