HYBRID ROUTE CHOICE PROCEDURES IN A TRANSPORT NETWORK CONTEXT

Glen D'Este
Senior Research Fellow
Transport Systems Centre
University of South Australia
Adelaide SA 5000
Australia
Fax: +61-8-8302-1880
E-mail: glen.deste@unisa.edu.au

abstract: Assignment methods for transport network models are typically based on finding preferred - usually shortest - paths on the basis of a single measure of the utility, such as generalised cost. However studies of the behaviour of transport users have shown that the decision process involves multiple criteria and is a hybrid of compensatory and non-compensatory stages with both optimising and satisficing behaviour. This paper reviews and extends the state of the art in network techniques and algorithms that can be used to implement a hybrid decision process.

1. INTRODUCTION

Traditional assignment methods for transport network models are based on finding shortest paths on the basis of a single measure of the utility of alternative paths through the network. This measure of utility may be the travel time, distance, or monetary cost, or a generalised cost based on a linear combination of several factors. This is still the most common approach despite the fact that many studies of mode and route choice behaviour have found that the decision process is much more subtle and complex than simply identifying the shortest path.

The prevalence of shortest path and related assignment models has arisen for a combination of reasons including computational efficiency and the dominance of modelling of private motor vehicle traffic in urban areas. In the past, the need to handle large networks in terms of numbers of links and nodes using limited computing power by present standards necessitated the use of assignment algorithms that are highly computationally efficient. As a result, it could be argued that the richness of the route choice process has been sacrificed in favour of simplicity and speed of computation. Conversely, it could be argued that it is reasonable to use a shortest path approach based on a single measure of utility for urban traffic assignment because the relevant characteristics, typically distance, time and cost, are essentially surrogates for each other. As the length of a path increases so does its duration and cost, and *vice versa*. Therefore minimising one characteristic minimises all and in practical terms, there is no need for more complex decision rules. However there are other transport applications that involve genuine trade-offs between route attributes, in particular, situations with charging for transport services, such as toll roads or long distance freight transport. In these situations, a faster route corresponds to a premium service and attracts a higher price.

The potential complexity of route choice behaviour in transport networks is well illustrated by the choice of carriers, modes and intermodal routes for long distance freight transport and international trade. Surveys of purchasers of international freight transport services (D'Este and Meyrick, 1992) have led to the development of conceptual models (Brooks, 1990; D'Este, 1992) which represent the decision process in terms of three basic steps:

STEP 1. Eliminate options that are infeasible for technological or other reasons
STEP 2. Eliminate the options that fail to meet prescribed performance criteria, such as maximum acceptable cost, maximum acceptable transit time, minimum levels of reliability or service quality
STEP 3. Select a transport carrier from the options that remain on the basis of a trade-off between relevant service characteristics

The number of options that are considered in the final trade-off (Step 3) has been found to be very small (Saleh and Lalonde, 1972; D'Este and Meyrick, 1992) - almost always less than ten and typically less than five.

It follows that in general, the route choice process is sequential and a hybrid of compensatory and non-compensatory stages with both optimising and satisficing behaviour. Satisficing is the process of establishing a standard of satisfactory performance and then accepting only those options that meets the standard. For example, a decision maker may prefer the cheapest transport option but only if that option ensures arrival within a prescribed time. All options that meet the travel time criterion are included in the choice set (others are eliminated), then the lowest cost option is selected from the choice set. Note that a low cost will not compensate for an unacceptably long travel time.

This decision mechanism includes the simple shortest path assignment as a special case but in general is much more complex and richer in its use of information. In particular, it does not engender two assumptions implicit in the shortest path approach with linear generalised cost function. The first is simultaneous comparison of all options and attributes. The choice set implicit in traditional shortest path assignment is the set of all possible paths through the network. This is a computationally convenient assumption but studies of human information processing (Newell and Simon, 1972) have found that sequential consideration of attributes is more common and that the final selection will usually involve a small number of options. This behaviour has been verified in the transport context by D'Este and Meyrick (1992). The second assumption implicit in the use of linear generalised cost functions is that attributes affecting route choice are compensatory, that is, exception performance in one attribute will compensate for poor performance in another. Sometimes this is the case but in many instances, attributes have absolute bounds of acceptable performance, for example, a maximum acceptable cost, or a maximum acceptable transit time, or minimum levels of reliability or quality of service.

With significant advances in computing speed and growing interest in modelling more complex transport decision situations, it is timely to reconsider options for implementing hybrid choice mechanisms in a transport network context. This paper reviews the state of the art in network techniques and algorithms that can be used to implement the sort of hybrid decision process described above. Further, the paper develops several procedures that apply network algorithms to eliminate unacceptable options then identify a preferred path that meets all relevant decision criteria. The significance of this paper is that it provides an algorithmic basis for implementing route choice techniques that are much more realistic and faithful to the process actually undertaken by decision makers.

2. THE PROBLEM

In a nutshell, the problem is to develop an efficient technique to implement a hybrid choice process across a transport network. There are several available techniques for modelling hybrid decision processes but these techniques assume that the choice set is small and fully enumerated. That is, it is assumed that all available options are known and fully defined, and that the choice set is an input to the decision model. Examples include the elimination by aspects model developed by Tversky (1972) and a similar sequential non-compensatory approach proposed by Recker and Golob (1979). These models assume that the choice set is given and that options are differentiated on the basis of several attributes associated with each option. Further it is assumed that these attributes can be ranked in importance and each attribute has a range of acceptable values. The choice mechanism commences by considering the most important attribute. Options that fail to meet the acceptance criterion are eliminated from the choice set then the process is repeated for the reduced choice set and the next most important attribute, and so on. The process continues until a single option remains or a small choice set of acceptable options is obtained.

In the case of transport networks, route options corresponds to paths and full enumeration of paths is impractical in all but the smallest and simplest networks. For even a modest sized network, there can be a very large number of available paths. The aim of the hybrid decision process is to select a preferred path (or paths) on the basis of one measure of utility while observing constraints on one or more other measures of utility. The closest network analysis equivalent is to find the shortest path under multiple side constraints. Algorithms exist for this problem, see Aneja et al (1983) and Jaffe (1984) but the shortest path problem with side constraints has been shown to be NP-complete (Garey and Johnson, 1979). Also, the problem is not amenable to solution using standard linear programming techniques. Therefore alternative approaches are preferable, both for computational efficiency and to incorporate more general hybrid choice processes. The problem addressed by this paper is flexible and efficient network algorithms for hybrid route choice mechanisms.

In more formal terms, the problem can be stated as follows. Consider a network $G = (V,E)$ where V is a set of n nodes and E is a set of m directed links and r is the maximum degree of nodes in G. Associated with each link are several non-negative attributes that measure the utility of the link according to different characteristics, such as distance, time, cost, reliability, or generalised cost. Let these attributes be denoted l and w_1, w_2,... Each attribute corresponds to a particular measure of the utility of a link as part of a path.

The problem considered by this paper is to select a preferred path (or paths) from a particular origin node s to a particular destination node t on the basis of attribute l while observing constraints on attributes w_i. Without loss of generality, all constraints can be considered to be "less than" constraints, and for convenience, attribute l will be referred to as the *optimising* attribute, and the attributes w_i subject to constraints will be referred to as the *satisficing* attributes. It will also be assumed that the satisficing attributes can be ranked in importance, with w_1 being the most important, and so on.

The preferred path may be the shortest path that meets the constraints or several paths may be used with probability that depends on the relative utility of each path according to a logit or similar choice process, for example Dial's (1971) multipath algorithm. These options for selecting a preferred path correspond to all-or-nothing and stochastic assignment techniques of transport network modelling, respectively.

Let $L(p;s,t,l)$ denote the length of path p from s to t calculated using attribute l, then in the case of an all-or-nothing assignment between the prescribed origin and destination, the problem can be expressed as

> Given a network $G(V,E)$ with attributes l, w_i on E and nodes s, t in V ; find a path p from s to t that minimises $L(p;s,t,l)$ subject to $L(p;s,t,w_i) \leq C_i \;\; \forall i$.

The range of hybrid choice mechanisms considered by this paper is broader than this definition but the definition provides a starting point from which the discussion will progress.

3. SHORTEST PATH ALGORITHMS

It has been argued that the shortest path method is an over-simplification of route choice procedures in transport networks. However shortest path algorithms provide the computational basis for most of the hybrid route choice models considered in this paper. Therefore it is worthwhile reviewing some aspects of the theory and application of shortest path techniques.

Firstly some background on shortest path algorithms. The algorithms are based on Bellman's (1958) equations applied using either node labelling (Dijkstra 1959) or matrix algebra (Floyd 1962). For a review and comparison of shortest path algorithms, see Gallo and Pallatino (1984). Node labelling shortest path algorithms find shortest paths from a single origin and have time complexity that is typically in the range $O(n^2)$ to $O(m \log n)$. According to Gallo and Pallatino (1984), the node labelling algorithm that provides best all round performance when applied to transport networks is the D'Esopo-Pape algorithm (Pape 1980). Floyd's algorithm has complexity $O(n^3)$ but it returns shortest paths between all pairs of nodes.

Two aspects of shortest path algorithms are directly relevant to the discussion of hybrid route choice mechanisms. These aspects are node labelling and ways in which the basic shortest path algorithm can be extended to a wider range of operations on link weights. As the name suggests, node labelling algorithms calculate a label for each node. On completion of the algorithm, the label λ_{sv} on any node v is the utility of the shortest path from the starting node s to node v on the basis of a particular link attribute. The algorithms can also be run backwards from the destination node t in which case the label λ_{vt} is the cost of the shortest path from node v to node t. The sum of the labels at node v from the forward and backward applications of the shortest path algorithm $(\lambda_{sv} + \lambda_{vt})$ is the utility of the shortest path from s to t through node v. This observation has direct implications for applying constraints to satisficing attributes since it sets a lower bound on the utility of all possible routes through a given node for a particular attribute.

Alternatively, the final matrix obtained using Floyd's algorithm can be used to calculate the shortest path from s to t through any given node. If Θ is the final matrix, then $\theta_{sv} + \theta_{vt}$ is the length of the shortest path from s to t through node v.

The other important property of shortest path algorithms is their ability to be extended to a much more general class of problems. The core of both node labelling and matrix algorithms is repeated application of a triple relation of the form

$$c_{ij} = \min[c_{ij}\,,\,(c_{ik} + c_{kj})]\,.$$

This triple relation is an example of the more general triple operation

$$z_{ij} = \mathrm{OPT}[z_{ij}\,,\,(z_{ik} \otimes z_{kj})]$$

where z is the path attribute to be optimised and $\otimes$ is a general operation. As long as the operation $\otimes$ satisfies the condition

$$z_{ij} = z_{ik} \otimes z_{kj}$$

then the algorithms provide a valid solution to the path finding problem implied by the triple operation. The significance of this observation is that it allows the shortest path techniques to be applied to a wider range of constraint types. For example, the shortest-path algorithm is equally valid for addition and multiplication, and for objectives such as reliability and capacity (Christofides, 1975).

It follows that shortest path techniques can be used to identify nodes that cannot support paths that meet specific constraints, and can be extended to incorporate a wide range of different constraint types.

4. ALGORITHMS FOR HYBRID ROUTE CHOICE MECHANISMS

We are now in a position to consider alternative approaches for addressing the hybrid route choice problem defined above. This paper considers three approaches; the first is an all-or-nothing assignment that involves screening shortest paths for compliance with the constraints; the second uses the constraints to create choice sets of acceptable options then applies the assignment process to a sub-network; and the third embeds the elimination of unacceptable options directly into the assignment algorithm.

4.1 Screening for Acceptable Paths

The first approach is a direct extension of the standard shortest path problem for all-or-nothing assignment. The aim of the hybrid route choice procedure is to select the shortest path on the basis of one measure of utility while observing constraints on one or more other measures of utility. The closest network analysis equivalent is the problem of finding the shortest path under multiple side constraints but as noted above, the problem has been shown to be NP-complete and is not amenable to solution using standard linear programming techniques.

Christofides (1975) has suggested that the best approach is to calculate the k shortest paths and then to enumerate and screen the paths. The process involves generating a choice set comprising the k shortest paths on the basis of the optimising attribute. Each of the k paths in the choice set is then screened for compliance with the constraints, starting with the 'best' path, then the next best and so on. The screening process terminates when first path is encountered that meets the constraints; this path becomes the preferred route.

The first step required to implement this approach is to identify the k shortest paths. There are two different types of kth shortest path problems; one allows paths to have cycles, and the other does not. In transport network problems, it is customary to restrict consideration to simple paths, that is, paths that do not return to the same node. There are several efficient algorithms available for identifying the k shortest simple paths. The best known algorithm, Yen (1971), has time complexity of $O(kn^3)$ but more recently Katoh et al (1982) have published an algorithm with running time $O(kn^2)$.

However it is possible to relax the requirement for simple paths and still construct paths that are acceptable in a transport network context. Fox (1979) has developed a fast algorithm for constructing the k shortest paths with the possibility of loops; the time complexity of the algorithm is $O(rn + kn \log r)$ plus one application of a node labelling shortest path algorithm. More importantly, Fox has also shown that imposing a 'reasonableness' condition as part of the algorithm has the effect of eliminating loops in networks with non-negative link costs. The reasonableness condition is equivalent to Dial's (1971) efficiency condition that paths do not backtrack. Therefore Fox's algorithm provides an efficient method for constructing a choice set of loopless paths that can then be screened for compliance with the side constraints.

Rather than generate a choice set then screen its elements, an alternative approach is to generate paths one by one and accept or reject each before generating the next element. This requires an efficient technique for constructing next best paths. Such a technique can be derived from the observation by Bellman and Kabala (1960) that "if a path ... is to be a second shortest path, then whatever the initial choice, the continuation must be either a shortest path or a second shortest path". In other words, the second best path is a first order deviation from the shortest path. Consider an edge $e(i,j)$ which is part of the second shortest path but not part of the shortest path from s to t then the second best path must reach i along the shortest path from s to i and continue from j along the shortest path from j to t.

The following algorithm for finding the second best path is based on Bellman and Kabala's observation:

STEP 1: Apply a label setting shortest path algorithm backwards from t to produce labels λ_{jt} which are the lengths of the shortest paths from node j to the destination node t.

STEP 2: Scan forwards from the origin node s along the shortest path identified in STEP 1 and construct labels λ_{si} which are the length of the shortest path from the origin node s to node i. At each node i on the shortest path, test deviations by calculating $\lambda_{si} + c_{ij} + \lambda_{jt}$ for each link $e(i,j)$ which is not on the shortest path. The best deviation is the second shortest path.

The algorithm has time complexity $O(rn)$ plus the complexity of an appropriate shortest path algorithm. Note that this approach incorporates the possibility of a path that shares no links with the shortest path since such a path can arise as a deviation at the starting node s. Further, the algorithm can be directly extended to the problem of finding the third, fourth, ... kth best paths. Having identified the best and second best paths, the third best path can be constructed by considering first order deviations from the best and second best paths in STEP 2 of the algorithm, and so on.

The task of calculating measures of utility required for the side constraints can be incorporated into the forward and backward passes in STEPS 1 and 2 for computational efficiency. This means that in STEP 1, each node will be labelled with the length of shortest path to the destination node t on the basis of the optimising attribute, and the cumulative values of the satisficing attributes along the same shortest path.

The advantage of progressive generation of next best paths is that the algorithm may terminate quickly; after generating the shortest path, or a small number of paths. In addition, there is no need to select a value for k, that is, to arbitrarily fix the number of shortest paths that will be generated. There is no theoretical basis for selecting an efficient value for k and there will be unnecessary computation if k is either too large or too small.

4.2 Elimination by Aspects

The screening process described in the previous section generates a preferred path then accepts or rejects the path on the basis of the constraints. It is also possible to reverse the process, that is, to eliminate unacceptable paths then select a preferred path from those already known to meet the constraints. This second approach has much in common with the elimination by aspects (EBA) model developed by Tversky (1972) and a similar choice process proposed by Recker and Golob (1979).

As mentioned above, the EBA and similar models assume that options are differentiated on the basis of several attributes associated with each option and that these attributes can be ranked in importance. Further, each satisficing attribute has a range of acceptable values so a rejection criterion can be associated with each attribute. The rejection criterion can be based on an absolute value, or on a critical tolerance with respect to the optimum value of the attribute. Recker and Golob (1979) have demonstrated a technique for estimating the mean values of acceptance tolerances from survey data and revealed preferences.

The EBA choice mechanism commences by considering the most important attribute. Options that fail to meet the acceptance criterion are eliminated from the choice set then the process is repeated for the reduced choice set and the next most important attribute, and so on. In a network context, this suggests a process of successive pruning of the network.

The EBA process can be efficiently implemented across a network $G(V,E)$ using shortest path techniques described in Section 3 of this paper. Consider the most important satisficing attribute in the choice process, w_1. Using node labelling based on forward and backward passes of a shortest path algorithm for that attribute, each node v can be labelled with the utility $(\lambda_{sv} + \lambda_{vt})$ of the shortest path from s to t through node v. With tolerance C_1 for attribute w_1, any node with $(\lambda_{sv} + \lambda_{vt}) > C_1$ cannot support paths that meet the critical tolerance. It follows that nodes that fail the acceptance can be eliminated from further consideration. Links that start or end at a node failing the criterion are also removed. The outcome is a sub-network $G_1 \subset G$ consisting of nodes that are visited by an acceptable route and their associated links. Note that the procedure guarantees that there is an acceptable path through every node in G_1.

The next step is to repeat the process using the sub-network G_1, the second most important attribute w_2 and its acceptance criterion C_2. Nodes that do not comply with the acceptance criterion are eliminated and the remaining nodes and associated links form the sub-network $G_2 \subset G_1$. The network is successively reduced by applying the constraints in order of attribute importance until either a single route remains or all constraints have been applied.

If more than one route option remains after all criteria have been applied, then the next step is to select the preferred path or paths. With realistic acceptance criteria the reduced network should be very small so the choice set of 'acceptable' routes should also be small. This provides an opportunity to select from a wide range of choice models since complete enumeration of paths becomes feasible and with full enumeration, it is viable to use assignment procedures that make full use of available information. Options range from all-or nothing assignment using a shortest path approach through to complex logit structures. If service frequency effects are significant, as in transit and freight transport networks, then the Spiess and Florian (1989) optimal strategies algorithm, or the Gallagher and Meyrick (1984) algorithm could be used. The Gallagher-Meyrick algorithm requires complete enumeration of paths but makes greater use of available information than does the Spiess-Florian method.

Note that the node elimination process does not guarantee that all possible paths in the reduced network will satisfy all criteria. Instead the process guarantees that there will exist an acceptable path through every node in the reduced network. Therefore some final screening of the choice set may be required. However with realistic criteria the reduced network should be very small and this property is of more theoretical than practical concern.

An alternative approach to selecting the preferred route is to progressively tighten the constraints on one or more attributes and repeat the network pruning process until a single route remains. This approach is usually used in the EBA process but Recker and Golob (1979) have suggested that it is more appropriate to apply a compensatory approach to select the preferred option from the "short list" generated by the initial EBA process.

The time complexity of the network-based elimination by aspects approach will depend on the method that is used to reduce the choice set to a single (or several) preferred routes.

However the overall complexity will be governed by the need for multiple applications of a shortest path algorithm.

Finally it is interesting to contrast some of the features of the approach outlined against corresponding aspects of the standard EBA process. In the normal EBA process, the choice set is an input to the process whereas in a network context, the choice set is not prescribed. It is generally impractical to fully enumerate routes so network algorithms implicitly construct route options as part of the algorithm. The standard EBA and the network EBA also differ in the elimination process. In the standard EBA process, options are directly eliminated whereas in the network algorithm, options are indirectly eliminated by removing nodes from the network.

4.3 Embedded Assignment to Acceptable Routes

The third approach involves an elimination mechanism that is embedded directly into the assignment algorithm. For the hybrid algorithms discussed in this paper, a particular path through the network is in the choice set if it is 'acceptable' in the sense that it meets criteria imposed on the satisficing attributes. The concept of an 'acceptable' route is similar to the notion of an 'efficient' route used by Dial (1971) in his multipath route assignment algorithm, and by Spiess and Florian (1989) for their transit assignment algorithm. In the context of the Dial and Spiess-Florian algorithms, an 'efficient' route is one that takes the traveller closer to the destination and contributes to a lowering of the expected travel time. However, efficiency is simply one of many non-compensatory acceptance criteria that could be imposed on route choice.

In Dial's algorithms, inefficient options are automatically assigned zero probability of selection and the consideration of paths is limited to efficient ones. The same concept can be applied to a hybrid decision process. Consider Dial's algorithm with the strong efficiency criterion, that is, a route must take the traveller further away from the trip origin *and* closer to the destination. Let $\underline{L}(v,w;l)$ be the length of the shortest path from node v to node w on the basis of attribute l then Dial's algorithm will assign a non-zero "likelihood" to a link (i,j) for trips from node s to node t if it satisfies

$$\underline{L}(s,i;l) < \underline{L}(s,j;l) \quad \text{and} \quad \underline{L}(j,t;l) < \underline{L}(i,t;l)$$

Instead of imposing constraints on the optimising attribute to limit consideration to "efficient" paths, we can impose constraints on the satisficing attributes to ensure that unacceptable paths have a zero probability of being selected. Dial's efficiency conditions can be replaced or augmented by constraints corresponding to the acceptance criteria on satisficing variables. The acceptance criteria for a link $e(i,j)$ can be written

$$\underline{L}(s,i;w_i) + L(i,j;w_i) + \underline{L}(j,t;w_i) \leq C_i \quad \forall i$$

where the lengths $\underline{L}(u,v;w_i)$ are constructed using shortest paths techniques as discussed in Section 3 of this paper. If the link fails the criteria then it will be assigned zero likelihood of being chosen.

Therefore non-compensatory elements can be embedded into Dial's algorithm to represent a hybrid route choice mechanism of the type being considered in this paper. The time complexity of the algorithm will be the same as for Dial's algorithm with the strong efficiency condition.

5. EXAMPLE

The screening, elimination, and embedded algorithms can be illustrated using a simple example. Consider the network of directed links shown in Figure 1. Associated with each link are two attributes which, for convenience, can be called cost and travel time. Each link in Figure 1 is labelled with the link cost and travel time, with the cost given first and stated in currency units, then the link travel time given in units of days. The direction of the link is indicated by the arrow head.

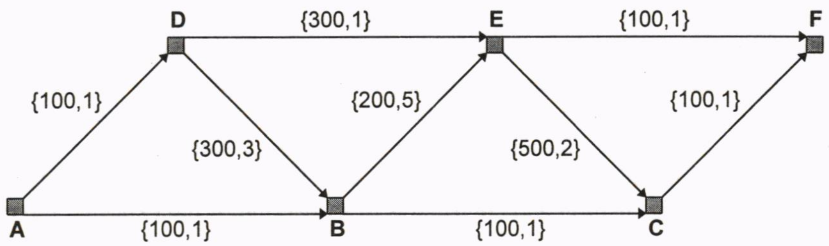

Figure 1: Example Network

Now consider the route choice problem of finding the preferred path or paths from node A to node F on the basis of cost, but subject to a maximum acceptable travel time of 6 days. Problems of this type are common in areas such as international freight movements where keeping cost to a minimum is an important consideration, but market forces or perishability of the cargo demand that the shipment must arrive within a given time.

There are eight possible paths from A to F and to assist with understanding this example, the paths are enumerated and their lengths in terms of cost and time are shown in Table 1.

Table 1: Enumerated Paths

Path from A to F	Total Cost	Total Time
A-B-C-F	300	8
A-B-E-F	400	7
A-B-E-C-F	900	9
A-D-B-C-F	600	11
A-D-B-E-F	700	10
A-D-B-E-C-F	1200	12
A-D-E-F	500	3
A-D-E-C-F	1000	5

5.1 Example of the Screening Method

The screening method for finding the preferred path involves constructing the least cost path then screening it for compliance with the constraint on time; then the second best path; and so on until a path is found that satisfies the constraint. Figure 2 shows link costs plus node labels after applying a suitable shortest path algorithm both forwards from node A, and backwards from node F using cost as the link impedance. In Figure 2, each node is labelled with three numbers; the first label is the length of the least cost path forwards from A to that node; the second label is the length of the least cost path backwards from F to that node; and the third label is the sum of the other two labels which corresponds to the length of the shortest path from A to F through that node.

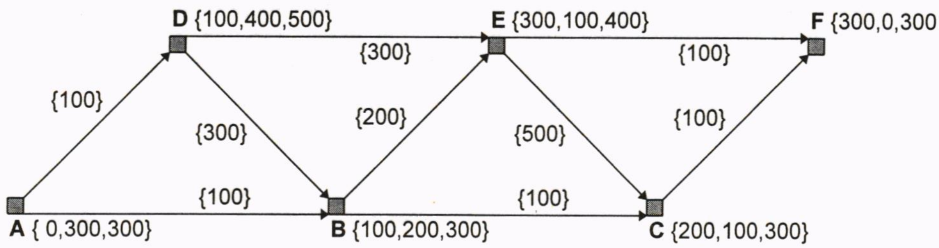

Figure 2: Node Labels - {Forward, Back, Total} Cost

The shortest path from A to F is A-B-C-F but as shown in Table 1 this path takes 8 days and hence is unacceptable. The next step is to construct and screen the next least costly path. The second best path can be constructed by considering first order deviations from the least cost path. Deviations are possible at A and B. The cost of a deviation at A is the sum of the forward label at A, the A-D link cost, and the backward label at D, that is, 0+100+400 = 500. Similarly, the cost of the deviation at B is the sum of the forward label at B, the B-E link cost, and the backward label at E, that is, 100+200+100 = 400. Therefore the second least costly path is A-B-E-F but this path has a travel time of 7 days and is also unacceptable. The next least costly path is constructed by considering first order deviations from the first and second least costly paths. Deviations are possible at A and E. As already found, the cost of a deviation at A is 500 units and the cost of a deviation at E is 300+500+100 = 900. Therefore A-D-E-F is the third least costly path. It has a travel time of 3 days which satisfies the constraint, so A-D-E-F is the preferred route.

5.2 Example of the Elimination Method

The elimination method involves constructing a sub-network by eliminating those nodes that are not visited by a route with acceptable travel time, then finding the least cost route through the sub-network. Figure 3 shows node labels after applying a suitable shortest path algorithm both forwards from node A, and backwards from node F using travel time as the link impedance. In Figure 3, each node labelled with the duration of the quickest path forwards from A to that node; the duration of the quickest path backwards from F to that node; and the duration of the least travel time path from A to F through that node.

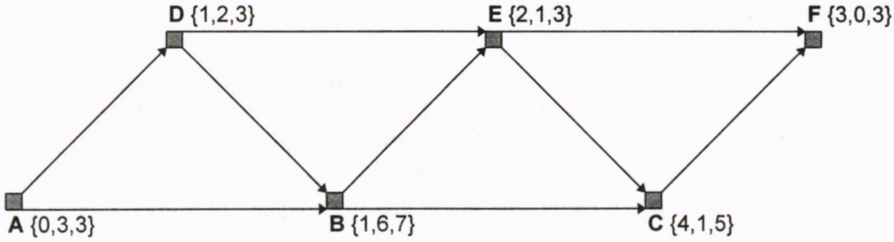

Figure 3: Node Labels - {Forward, Back, Total} Travel Time

The third node label is the duration of the least travel time path from A to F through that node, so all nodes support paths with acceptable travel time, except node B. The duration of the path with least travel time from A to F that passes through node B is 7 days. Therefore node B and all associated links can be removed from the network. The resulting sub-network with its link costs and travel times is shown in Figure 4.

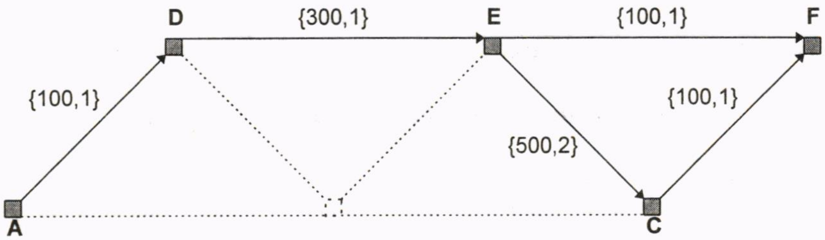

Figure 4: Sub-Network of Acceptable Nodes and Links

It is easily verified that the least cost path through this network is A-D-E-F and that this path has travel time of 5 days. Therefore A-D-E-F is the preferred route.

5.3 Example of the Embedded Assignment Method

This method involves a stochastic assignment to acceptable routes. It can be illustrated by applying a variant of Dial's algorithm to assign selection probabilities to paths in the example network on the basis of cost, while ensuring that unacceptable paths (those with travel time greater than 6 days) have zero probability of being chosen. The first step is to apply a shortest path algorithm forwards from A to set forwards node labels on the basis of both cost and travel time. Then apply a shortest path algorithm backwards from F to set backwards labels on the basis of travel time. Each node now has three labels; least cost from A to the node; least travel time from A to the node; and least travel time from the node to F. The labels are shown on Figure 5.

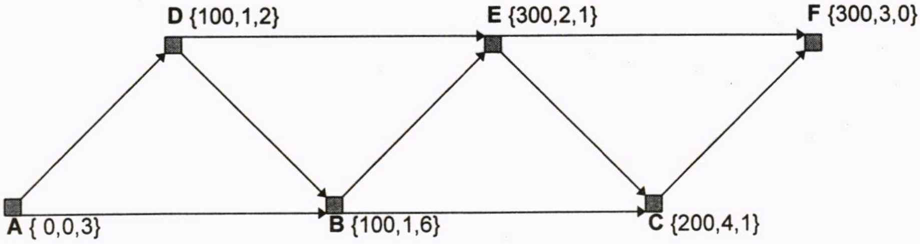

Figure 5: Node Labels - Cost and Travel Time

Dial's algorithm can now be applied in a forward sweep from A to set link weights w_e on the basis of cost. With Dial's algorithm adjusted to assign non-zero probabilities only to acceptable routes, the weight for link *e(i,j)* of the example network can be written as

$$w_e \;=\; \exp\{\theta[\underline{L}(A,j;Cost) - \underline{L}(A,i;Cost) - L(i,j;Cost)]\}$$

if $\underline{L}(A,i;Time) + L(i,j;Time) + \underline{L}(j,F;Time) \leq 6$, or $w_e = 0$ otherwise.

Let the diversion parameter be $\theta = 0.01$, then for link A-D

$$\underline{L}(A,A;Time) + L(A,D;Time) + \underline{L}(D,F;Time) = 0 + 1 + 2 \leq 6$$

so

$$\begin{aligned} w_{AD} &= \exp\{0.01*[\underline{L}(A,D;Cost) - \underline{L}(A,A;Cost) - L(A,D;Cost)]\} \\ &= \exp\{0.01*[100 - 0 - 100]\} \\ &= \exp(0) \;= 1 \end{aligned}$$

For link A-B,

$$\underline{L}(A,A;Time) + L(A,B;Time) + \underline{L}(B,F;Time) = 0 + 1 + 6 \;= 7$$

which does not satisfy the acceptance criterion so $w_{AB} = 0$. Using the same technique, weights can be calculated for each link as shown in Figure 6.

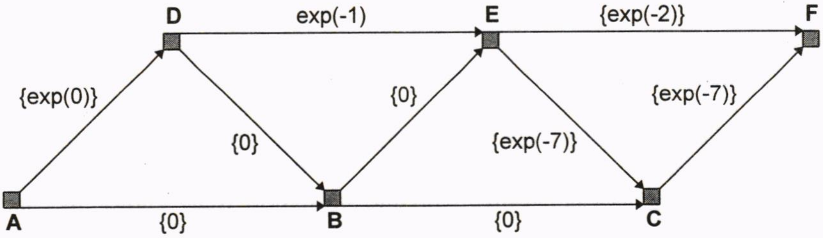

Figure 6: Link Weights

The second stage of Dial's algorithm is to apply a backward pass from F to convert the link weights to probabilities. The results are shown in Figure 7.

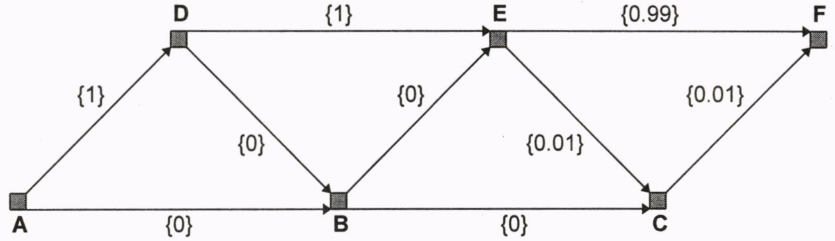

Figure 7: Link Probabilities

Therefore the path A-D-E-F would be preferred 99% of the time and A-D-E-C-F would be preferred 1% of the time.

6. CONCLUSIONS AND FURTHER RESEARCH

This paper has presented several approaches that can be used for implementing hybrid route choice mechanisms in a network context. The first is an all-or-nothing assignment that involves screening shortest paths for compliance with the constraints; the second uses a network equivalent of the EBA process to create choice sets of acceptable options then applies an assignment process to the sub-network; and the third embeds the elimination of unacceptable options directly into an assignment algorithm, such as Dial's algorithm.

In computational terms, these techniques are much less efficient than unconstrained all-or-nothing or stochastic assignment methods but there are many situations in which the behavioural richness of a hybrid choice process can justify the additional computational overhead. Studies of the actual behaviour of decision makers faced with complex transport decisions, particularly those involving direct payment for transport services, have shown that a hybrid choice model provides a much more realistic and faithful representation of the process actually undertaken by decision makers.

The aim of this paper has been to present the hybrid route choice algorithms without detailed discussion of their implementation. The next step is to further investigate the algorithms, and test their relative computational efficiency and performance for realistic transport networks. There appears to be considerable scope to refine and extend the algorithms, in particular, there are opportunities for the development of efficient data structures and efficient techniques for applying the hybrid route choice mechanism to consideration of multiple origins and destinations. Testing on realistic transport networks is essential because experience with unconstrained shortest path algorithms has shown that the theoretical worst-case efficiency of algorithms is not necessarily a good indicator of the algorithm's performance with networks of practical interest. These implementation issues are an important direction for further research.

REFERENCES

Aneja, Y.P., Aggarwal, V., and Nair, K.P.K. (1983) Shortest chain subject to side constraints. **Networks 13**, 295-302.

Bellman, R. (1958) On a routing problem. **Quart. Appl. Math. 16**, 88-90.

Bellman, R. and Kabala, R. (1960) On *k*th best policies. **J. SIAM 8**, 582-585.

Christofides, N (1975) **Graph Theory: An Algorithmic Approach**. Academic Press, New York.

Brooks, M.R. (1990) Ocean carrier selection criteria in a new environment. *Logistics and Transportation Review,* **26**, 339-355.

D'Este, G M and Meyrick, S (1992) Carrier selection in a RO/RO ferry trade. Part 1: Decision factors and attitudes. **Maritime Policy and Management 19**, 115-126.

D'Este, GM (1992) Carrier selection in a RO/RO ferry trade - 2. conceptual framework for the decision process. **Maritime Policy and Management 19**, 127-138.

Dial, R.B. (1971) A probabilistic multipath traffic assignment model which obviates path enumeration. **Transportation Research 5**, 83-111.

Dial, R.B. (1996) Bicriterion traffic assignment: basic theory and elementary algorithms. **Transportation Science 30**, 93-111.

Dijkstra, E.W. (1959) A note on two problems in connection with graphs. **Numerische Mathematik 1,** 269.

Floyd, R.W. (1962) Algorithm 97: shortest path. **Communications of the ACM 5**, 345.

Fox, B.L. (1973) Data structures and computer science techniques in operations research. **Operations Research 26**, 686-717.

Gallagher, F.D. and Meyrick, S.J. (1984) ASEAN-Australia liner shipping: a cost-based simulation analysis, **ASEAN-Australia Economic Papers 12**, ASEAN-Australia Joint Research Project, Canberra.

Gallo, G. and Pallatino, S. (1984) Shortest Path methods in transportation models. **Transportation Planning Models** (Edited by M. Florian), Elsevier, New York.

Garey, M.R. and Johnson, D.S. (1979) **Computers and Intractability: A Guide to the Theory of NP-Completeness.** Freeman, San Francisco.

Jaffe, J.M. (1984) Algorithms for finding paths with multiple side constraints. **Networks 14**, 95-116.

Katoh, N., Ibaraki, T. and Mine, H. (1982) An efficient algorithm for K shortest simple paths. **Networks 12**, 411-427.

Newell, A. and Simon, H.A. (1972) **Human Problem Solving.** Prentice-Hall, Eaglewood Cliffs, New Jersey.

Pape, U. (1980) Algorithm 562: shortest path lengths. **ACM Transactions on Math. Software 6**, 450-455.

Recker, W.W. and Golob, T.F. (1979) A non-compensatory model of transportation behaviour based on sequential consideration of attributes. **Transportation Research 13B**, 269-280.

Saleh, F. and LaLonde, B.J. (1972) Industrial buying behaviour and the motor carrier selection decision. **Journal of Purchasing 8,** 18-33.

Spiess, H. and Florian, M. (1989) Optimal strategies: a new assignment model for transit networks. **Transportation Research 23B**, 83-102.

Tversky, A. (1972) Elimination by aspects: a theory of choice. **Psychological Review 79,** 281-299.

Yen, J Y (1971) Finding the *k*-shortest, loopless paths in a network. **Management Science 17**, 712.