The University of Michigan

IVHS

Intelligent Vehicle-Highway Systems

Fastest Paths in Dynamic Networks with Application to Intelligent Vehicle-Highway Systems

David E. Kaufman
Robert L. Smith

August 1990

IVHS Technical Report-90-4

UNIVERSITY OF MICHIGAN
TRANSPORTATION RESEARCH INSTITUTE • (313) 936-1066
2901 Baxter Road, Ann Arbor, MI 48109-2150, and

COLLEGE OF ENGINEERING • (313) 764-4333
4112 EECS, Ann Arbor, MI 48109-2122

Fastest Paths in Dynamic Networks with Application to Intelligent Vehicle/Highway Systems

David E. Kaufman
Robert L. Smith
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan 48109

August 26, 1990

Abstract

We consider a class of cyclic dynamic networks which is characterized by (possibly random) arc travel times which depend on the time when travel on the arc is begun. We provide a condition which is sufficient for a Dynamic Principle of Optimality to hold, so that the natural extension of conventional shortest-path methods solve for the fastest path with no more computational effort in the worst case than in an ordinary (i.e., static) cyclic network. We discuss how static shortest-path algorithms may be used to improve previously proposed algorithms for cases in which the Principle of Optimality does not hold. The paper is motivated by issues arising from routing drivers using real-time information in an Intelligent Vehicle/Highway Systems (IVHS) environment. The potential benefits of dynamic network modelling in this environment are demonstrated through modifications to the INTEGRATION traffic simulation program.

1 Introduction

An element of the emerging field of Intelligent Vehicle/Highway Systems (IVHS) is the enhancement of travel quality by improving drivers' route choice. The traffic network can be modeled as a directed graph, with some generalized travel cost associated with each arc of the graph, corresponding to a link in the network. Ideally, IVHS would monitor the network, measuring these travel costs in real time and disseminating them to hardware installed in individual vehicles. Then all that would remain is for the driver to program his destination (and perhaps his current location, if Automatic Vehicle Location technology lags) and then his personal guidance unit would employ a shortest-path algorithm to provide him with an optimal, i.e., least-cost route.

This approach overlooks the rapid change which often characterizes road networks, particularly in rushhour conditions when benefits of IVHS may be greatest. Changing traffic volumes and incidents such as lane blockages will alter the time required to travel links of the network, and this travel time will generally be a primary, if not the only, component of travel cost. One way to approach this dynamic character of traffic networks is to update the system measurements very frequently, so that the inputs to the shortest-path algorithm are as recent as possible. However, this does not address the basic inaccuracy of the model. Since travel costs will dynamically change, the costs should be modelled as varying with time. (Of course, this creates a related problem, not addressed in this paper, of forecasting these costs for each link of the network over an appropriate horizon.) In this paper we address this need for the case when cost consists solely of travel time. We will provide a rigorous foundation for the modelling and solution of these networks, and we will demonstrate potential benefits to drivers who have the opportunity to determine fastest paths given the dynamic network model. Although discussion centers on the context of IVHS, the results presented are general to networks with costs known to vary.

2 Literature Review and Synopsis

The fastest path problem with time-dependent travel times was first considered by Cooke and Halsey [2]. They cited Bellman's Principle of Optimality [1] to write a functional equation which implicitly defined a network having states incorporating not only the current location in the network (i.e., node), but also the current time. Travel times are assumed to be known for each arc at times $t_0, t_0 + \Delta, t_0 + 2\Delta, \dots$ and are multiples of Δ, for some $\Delta > 0$. Their solution algorithm proceeds recursively on the maximum number of nodes visited in a path, and if some upper bound $T_{max}\Delta$ can be found for the optimal trip time, then the computational effort is $O(N^2 T_{max}^2)$.

Dreyfus [4] observed that Cooke and Halsey's implicit expansion of the state space and restriction to discrete time intervals can be avoided, and the problem solved by a generalization of Dijkstra's method [5], as efficiently as for static shortest path problems (constant link travel times). However, we will provide a counterexample, and then rigorously establish conditions under which generalizations of conventional static shortest-path algorithms may be applied. We will also discuss how Cooke and Halsey's algorithm may be accelerated by the application of a generalized static shortest-path algorithm when the latter is not guaranteed to provide an optimal solution.

Hall [6] considers networks with random time-dependent travel times and demonstrates that in general, static algorithms cannot be applied. He shows that adaptive routing is necessary, since policies in which the path chosen from an intermediate node depends on the time of arrival at that node may have expected travel times less than that of any fixed routing policy. We will furnish conditions which allow these networks to be solved by static algorithms at computational cost similar to that of static deterministic shortest path solutions. Because these conditions are somewhat restrictive, we will discuss how static methods can be used as heuristics to improve performance of the general-case algorithm proposed by Hall.

Finally, we will present computational results demonstrating the travel-time reduction which can be realized by a vehicle selecting its route in a road network by anticipatory fastest-path calculation. These results are the product of modifications to Dr. Michel Van Aerde's INTEGRATION traffic simulation program, which is ideally suited to investigation of varying methods of route choice for individual drivers.

3 Modelling and Solution in Deterministic Dynamic Networks

Consider a network $(\mathcal{N}, \mathcal{A})$ with node set $\mathcal{N} = \{1, \ldots, N\}$ and arc set $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. Let $c_{ij}(t)$ be the travel time on arc (i, j) departing node i at time t for $t \in T$, where T and $c_{ij}(\cdot)$ satisfy:

1. $0 \in T$

2. $t + c_{ij}(t) \in T$ for all $t \in T$.

We assume the absence of negative cycles, i.e., sequences of nodes $i_1, i_2, \ldots, i_M, i_{M+1}$ with $i_{M+1} = i_1$ such that $\sum_{k=1}^{M} c_{i_k i_{k+1}}(t_k) < 0$ where $t_{k+1} = t_k + c_{i_k i_{k+1}}(t_k)$ for $k = 1, \ldots M$ and $t_1 \in T$. We may have $T = \{t : t \geq 0\}$ and $c_{ij}(\cdot) \geq 0$, or we may select some arbitrary time unit and set $T = \{0, 1, 2, \ldots\}$ in these units if we require that $c_{ij}(t)$ is given in integer numbers of these units. Such a network will be called *dynamic*. Our task is to determine a

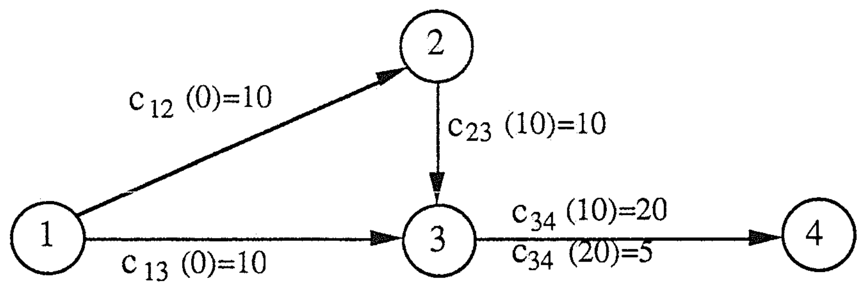

Figure 1: Violation of the Dynamic Principle of Optimality

fastest path from node 1 to node N, assuming that we begin the trip at the current time $t = 0$.

Define the optimal value function $f_i(t)$ to be the minimum trip time over all paths from node i to node N departing node i at time t. Then Cooke and Halsey's functional equation is:

$$f_i(t) = \begin{cases} \min_{j \neq i}\{c_{ij}(t) + f_j(t + c_{ij}(t))\} & \text{for } i = 1, \ldots, N-1;\ t \in T \\ 0 & \text{for } i = N; t \in T \end{cases} \tag{1}$$

and the problem is to find $f_1(0)$. Implicitly, this defines a network with node set $\mathcal{N}' = \{(i,t) : i \in \mathcal{N}, t \in T\}$ and arc set $\mathcal{A}' = \{((i,t),(j,u)) : (i,j) \in \mathcal{A}; t,u \in T; c_{ij}(t) = u - t\}$. Since this network has fixed travel times, we will refer to it as the *expanded static network*. The expansion of the state space increases the computational effort required to solve the problem; Cooke and Halsey's solution algorithm requires T to be finite, with a known upper bound T_{max} on $f_1(0)$, and requires $O(N^2 T_{max}^2)$ time. In comparison, a network with fixed travel times can be solved in $O(N^2)$ time, or $O(N \log N)$ time for sparse networks [3].

Dreyfus [4] observed that the problem could be solved by generalizing the method of Dijkstra [5]. He proposes that labels v_i be maintained, and at each iteration the node i with minimum v_i be *expanded*; that is, v_i is made permanent and the labels v_j are updated if appropriate, for each j such that $(i,j) \in \mathcal{A}$. The claim is that at termination, v_i represents the optimal trip time from node 1 to node i for all $i \in \mathcal{N}$. However, Figure 1 demonstrates that this procedure may fail in dynamic networks. The procedure suggested by Dreyfus would identify (1,3,4) as the optimal path from node 1 to node 4, with trip time 30, but the path (1,2,3,4) has a trip time of 25.

Dreyfus' suggestion implicitly assumes the following functional equation:

$$f_j = \begin{cases} \min_{i \neq j}\{f_i + c_{ij}(f_i)\} & \text{for } j = 2, \ldots, M \\ 0 & \text{for } j = 1 \end{cases} \tag{2}$$

where f_i is the minimum trip time over all paths from node 1 to node i departing node 1 at time 0. This assumes Bellman's Principle of Optimality [1] in its forward-recursive form.

For (2) to hold, it must be true that on an optimal path, each intermediate node is reached as soon as possible. Clearly, this is not the case in Figure 1.

However, the following assumption [8] will suffice to imply (2).

Assumption 1 *$s + c_{ij}(s) \leq t + c_{ij}(t)$ for all $s, t \in T$ such that $s \leq t$.*

In Figure 1, $c_{34}(\cdot)$ violates Assumption 1. If we arrive at node 3 at time 10, we travel link (3,4) in 20 minutes, arriving at time 30. By arriving 10 minutes later, we can travel the link in only 5 minutes, arriving at time 25. In the context of vehicle routing, this requires us to believe that if a driver leaves node 3 at time 10 and another driver leaves node 3 10 minutes later, the second driver necessarily *passes* the first. However, both of these drivers represent the single driver whose optimal path is being sought, and so violation of the assumption seems to require the driver's behavioral characteristics to be a function of time. In the absence of such an unlikely effect, Assumption 1 does not appear to be overly restrictive.

There can be situations within the traffic environment in which the assumption is violated due to non-behavioral causes. In a surface street area, stoplights may be coordinated so that traffic moving in in a given direction without turning will hit a string of green lights, avoiding stop-and-go conditions. Suppose one car is stopped at the first of such a string of lights waiting for the light to change from red to green, and a second car reaches the intersection just after the change occurs. The second car will not have to stop, and will pass the first car easily, reaching the next light (*i.e.*, the end of the next link) earlier even though he reached the beginning of the link later [9]. This effect may be small because it will be difficult to predict at what part of the stoplight cycle a driver will reach an intersection.

When Assumption 1 does hold, we may prove the desired principle of optimality by working with the expanded static network $(\mathcal{N}', \mathcal{A}')$ defined by (1).

Theorem 3.1 *(Principle of Optimality for Dynamic Networks) Under Assumption 1, for all $j \in \mathcal{N}$ there exists an optimal path $(1,0), \ldots, (i, s^*), (j, t^*)$ in the expanded static network from node 1 at time 0 to node j at a minimal time t^*, such that the truncation $(1,0), \ldots, (i, s^*)$ is an optimal path in the static network from node 1 at time 0 to node i at a minimal time s^*.*

Proof: Consider any optimal path $P' = (1,0), \ldots, (i, s'), (j, t')$ in the static network. (Such a path must exist in the absence of negative cycles.) Consider an optimal path $P_i^* = (1,0), \ldots, (i, s^*)$ from node 1 to node j. (Unlike the notation in the statement of the theorem, P' and P_i^* are not assumed to have the same intermediate nodes.) Let

$$t^* = s^* + c_{ij}(s^*)$$

The claim is that the path $P^* = (1,0), \ldots, (i, s^*), (j, t^*)$ travelled by following P_i^* and then proceeding directly from node i to node j is an optimal path from node 1 at time

0 to node j.
Because P_i^* is optimal, $s^* \leq s'$. Then

$$\begin{aligned} t^* &= s^* + c_{ij}(s^*) \\ &\leq s' + c_{ij}(s') \qquad \text{by Assumption 1} \\ &= t' \end{aligned}$$

By assumption, t' is a minimal time to reach node j. Therefore, t^* is also minimal, and hence P^* is optimal. ∎

Therefore, under Assumption 1, it is optimal to reach each intermediate node of a path as early as possible, and (2) is valid. Note how the optimal value function in (2) serves the function of the state variable t in (1), thus allowing for a more computationally efficient solution. In fact, when $c_{ij}(\cdot) \geq 0$, as is typical for transportation networks, any algorithm which suffices to solve a deterministic shortest path problem will solve (2).

To see this, consider such algorithms as being divided into two classes, label-setting and label-correcting (also called best-first and list-search, respectively). Label setting algorithms such as Dijkstra's method [5] select a label v_j to make permanent, and guarantee that when this is done, the label is indeed at the optimal value for node j. Therefore, the only arc travel time data required are $c_{ij}(f_i)$ for all $i, j \in \mathcal{N}$. Hence the network may be considered to have time-invariant travel times, and problem is an ordinary deterministic shortest path calculation.

In contrast, label-correcting algorithms may require $c_{ij}(t)$ for multiple values of t. Each time the expansion of a node i causes a decrease of a label v_j (indicating the discovery of a superior path to node j), node j is added to a list of nodes to be expanded, and the algorithm terminates only when the list is empty. When a node i is expanded, the arc travel time $c_{ij}(v_i)$is used. There are two cases. Firstly, v_i may equal f_i. This case reduces to that of the label-setting algorithms. Secondly, we may have $v_i > f_i$, and hence the value $c_{ij}(v_i)$ is not the arc travel time experienced on an optimal route which traverses arc (i, j). Suppose the algorithm terminates in this condition, and let k be the node preceding i on the optimal path from node 1 to node i. If $v_k = f_k$, then node i must not have been expanded since the last update of v_k, which contradicts the stopping rule. So $v_k > f_k$. We may continue this argument inductively on the number of nodes visited on a path until we find that $v_1 > f_1$. But since both of these values must be 0, a contradiction ensues, and hence the label-correcting algorithms succeed.

When Assumption 1 does not hold, the application of standard shortest path algorithms is still useful. Cooke and Halsey's algorithm requires an upper bound T_{max} on the optimal trip time, and the tighter the bound, the fewer operations required in the algorithm. They propose that the value $c_{1N}(0)$ be used, and when this value is infinite (i.e., $(1, N) \notin \mathcal{A}$), their scheme for devising T_{max} is enumerative. In contrast, applying a standard shortest path algorithm may provide a tight upper bound.

4 Stochastic Dynamic Networks

Travel time on a link of the road network is is subject to uncertainty for many reasons. These include the chance of an incident occurring to reduce link flow capacity, uncertainty in the volume of traffic which will be on the link, random interaction of the drivers on the link, and delays caused by drivers making turns, onto links which may themselves be congested, causing spillback. Modelling travel times as stochastic quantities may significantly improve routing quality.

In keeping with the deterministic development, let the random variables $C_{ij}(t)$ be the travel time for link (i, j) departing node i at time t, for $t \in T$, where T and $C_{ij}(\cdot)$ satisfy

1. $0 \in T$
2. With probability 1, $t + C_{ij}(t) \in T$ for all $t \in T$.

Constructing the corresponding static stochastic network, we define the optimal value function $f_i(t)$ to be the minimum expected trip time over all paths from node i to node N departing node i at time t. Then we have the functional equation

$$f_i(t) = \begin{cases} \min_{j \neq i}\{E[C_{ij}(t) + f_j(t + C_{ij}(t))\} & \text{if } i = 1, \ldots, N-1;\ t \in T \\ 0 & \text{if } i = N; t \in T \end{cases} \tag{3}$$

where E denotes expectation with respect to $C_{ij}(t)$. Then the problem is again to find $f_1(0)$. As before, we can make an assumption which will allow us to compress the state space of the static network.

Assumption 2 *$E[S + C_{ij}(S)] \leq E[V + C_{ij}(V)]$ for all random variables S and V whose values are restricted to T with probability one such that $E[S] \leq E[V]$, where in the first expression E represents expectation with respect both to V (or S) and to $C_{ij}(\cdot)$.*

Theorem 4.1 *(Principle of Optimality for Stochastic Dynamic Networks). Under Assumption 2, for all $j \in \mathcal{N}$ there exists an optimal path $(1,0), \ldots, (i, S^*), (j, T^*)$ in the static network from node 1 at time 0 to node j at a random time T^* with minimal expectation, such that the truncation $(1,0), \ldots, (i, S^*)$ is an optimal path in the static network from node 1 at time 0 to node i at a random time S^* with minimal expectation.*

Proof: Follows the proof of Theorem 3.1 ∎

As before, we can replace the backward recursion of (3) by a forward recursion with a reduced state space. Define f_i to be the minimum expected trip time over all paths from node 1 to node i departing node i at time 0. The functional equation is

$$f_j = \begin{cases} \min_{i \neq j} \{f_i + E[C_{ij}(f_i)]\} & \text{for } j = 2, \ldots, N \\ 0 & \text{for } j = 1 \end{cases} \tag{4}$$

We may consider $E[C_{ij}(f_i)]$ to be a deterministic function of f_i, and hence this functional equation has an identical structure to the deterministic functional equation (2), so we may use identical solution procedures.

Assumption 2 is more restrictive than its deterministic counterpart, and when the condition holds, it will be difficult to verify. Hall's counterexample [6] shows that a simple network operating in a reasonable fashion can violate Assumption 2 and hence prevent optimal solution by static shortest-path algorithms. To solve the problem, he constructs a revised network with identical topology but with a fixed travel time $\hat{c}_{ij}$ for each link (i, j) equal to the minimum possible travel time on that link over all times at which travel on the link may be begun; that is, $\hat{c}_{ij} = \inf_{t \in T}$ {infimum of the support of $C_{ij}(t)$}. The mth iteration of the solution algorithm identifies the mth shortest path in the revised network and calculates the expected trip time T_m for this path. T_m serves as an upper bound on the optimal expected trip time, and the trip time for this path in the revised network serves as a lower bound for all paths not analyzed in iterations $1, \ldots, m$. The algorithm terminates when the lower bound is greater than or equal to the upper bound.

Even when Assumption 2 does not hold, applying static methods based on functional equations (2) and (4) will improve the algorithm by tightening both the upper and lower bounds, and hence terminating as early or earlier. Before the algorithm begins, a static shortest-path algorithm should be applied, serving as a heuristic to identify a greedy path, providing an upper bound which otherwise might not be identified until Hall's algorithm has iterated a number of times. The effort for this preprocessing is that of an ordinary shortest path calculation. Within the algorithm itself, the calculation of mth shortest paths should be time-dependent. That is, since $\hat{c}_{ij}$ is an infimum over all times, the corresponding time-dependent paths based on $\hat{c}_{ij}(t) =$ {infimum of the support of $C_{ij}(t)$} must be no shorter, providing lower bounds which are as tight or tighter. Since these time-dependent paths in the revised network are deterministic, Assumption 1 is not restrictive, and hence the Principle of Optimality holds, allowing the use of static k-th shortest path methods, at no additional computational effort. In particular, the assumption is satisfied in the bus network on which Hall demonstrates the algorithm.

5 Computational Results: Anticipatory Route Choice in INTEGRATION

The benefits available to individual drivers using a dynamic network model have been investigated through the INTEGRATION traffic simulation [10], developed by Professor Michel Van Aerde at Queens University, Kingston, Ontario, Canada. INTEGRATION models traffic microscopically, maintaining location and destination information for each vehicle in the simulated network. The main issues addressed by INTEGRATION are first, interaction of freeway corridors and signalized networks, and second, opportunities for optimal route choice by individual drivers given real-time information. The latter makes INTEGRATION well-suited to examine the potential benefits of a dynamic network model.

INTEGRATION maintains one category of vehicles which may be considered background traffic; these drivers follow a fixed route from their origin to their destination regardless of network conditions. Vehicles in a second category are provided with an "optimal" routing each time they reach nodes, hence having decision opportunities. This optimal routing is calculated by the application of a static shortest-path algorithm to a network model with fixed travel times determined from the latest real-time measured travel times within the simulation. These travel times are determined by the difference between link entry and departure times for vehicles which have traveled the link, and hence include not only the time required to travel the link distance at some projected speed but also the delay incurred due to stoplights, incidents and lane blockages, and queueing caused by congestion. However, this "dynamic" routing optimization depends on a static network model with fixed link travel times.

We have modified INTEGRATION to add a third category, which performs personal routing optimization on a dynamic network model. The second class of drivers may be routed to links which are currently uncongested, only to discover that the links have become congested by the time the drivers reach that link. In contrast, drivers in the third category look ahead in time to avoid links which will become congested. We describe this dynamic-network route choice as *anticipatory.* The benefits of anticipatory routing are demonstrated in the network shown in Figure (). Due to the loading over time in this network, second-category drivers actually experience a longer average trip time, because their route-choice model assumes that the latest measured travel times will hold permanently.

However, if we execute the simulation a second time, allowing newly introduced anticipatory drivers to perform fastest-path calculations based on the actual measured link travel times from the entire first simulation (i.e., $c_{ij}(t)$ data for the entire time horizon), we see that the anticipatory drivers avoid divert from the primary route only when it would be slower than the secondary route *during the trip*, not at the time the decision is made.

This testing procedure requires that the proportion of anticipatory drivers be kept very small in the second simulation, so that the link travel times observed in the first simluation may be considered accurate forecasts of future link travel times during the second simulation. If there are many anticipatory drivers, their improved route choice will cause the network to be differently loaded over time, resulting in different observed link travel times, and invalidating the route choice of the anticipatory drivers. Operating an IVHS system which disseminates short-term travel time forecasts to large numbers of anticipatory drivers in a physical road network will require substantial research on the subject of determining the future of a traffic network based on real-time traffic conditions, incident modelling, stochastic driver behavior and route choice, time-varying travel demands, and so on [7].

References

[1] Bellman, R., *Dynamic Programming*, Princeton University Press, Princeton, N.J., 1957.

[2] Cooke, K. L. and E. Halsey, "The Shortest Route Through a Network with Time-Dependent Internodal Transit Times", *Journal of Mathematical Analysis and Applications* **14**, 493-498 (1966).

[3] Denardo, E. V., *Dynamic Programming: Models and Applications*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1982.

[4] Dreyfus, S. E., "An Appraisal of Some Shortest-Path Algorithms", *Operations Research* **17**, 395-412 (1969).

[5] Dijkstra, E. W., "A Note on Two Problems in Connexion with Graphs", *Numerische Mathematik* **1**, 269-271 (1959).

[6] Hall, R. W., "The Fastest Path through a Network with Random Time-Dependent Travel Times," *Transportation Science* **20**, 182-188 (1987).

[7] Kaufman, D. E., Lee, J., and Smith, R. L., "Anticipatory Traffic Modelling and Route Guidance in Intelligent Vehicle/Highway Systems", Technical Report 90-08, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI, 1990.

[8] Prakash, A. *Private communication*, 1989.

[9] Underwood, S. *Private communication*, 1990.

[10] Van Aerde, M., and Yagar, S., "Dynamic Integrated Freeway/Traffic Networks: A Routing-Based Modelling Approach", *Transportation Research A* **22A** 6, 445-453 (1988).