TECHNIQUES FOR PREDICTING FLOW IN A TRAFFIC NETWORK

L.R. FOULDS

UNIVERSITY OF CANTERBURY, CHRISTCHURCH NEW ZEALAND.

SUMMARY

This paper surveys techniques developed to date on the problem of predicting flow assignment in a traffic network. The paper examines the problem, presents a standard mathematical formulation, and considers extensions to capacity constraint, variable demand, and stochastic techniques. Some counter-intuitive phenomena in traffic assignment are documented.

1. INTRODUCTION

This paper surveys work done to date on the problem of predicting flow in a traffic network. This prediction process constitutes a vital intermediate step in the exercise of transport planning in any metropolitan area. It is used to help design traffic networks, assessing the impact of changes in networks and the performance of networks.

Let us consider a traffic network containing drivers who are each travelling from an origin to a destination. If each of their paths is assigned with the aim of minimizing congestion the resulting flows are termed system-optimal flows. Unfortunately, a traffic engineer cannot assign drivers to paths of his own choice. It is more realistic to assume that drivers choose paths they perceive to be least costly and that Wardrop's first principle [85] holds for each origin-destination pair p - q. This principle states:

The travel costs from p to q of any paths actually used are equal and no greater than the cost of any unused path.

The objective is to predict the user-optimal flows (also called equilibrium flows) in the network which will occur assuming these assumptions of driver behaviour. Of course, if the arc capacities are sufficiently generous and travel costs are constant per unit distance, then each driver will simply choose the least cost path from his origin to his destination. However, in practice each arc often has a tight bound on its capacity to accommodate flow. This means that travel cost in the arc is nonlinearly dependent upon the level of its flow. This causes congestion and makes it hard to predict the equilibrium flows. This is because the travel costs remain unknown until flow levels are estimated. Beckman et al. [3] have shown that the problem of finding the user flows is the following variational problem:
The Traffic Assignment Model (TAM)

We are dealing with a network $N=(V,A)$ with node set $V$ and arc set $A$. Let

- $A_{ij}(t) =$ the unit travel cost in arc $(i,j)$ when it contains $t$ units of flow,
- $x_{ij} =$ the level of flow in arc $(i,j)$,
- $x_{ij}^s =$ the number of units of flow in arc $(i,j)$ which have ultimate destination node $s$,
- $D(j,s) =$ the travel demand with origin $j$ and destination node $s$,
- $n =$ the number of nodes in $V$.

Then the problem is

$$\text{Minimize} \sum_{(i,j) \in A} \left[ \int_0^{x_{ij}} A_{ij}(t) \, dt \right] x_{ij}$$

Subject to

$$D(j,s) + \sum_{i=1}^n x_{ij} = \sum_{s=1}^n x_{ij}^s, \quad j=1,2,...,n,$$

$$\sum_{s=1}^n x_{ij}^s = \sum_{s=1}^n x_{ij}, \quad j=1,2,...,n, \quad s \neq j.$$

$$\sum_{s=1}^n x_{ij}^s = x_{ij}, \quad (i,j) \in A$$

Constraint (2) ensures that each driver begins at his origin $j$ and ends at his destination $s$. (3) is a definitional constraint which ensures that all drivers in arc $(i,j)$ have an ultimate destination $s$. (4) is the usual nonnegativity constraint.

A number of versions of the function $A_{ij}(x_{ij})$ are in existence. The following four from the National Co-operative Highway Research Program [60] have been adapted to conform with the present notation. Let

- $A_{ij}(0) =$ the free flow travel cost in arc $(i,j)$. For brevity this term will be denoted by $A_{ij}$,
- $u_{ij} =$ the capacity of feasible flow in arc $(i,j)$.

A cost function $A_{ij}(x_{ij})$ for any arc $(i,j)$ should have the following properties.

1) The cost should be nonnegative for all flows.
2) For flow levels between zero and the given capacity the cost should increase.
3) For flow levels increasing beyond $u_{ij}$ the cost should increase rapidly, i.e.,
1') \( A_{ij}(x_{ij}) \geq 0 \), \( x_{ij} \geq 0 \).

2') \( A_{ij}(x_{ij}) > A_{ij}(x_{ij}^2) \) if \( u_{ij} > x_{ij}^1 \geq x_{ij}^2 \).

3') \( A_{ij}(x_{ij}^1) > A_{ij}(x_{ij}^2) \) if \( x_{ij}^1 > x_{ij}^2 \geq u_{ij} \).

Smock: \( A_{ij}(x_{ij}) = A_{ij}\exp(x_{ij}/u_{ij}-1) \) (5)

Bureau of Public Roads: \( A_{ij}(x_{ij}) = A_{ij}[0.87 + 0.13(x_{ij}/u_{ij})^4] \) (6)

Bureau of Public Roads: \( A_{ij}(x_{ij}) = A_{ij}[1.00 + 0.15(x_{ij}/u_{ij})^4] \) (7)

Schneider: \( A_{ij}(x_{ij}) = A_{ij}[2(x_{ij}/u_{ij})^{-1}] \) (8)

Many authors now use a cost function based on the preceding ideas of the form

\[ A_{ij}(x_{ij}) = A_{ij} + b_{ij}(x_{ij}/u_{ij})^4, \]

where \( b_{ij} \) is a congestion constant for arc \((i,j)\). Branston [6] and Akcelik [1] have discussed various other cost functions. Boyce et al. [4] have studied the effect of different arc cost functions on equilibrium assignment for a large regional network. They compared the resulting equilibrium flows with each other and with observed traffic counts to evaluate the performance of the functions.

We now review the literature on the TAM. The reader should also refer to the relevant sections in Potts and Oliver [68], Steenbrink [75], to a previous survey by Nguyen [63], and to the state-of-the-art papers in [93].

2. TRAFFIC ASSIGNMENT TECHNIQUES

We begin with earlier models and techniques for them, which lead to the development of the TAM.

2.1 Least-Cost Route Assignment

Early work on this approach is surveyed in Martin et al. [51]. The reader should refer to the FHWA Traffic Assignment Manual [9] for a description of the methods mentioned. As network size increases many problems with the approach emerge. It is not clear that computer programs can deal with realistically-sized networks in a useful way. Stover [77] described a package, TEXAS-BIGSYS, which will handle up to 16,000 nodes and 64,000 links. However, gross simplifications to the network sometimes have to be made, and this approach has never been widely used. The simple all-or-nothing strategy is certainly the most popular and straightforward least-cost-route method. All flow from each origin to each destination is sent along the path of least cost joining the origin and the destination. Arc capacity is ignored, so that cost is assumed to be independent of flow. The network is loaded in this manner by using a tree-building algorithm. Loubal's method [50] makes it possible to consider only the effects of new or improved network links. The necessary shortest paths can be found by Dijkstra's
method [22], or more efficiently by Murchland's algorithm [59]. Unfortunately, since travel costs are assumed constant, the strategy usually fails to reflect driver behaviour accurately. Further, the strategy is unstable in the sense that an apparently insignificant change in data input can lead to a significantly different output.

2.2 Incremental Capacity Restraint Assignment

These models assume that travel costs depend upon flow levels. The network is loaded with multi-path flow in stages. At each stage, the new assignment is based on updated travel costs which are calculated on the basis of the flow accumulated in each arc so far. Van Vliet [83] reported that the final flows are not greatly affected by how much flow is loaded at each iteration. The strategy was first used by the Chicago Area Transport Study [12] and variations have been presented by Irwin et al. [42], The Bureau of Public Roads [9], Homburger [39], Smock [72], Schneider [71], Steel [74], Almond [2], Wilson et al. [90], and Martin and Manheim [52].

The concept of diversion assignment, as presented by Irwin and Von Cube [41], McLoughlin [54], and Overgaard [65], represents a slight improvement over the simple all-or-nothing assignment. The U.S. Bureau of Public Roads Traffic Assignment computer package [9] employs diversion assignment and allows for turn penalties and prohibitions. It allows for a second route to be chosen and is thus more useful than all-or-nothing. Michaels [55] reports on a study of the diversion of traffic on an arterial road to a parallel toll-road. Van Falkenhausen [84] and Burrell [10] were among the first to make the even more realistic assumption that multiple routes between each origin and destination pair may be used. It appears that their assignment procedures are significantly superior to the simple all-or-nothing strategy.

The most important capacity restraint method is the FHWA procedure. It is multipath if the results of successive assignments are averaged or otherwise combined. If they are combined in one of a number of ways, convergence to equilibrium can be demonstrated.

There have been a number of papers reporting efforts to ascertain the accuracy of these methods. Tagliacozzo and Pirzio [78] discovered that users are not, in general, sufficiently acquainted with the topography of the urban network and hence not capable of evaluating the parameters involved in the path selection. Judge [44] concludes that the accuracy of path choice predictions is a sufficient condition for assignment accuracy. Moreover, this condition subsumes the condition of accurate arc loadings since an accurate route choice prediction will of itself produce accurate arc loadings.

Ferland et al. [24] have come to the conclusion that incremental capacity-restraint methods do not satisfy Wardrop's first principle, i.e., not produce user-optimal flows; but it is intuitively clear that the smaller the increment the closer to equilibrium will the resulting flow pattern be. This supports the
results of experiments by Nestle [61]. However this is at variance with Van Vliet's [83] experimental evidence. Far more serious is the deficiency that once an increment of flow is assigned, it cannot be reassigned to another path. Ferland et al. go on to display some discrepancies that may result due to this characteristic of the method. It is clear that the final flow pattern obtained from use of incremental capacity restraint methods is quite sensitive to the manner in which the network is loaded.

2.3 Mathematical Programming Methods

There have been a number of attempts to predict traffic flow with mathematical programming. Charnes and Cooper [11] formulated a multicycle traffic assignment model which turns out to be a linear programming problem. Pinnell and Satterly [67] presented a detailed example of multicycle assignment on a network with 25 nodes, 176 arcs, 12 origins and 3 destinations. This example was also solved using a discrete version of Pontryagin's maximum principle (see Hwang and Fan [40]), by Yang and Snell [92] and Snell et al. [73]. Tomlin [80] developed a network programming model which can be solved by a multicommodity flow algorithm. The approach is able to overcome problems in efficiency by using column generation. (See Leventhal et al. [49] for another column generation method.) Even so, despite the efforts of others, including Gibert [35], the method seems to be practical only for small networks. Tillman et al. [79] proposed a dynamic programming algorithm to solve the multicycle problem which does not require the direction of flow to be fixed before each copy. A survey of the state of the art in 1967 appears in Jewell [43].

2.4 Equilibrium Methods

The third generation methods have proven themselves far more efficient and to produce better assignments than the earlier methods. They use various aspects of mathematical programming when the relationship between system-optimal and user-optimal solutions can be exploited. They sometimes attempt to simulate what drivers actually do in practice in trying to find satisfactory paths. Mosher [56] was one of the first to formulate the problem of finding user-optimal flows explicitly. Dafermos and Sparrow [14] and Dafermos [13], [15] generalized Mosher's work to include the cases where travel costs in an arc depends upon the entire flow pattern (not just flow in the arc) and where there are several classes of users. Bruynooghe et al. [7] have also presented algorithms. Their main contribution is that, with their procedures, there is no need to calculate shortest paths initially. They are found as needed. (See Ruiter [70] for further details.)

Based on Murchland's work [58], Ruiter [70] has described a general framework for equilibrium algorithms for user-optimal solutions guaranteeing convergence. Further breakthroughs have been achieved in the methods of Nguyen [64], Le Blanc et al. [48], Evans [23] and others.

The method named TRAFIC was developed by Nguyen and James and uses an adaptation of the Frank-Wolfe method. Specifically, an initial solution is developed (zero flow if nothing more suitable is
available). Then a linear approximation to (1) is made in order to find a best direction in which to proceed to a new solution. Dijkstra's [22] method is used to calculate shortest paths at this stage. Golden Section search is used to determine the step size taken along the direction chosen. A heap storage system is used for node and arc indexing and retrieval. Function (7) was used to define arc costs. The algorithms were programmed in FORTRAN on a CDC CYBER 74 and tested on a network with 376 arcs, 155 nodes, 27 zones, and 690 origin-destination pairs with encouraging results.

Le Blanc et al. [48] also developed a similar algorithm for the TAM. They show that in using their technique none of the constraints need be considered explicitly - they are satisfied automatically. The Frank-Wolfe algorithm [33] is used iteratively to calculate feasible directions. The method was coded in FORTRAN on a CDC 6400 on a network with 24 nodes and 76 arcs with promising computational time.

Wigan and Luk [87] compared the three assignment techniques: all-or-nothing, incremental loading, and equilibrium assignment. It is reported that the first two techniques did not produce useful solutions and were insensitive to changes in trip demands, unlike the equilibrium technique TRAFFIC. A perceptive analysis of techniques and computer packages available until 1977 is summarized by Wigan [88].

The methods mentioned so far all require that the origins and destinations of the trips be represented as points or centroids. This misrepresents actual origins if they are located in the middle of arcs. Daganzo [18] attempted to overcome this problem by introducing a procedure which is designed to deal with substantially larger numbers of centroids to the stage of handling a continuous distribution of population. See also Buckley [8] for an analysis of traffic assignment in a two-dimensional continuous representation. There has been some recent interest in incorporating explicit bounds on arc capacity. That is, as well as introducing the concept indirectly as a penalty in an objective function such as (5)-(8), bounds of the form

$$ x_{ij} \leq u_{ij} \quad (i,j) \in \Lambda $$

are added to (1)-(4). Wardrop's principle needs to be modified in this case. The addition of (10) raises both theoretical and practical considerations, some of which have been examined by Hearn [36], Payne and Thompson [66], and Hearn and Ribera [37], [38]. Daganzo [16], [17] attempted to show how the algorithms of Le Blanc et al. [48] and of Nguyen [64] can be extended to include (10). Unfortunately, in order to guarantee convergence it was necessary to assume that

$$ \lim_{t \to \infty} A_{ij}(t) = \omega, \quad (i,j) \in \Lambda $$

Hearn and Ribera [38] prove a theorem which establishes convergence under much weaker conditions.
For a study of analytic methods on networks of very simple geometry see Lam and Newell [47].

2.5 Varying Demands

This section is a summary of the material in [70]. Since 1967 a number of algorithms have appeared which obtain flow assignments when travel demand varies with cost. The first such algorithm with guaranteed convergence was presented by Gibert [35]. Murchland [58] refined this work to take advantage of the relationships between system-optimal and user-optimal solutions. He stated broad guidelines which he felt should be taken into consideration in algorithm development and used them himself in developing a computer program for producing equilibrium assignment. Wilkie and Stefanek [89] use control theory to solve their nonlinear model, but Kulash [46] has only a linear model. Netter and Sender [62] have an algorithm with proven convergence which allows for multiple user groups and multiple-dimensional variable demands. They show that multiple solutions may exist, the algorithm converging to one of them. Which one depends upon the chosen starting point.

Yagar [91] has extended Homburger's capacity-restraint algorithm [39] to allow for time-varying demands and queued-up demands. This is achieved by breaking up the time horizon of the scenario into sufficiently small segments so that each segment can be assumed to have non-varying demands. The method was tested on an existing roadway corridor in the San Francisco Bay area.

Merchant and Nemhauser [53] have presented a discrete time model for TAM with varying demand. Congestion effects appear in the flow equations of their nonlinear, nonconvex model. They make a piecewise linear approximation which can be solved optimally by a one-pass simplex algorithm. Due to its special structure it can be solved by decomposition techniques or compactification methods for sparse matrices. Gartner [34] has surveyed methods for assignment under varying demand through to 1980.

3. STOCHASTIC MODELS

It seems that the flow assignments produced by the methods mentioned so far do not always accurately reflect actual driver behaviour. The actual arc volumes may be so far astray as to compromise the transport designer's decisions. It would seem more sensible to model the somewhat arbitrary and suboptimal habits of drivers. That is, apportion trips to paths in a stochastic manner according to their probability of actually being used. Unfortunately, there are just too many possible paths to make the computational effort worthwhile. Dial [21] has presented a stochastic multipath method which attempts to circumvent the problem of the explosive number of paths. It assigns trips to the reasonable paths simultaneously in an efficient manner. Dial calls it a two-pass Markov model. It finds node/arc transition probabilities in the first pass and assigns trips in the second, when it diverts flow entering a node to all reasonable arcs ending at the node. In this manner it assigns trips simultaneously to an entire set of reasonable paths.
The method has been extended by Florian [28] to the case of dynamic assignment of trips, including the determination of the distribution of trips through the network nodes. The extensions attempt to overcome the undesirable features of sometimes unreasonable flows produced by Dial's method. However, Florian and Fox [29] have concluded that the shortcomings of Dial's method are extremely difficult to overcome.

Other early methods which do not necessarily allocate all flow to least cost routes were presented by Burrell [10] and Von Falkenhausen [84]. Daganzo and Sheffi [19] critically appraise the methods just mentioned and propose an alternative formulation. They enunciate a modification of Wardrop's user-equilibrium principle, namely:

No user believes he can improve his travel time by unilaterally changing routes.

A stochastic model with constant link costs is then analysed in detail and an expression for the probability of route choice based on the following two postulates:

1) Nonoverlapping sections of road are perceived independently by the tripmaker.

2) Sections of the road of equal travel cost are perceived in identical fashion.

Existing stochastic methods are compared to this suggested approach. The paper also discusses two proposed techniques which can be used to estimate flow assignment, based on the model in large networks. Numerous extensions of this work have been reported by these authors.

Another stochastic technique, which complements the just-mentioned method, was presented by Dial [20]. It provides an idea for a practical approach for the determination of the relatively few paths that a driver will seriously consider.

The algorithm of Robillard [69] combines the concepts of varying demand and stochastic techniques. Dial's probabilistic multipath method [21] is extended to the case of dynamic demand. Robillard's algorithm uses an adapted version of Dial's method repeatedly along with a fast Fourier transform procedure. His model does not allow for explicit arc capacity constraints. Trahan [81] has described improvements to Dial's method and illustrated the advantages of Robillard's refinements over the original Dial approach.

Finally, Fisk [26] has formulated a network optimization problem which yields a stochastic equilibrium traffic assignment model incorporating congestion effects. She also demonstrates the similarity between some fixed demand incremental methods and TAM.
4. COUNTER-INTUITIVE RESULTS IN TRAFFIC FLOW PREDICTION

When a network is uncongested, travel costs are relatively insensitive to changes in flow levels. In this case the addition of a new arc will not increase the individual driver's cost and hence will not increase total user cost. Intuitively it seems likely that this will also be true when congestion is present and travel costs are very sensitive to flow levels. Such is not the case. Braess [5] constructed a small numerical example in which the addition of a new arc significantly increased all drivers' costs and therefore increased congestion. This result was made more accessible by Murchland [57] who dubbed it Braess's paradox and commented on its implications. A slightly more elaborate example is given by Stewart [76]. All authors are in agreement that this result is not really paradoxical in the sense that there is no scientific reason why arc addition should lessen congestion, given the assumption of user optimal flows. (Of course under the assumption of system optimal flows, an arc addition would never increase congestion.) This result is further discussed by Florian [28]. Other counter-intuitive results appeared recently. Fisk [25] showed that congestion may increase when some flow is removed from the network.

The papers referred to so far have considered only small, contrived examples. However, Knodel [45] reported that Braess's paradox may have occurred in actual practice in the city of Stuttgart. After network improvements failed to achieve the desired results, an arc (representing the lower part of Königstrasse) was eliminated. This resulted in a reduction in congestion. A result which is similar in spirit has been reported by Fisk and Pallottino [27] who found that, using city of Winnipeg network data, total travel cost in a user-optimal assignment may decrease when some travel demands are increased.

5. SUMMARY AND CONCLUSIONS

This paper has introduced some of the models and techniques for predicting traffic flow in an urban network. It began with a standard formulation of the problem and a description of the various methods in the three generations of approaches to traffic flow prediction: least cost-route assignment, capacity restraint, and equilibrium assignment. Next, the extension to explicit arc capacity constraints and varying demands were considered. A discussion of stochastic techniques and counter-intuitive results completed the paper.

Of the three approaches for the deterministic traffic assignment model with fixed demand: least cost-route assignment, incremental loading/capacity restraint, and equilibrium methods, only the last two approaches appear to truly satisfy Wardrop's first principle and produce flows that approximate reality to any significant degree. Even these methods are cumbersome to use on realistic networks (often due to the slow convergence of the Frank-Wolfe algorithm), and do not allow for the arbitrary and suboptimal behaviour of drivers.
It appears that the probabilistic multipath approaches sometimes overcome the criticism of assuming drivers are infallible. Yet none of these methods have been shown to be entirely satisfactory. Thoroughly documented experience with the methods along the lines of Florian and Nguyen [30], Overgaard [65], and Whiteford [86] with equilibrium assignment would be a most useful exercise.

REFERENCES

175


[Fox, B. Probabilistic Networks, Publication 33, Centre de Recherche sur les transport, Université de Montréal, Montreal, 1975.]


[71] Schneider, M. A Direct Approach to Traffic Assignment. Highway Research Record No. 6 (1963) 71-75.


The author would like to express his sincere thanks to an anonymous referee who made valuable suggestions leading to an improved paper.