The Network Signal Design Problem for Long-Range Travel Forecasting

Alan J. Horowitz, Member, ASCE
University of Wisconsin—Milwaukee
Professor of Civil Engineering
PO Box 784
Milwaukee, WI 53201
E-mail: horowitz@uwm.edu
Voice: 414-229-6685
Fax: 414-229-6958

Minnie H. Patel
Associate Professor of Industrial Engineering
San Jose State University
One Washington Square
San Jose, CA 95192-0085
E-mail: Minnie.Patel@sjsu.edu
Voice: 408-924-4152
Fax: 408-924-4040

Revised June 21, 2004

TE-22737

Key Words: Travel Demand; Traffic Signals; Transportation Planning; Optimization; Network Design

The Network Signal Design Problem for Long-Range Travel Forecasting

Abstract

The network signal design problem (NSDP) seeks the optimal deployment of traffic signals in a growing urban area. This paper is especially concerned with how signals may be optimally deployed over a very long period of time for the purpose of creating realistic networks for travel forecasting. The NSDP is very difficult to solve for long-range problems because of the large number of possible solutions, the high cost of evaluating the merits of just a single solution, and the complexities of how signal delay affects traffic patterns and how traffic patterns affect signal delay. The paper describes the NSDP, introduces a reasonable set of simplifications based on transportation planning and traffic engineering practice, describes experiences with a possible heuristic algorithm for problem solution, and contrasts this method with current planning practice and other research. The long-range algorithm embeds a "strategic" algorithm for finding an optimal deployment for a single time period with constant travel demands. The strategic algorithm draws upon two well-known techniques of combinatorial optimization: a greedy constructive search coupled with a restricted neighborhood search. The strategic algorithm was able to find exact solutions on a small test network with eight stop-controlled intersections. The long-range algorithm is demonstrated on a full-sized planning network with about 380 stop-controlled intersections that could be signalized.

The Network Signal Design Problem for Long-Range Travel Forecasting

Alan J. Horowitz[1] and Minnie H. Patel[2]

Introduction

The network signal design problem (NSDP) is rather easy to state. Given a city undergoing growth, what is the optimal arrangement of signalized intersections in some distant future year? There are two variants of this problem that are described in this paper: the strategic NSDP and the long-range NSDP. The strategic NSDP seeks to find the optimal arrangement of signalized intersections at a single point in time. The long-range NSDP seeks to find the staged optimal deployment of signals over a long period of time. The only previous study of the long-range NSDP was conducted by Horowitz and Granato (2000), who developed an algorithm that simulated the actions of a traffic engineer while making decisions to upgrade deficient unsignalized intersections.

The strategic NSDP is a combinatorial problem. In any given network there is a certain set of intersections that are already signalized and another set of intersections that could reasonably be signalized. Potentially, there are as many feasible solutions to the problem as 2 to the power of the number of such intersections. The number of possible solutions is extremely large in all but the smallest cities. Complicating the problem still further is the difficulty of evaluating the merits of just one solution. Each evaluation requires almost a complete traffic forecast, usually involving an equilibrium traffic assignment where delays are properly sensitive to the various traffic controls. Each traffic forecast can take between several minutes for a small city to several

[1] Alan J. Horowitz, Professor of Civil Engineering, University of Wisconsin—Milwaukee, PO Box 784, Milwaukee, WI 53201.

[2] Minnie H. Patel, Associate Professor of Industrial Engineering, San Jose State University, One Washington Square, San Jose, CA 95192-0085.

hours for a big city. Thus, there is a high premium associated with any inefficiency in an algorithm intended to solve this problem, and judicious compromises are necessary to achieve reasonable solutions.

Until recently, those planners doing travel forecasts were blissfully unaware of the need for the NSDP because nearly all travel forecasting networks either ignored the existence of traffic controls or represented them very crudely. However, in the last few years software packages have emerged that explicitly recognize the properties of traffic controls within a forecasting network. In order to properly perform a long-range (e.g., 20 to 30 years) travel forecast with one of these packages the analyst would need to first determine the location of all the signals. Unfortunately, the location of a new signal is dictated largely by the pattern of traffic on the network, which is unknown for a future year without a forecast and strongly influenced by the location of any existing signals within a forecast.

The NSDP has many similarities to and many differences from the network design problem (NDP) that has been the focus of many studies over the past thirty years (see Chen and Alfa (1991) or Solanki, Gorti and Southworth (1998) for recent examples.) The principle difference is that the NDP focuses on links – their presence, absence or characteristics – while the NSDP focuses on nodes. Network design issues relating to nodes are inherently more complex than those related to links. Beyond philosophy, approaches to the NDP are unlikely to be pertinent to the NSDP.

This paper will principally report on experience the authors have had with a heuristic algorithm for finding the solution to the NSDP. The heuristic is a variant of a greedy constructive search, a well-known technique for solving a certain class of combinatorial problems. A pure greedy search would evaluate each unsignalized intersection for entry into the set of signalized intersections on its own merits without regard to its interaction with other intersections in the set. However, a pure greedy search performed poorly on test networks, so the algorithm was augmented with a limited amount of information about the interactions between intersections. The heuristic is demonstrated to be good on a small test network for which an exact solution can be obtained for a strategic NSDP, and then it is tried on a full-sized urban network for evaluation of its computational properties for the long-range NSDP.

Overview of the Network Signal Design Problem

The strategic network design problem can be slightly more formally stated for application in the United States as:

Minimize: a measure of effectiveness

Subject to: (a) each unsignalized intersections meeting a Manual on Uniform Traffic Control Devices (MUTCD) warrant

(b) all link flows and turning movements consistent with

(i) principles of travel behavior

(ii) principles of network equilibrium, including flow conservation

(iii) demographic and socioeconomic conditions at all origins and destinations

(c) delays consistent with all relevant flows, turning movements highway geometry and the nature surrounding traffic controls

(d) any given intersection may assume one of two states – signalized or unsignalized

A formal mathematical statement of constraints (a) to (c) will be bypassed here for the sake of brevity. Instead, citations will be made to sources detailing these constraints and descriptions will be given on how the constraints are specifically implemented for our version of the NSDP.

There is a wide variety of possible measures of effectiveness (MOE). For simplicity and for illustrative purposes, this research concentrated on the total vehicle-hours-travel (VHT) over a whole network, although other MOE's, such as total fuel consumption, air pollution emissions, safety indexes, benefit-cost ratios or some combination, could also be implemented within the NSDP. VHT was selected here as an example MOE because of its prominent place among MOEs used in selecting alternatives for long-range transportation plans and because it allowed comparisons with previous research, but VHT by itself is not necessarily consistent with the judgments of a particular traffic engineers who are deploying signals at intersections. A deployment decision about any given signal is more complex than can be represented within a single MOE, but the long-range NSDP only has available the information that can be reasonably assumed from a long-range transportation plan and travel forecast.

Selection of a Greedy Search

Both the strategic NSDP and long-range NSDP are combinatorial in nature. Two approaches; namely, construction and improvement are straightforward to develop and implement for combinatorial problems (see the comprehensive survey article by Zanakis et. al. (1989)), but customization to the specific problem is essential. Construction algorithms generate a solution by adding individual components (in our case intersections) one at a time until a feasible solution is found. One class of constructive heuristics is greedy algorithms that seek to maximize improvements in each step. The improvement heuristics maintain feasibility and successively improve the solution by sequences of exchanges and merges. All neighborhood search algorithms can be classified as improvement heuristics. Other combinatorial algorithms are available but are not explored here.

Network Issues

Some of the most important MOEs that might be used for signal selection involve vehicular delay. Viewed myopically, one would expect that a delay reduction at one intersection would result in a similar overall delay reduction for the network as a whole. However a direct relationship between intersection and network delay cannot be assumed, as intersections are interconnected in rather complex ways.

First, traffic volumes at many intersections are affected by delays at a single other intersection. This paper makes the broad and common assumption, typical of long-range travel forecasts, that traffic volumes rapidly achieve a user-optimal equilibrium condition. When delay increases or decreases at an intersection, traffic is diverted to or from other intersections. The diverted traffic has significant implications for delay throughout the network.

Second, delays at one intersection are affected by traffic volumes and characteristics of neighboring intersections by a process known as "upstream filtering". Upstream filtering is described in the 1997 and 2000 HCM editions for both signalized and two-way, stop-controlled intersections. For signalized intersections, the delay calculation at an intersection approach involves consideration of the degree of saturation of the upstream signalized intersection, as measured by its overall volume-to-capacity ratio. In the case of two-way, stop-controlled intersections, signalized intersections within 0.40 km introduce platoons that change the character of gaps experienced by drivers attempting to turn onto or cross uncontrolled streams of traffic, thereby affecting delay at signed approaches (Highway Capacity Manual 2000).

Third, Braess' Paradox says that an improvement to an individual network element in a user-optimal traffic assignment does not necessarily improve the network as a whole. The principle reason for Braess' Paradox is that a user-optimal assignment can differ considerably from a system-optimal assignment. It stands to reason that the predicted improvement at an individual intersection may not be a very accurate indicator of the improvement in the network as a whole.

Fourth, any given network with traffic controlled intersections can have multiple, valid equilibrium assignments (Meneguzzer 1997; Horowitz 1992), but only one assignment can be considered in the selection of a signal. Sometimes a comparatively small change to the network can cause the assignment to jump to a state that is quite different.

Fifth, signals can have synergistic effects by their ability to be timed to create higher volume corridors through the network. In this paper, variable green times and phasing plans contribute to this synergy. Signal coordination, handled here by rule (not optimized), can also create synergy.

Practical Simplifications

Without some simplifications, even the strategic NSDP would be impossible to solve for any city of moderate size or larger. Horowitz and Granato (2000) provide considerable guidance as to practical simplifications that would greatly reduce the number of solutions that need to be investigated.

First, it makes sense to consider together all signals that might be required within a limited period of time, for example 4 years or less.

Second, signals are almost never removed once installed. Consequently, signals can be realistically grandfathered into the problem. Grandfathering would mean that the algorithm does not need to consider removal of already installed signals from the solution set for any strategic problem. Grandfathering also would mean that signals accumulate across multiple time periods for any long-range problem.

Third, signals should never be installed unless the intersection meets a MUTCD warrant at the time the signal is being considered. Either (1) warrants can be checked for an individual intersection at the time that intersection is considered for signalization or (2) warrants can be checked for all intersections ahead of considering any intersections within a time period. These two strategies will produce different results, but the number of intersections needing evaluation is greatly reduced in either case. The latter strategy, adopted for this research, has the computational advantage of exactly defining those intersections that could feasibly be signalized prior to the start of the optimization.

For peak hour traffic, which is the subject of the tests described later in this paper, Warrant #3a from the MUTCD Millennium Edition (2001) is most relevant. A signal is warranted if total delay on one minor street exceeds 4 vehicle hours for a one-lane approach while carrying a minimum of 100 vph or exceeds 5 vehicle hours for a two-lane approach while carrying a minimum of 150 vph. In addition, total entering volume must be greater than 800 vph for an intersection with 4 or more approaches or 650 vph for an intersection with three approaches.

Fourth, traffic delay at signals can be adequately calculated with macroscopic relations from the HCM. The HCM is authoritative, but lacks subtleties that may be better captured by some available microscopic traffic simulation models. The primary reason for selecting a macroscopic model is the relative high speed of calculating delay, which is necessary given the extremely large number of delay calculations required by embedding a user-optimal equilibrium traffic assignment within the optimization. Furthermore, travel forecasting networks most often lack the geometric and environmental details that would give a clear advantage to microscopic simulation.

A Practical Heuristic Algorithm for the Strategic NSDP

Figures 1 and 2 briefly describe a heuristic algorithm for the strategic NSDP, which is needed for the long-range problem described later. The strategic NSDP is applied to a single time point in the future (e.g., 4 years hence). A traffic network is created that contains all important road segments and intersections for a city. There are four types of intersections: uncontrolled, two-way (or one-way) stop controlled, all-way stop controlled and signalized. The algorithm selects those two-way and all-way, stop-controlled intersections that should be signalized. An intersection is selected if it has the maximum value among all intersections of a "greedy function". The greedy function is a guess as to the decline or increase in an MOE, in our example a decline in total system VHT, that would occur if the intersection were to be signalized. This greedy function has components of (1) delay at the intersection while unsignalized, (2) delay if the intersection were signalized but with the same volumes, (3) elasticity of intersection volume with respect to intersection delay, and (4) delay effects on neighboring intersections.

The problem is started by estimating peak-hour traffic demands for the time period in the form of a trip table, i.e., the number of vehicle trips between all possible origins and destinations in the network. This trip table is sensitive to existing delays on the network. Once the trip table has been determined, then an initial traffic assignment is performed to estimate volumes, turning movements and delays. The volumes and delays are then checked against MUTCD warrants, in our example just Warrant #3a, Peak Hour Delay, to determine the set of feasible unsignalized intersections. Each feasible intersection is temporarily signalized and a traffic assignment is again performed. This set of traffic assignments allows an initial calculation of the elasticity of total intersection volume with respect to average intersection delay and an initial calculation of the delay effects on neighboring intersections. Given this information, a greedy function can be computed for each intersection.

The unsignalized intersection with the largest greedy function is again temporarily signalized. If the intersection successfully reduces VHT as demonstrated by a completely new traffic assignment, then it is kept. Otherwise, it is rejected and must wait for other intersection(s) to be signalized before being reconsidered. There are two means by which an intersection can be reconsidered. However, the process of rejecting an intersection gives enough information to revise its demand/delay elasticity and effects on neighboring intersections. Thus, any reconsideration of the intersection is done with a completely updated greedy function.

Figure 2 reflects the need for checking each solution with a traffic assignment. The traffic assignment allows determination of revised volumes and turning movements at every intersection and computation of delays resulting from the revised volumes. Thus, the traffic assignment must be capable of calculating delay in accordance with good traffic engineering principles. The traffic forecast also generates a substantial amount of information that can be fed into a greedy function for all remaining unsignalized intersections.

If an intersection is kept, then the remaining intersection with the largest greedy function is considered. The optimization stops when any one of the following happens.

a. There are no more intersections to be considered;
b. All remaining intersections have values of the greedy function that are very negative and below a lower-bound "threshold", that checking any of them would be pointless.
c. There have been a certain number of rejections (e.g., 2 or 3 for this research) in a row.

Figure 2 refers to rejections as "exclusions", because rejected intersections must wait for another intersection to be accepted into the solution before being again considered.

It is possible for a network to have an intersection that when signalized effectively blocks all or nearly all remaining intersections from becoming signalized. Thus, stopping criterion (c) could occur because there are no more worthy unsignalized intersections or because one of the last intersections to be signalized is blocking. To check for blocking, the last one or last two intersections to be signalized are temporarily removed from the solution and the algorithm is restarted. This blocking check produces three solutions or “branches”:

0. Original solution;
1. Solution with the last intersection from the original removed; and
2. Solution with the last two intersections from the original removed.

The solution across branches with the smallest VHT is the overall solution.

Each traffic assignment seeks a user-optimal equilibrium, with paths between origins and destinations that are sensitive to the amount of delay at all traffic controlled intersections. This means that the HCM delay procedures must be embedded into the traffic assignment algorithm.

A Practical Heuristic Algorithm for the Long-Range NSDP

The strategic NSDP does not by itself have much practical relevance for short-term problems because traffic engineers can integrate a far greater amount of local knowledge into the decision making process without it. However, the long-range NSDP would be especially germane to determining signal configurations for long-term transportation plans, where the planning horizon would be many years hence and where local knowledge is obsolete. A heuristic algorithm for the long-range NSDP that embeds the strategic NSDP is shown in Figure 3. It is seen that this algorithm simply divides the time between now and the planning horizon into short time periods. The strategic NSDP is run for each time period, and the signals accumulate as they become part of the end solution to the strategic problem, thereby implementing the grandfathering assumption.

This algorithm, even if an optimal solution could be guaranteed for a individual time period, does not necessarily result in the lowest possible VHT over all time periods, because an intersection that is signalized in an earlier time period might have a detrimental effect on VHT in a later time period.

Discussion

Although this research concentrated on Warrant #3a to select feasible intersections, other warrants might pertain. These warrants would include any that depend directly or indirectly upon on single-hour volumes: Peak Hour Volume, Minimum Pedestrian Volume or School Crossing. More warrants would tend to increase the size of feasible set of intersections at the beginning of each interval, so the research focused on Warrant #3a for computation efficiency. Otherwise, including additional warrants is straightforward.

An elastic-demand trip table is now considered “best practice” for long-range travel forecasts. In the implementation here, demand is elastic at the start of the strategic problem, then fixed thereafter. Elastic demand, in our case studies, means that the trip table, as determined by the trip distribution step, is sensitive to and consistent with the delays along links and at traffic controlled intersections under equilibrium conditions. Locking the trip table before the start of the strategic problem seems to improve the performance of the algorithm. Additional tests would be needed to determine if locking the trip table at the beginning of an interval appreciably affects the solution.

Signalizing one intersection could divert traffic, thereby causing another intersection not in the feasible set to now meet the warrant. Such intersections are ignored until the next time period. Conversely, another intersection currently feasible could now fail the warrant. Such intersections remain in the feasible set.

Signalizing an Intersection

The act of signalizing an intersection entails distinguishing the intersection from being stopped controlled as defined by the Highway Capacity Manual (HCM). Procedures from the 1997 and 2000 HCM are used for calculating intersection delay. In this research, the only change to the intersection is the form of traffic control. Lane geometry or other physical features of the intersection that affects its capacity remain unchanged. Lane geometry is described in terms of the number of exclusive left lanes, number of exclusive right lanes and the number of shared or through lanes.

All new signals are given a preset cycle length (e.g., 90 seconds), without any attempt to adjust the cycle length for best performance of the intersection, as was done by Horowitz and Granato (2000). Such adjustments are easily accomplished, but were bypassed in this research for the sake of cleanliness. Phasing plans and green splits are determined by conventional traffic engineering practice, as described by Horowitz (1997). Briefly, four different phase plans can be selected by rule for pairs of opposing approaches:

[LTR, LTR] (Left-Through-Right, Left-Through-Right)
[L, L] then [LTR, LTR]
[LTR, ---] then [LTR, LTR]
[---, LTR] then [LTR, LTR]

Once the phase plan is determined, green times are set by apportioning the available cycle time (excluding lost time) across phases according to the flow ratio (i.e., the ratio of volume to saturation flow rate), as recommended by the HCM. A floor on the flow ratio is enforced in the green time calculation for through phases so that all approaches get a reasonable amount of green. Phasing plans and green splits are revised after every MSA average of the user-optimal equilibrium traffic assignment (see later discussion on MSA).

All approaches at the signalized intersection are given an arrival type of 4, defined in the HCM to mean a little better progression than random arrivals, as recommended by Horowitz and Granato (2000). No attempt was made to determine optimal progression, which is a whole field of study onto itself.

Traffic engineers often make geometric changes to intersections when introducing signals. Such geometric changes should be considered when solving the NSDP in practice. Geometric changes were not considered in the tests here to maintain a slightly cleaner algorithm, but could be easily added later.

Travel Forecasting for the NSDP

Demands

For the two case studies, demands are calculated by the traditional trip generation and trip distribution steps of the four-step model. Trip generation was performed with the procedures and transferable parameters from NCHRP Report #365 (Martin and McGuckin 1998). Trip distribution was performed with the gravity model, using an exponential friction-factor function of path travel time. Trip distribution was calculated with feedback from traffic assignment with the MSA (method of successive averages) (Powell and Sheffi 1982), such that the resulting trip table was consistent with loaded path times across the network (Horowitz 1989). For the small network used for testing the strategic algorithm, it was found that 160 MSA averages were required to achieve reasonably precise user-optimal, equilibrium results. For the large network for testing the long-range algorithm, only 10 MSA averages were used to save time; more averages are preferred.

The trip table is calculated exactly once for each strategic problem using a network without any new signalized intersections. Any possible changes to the trip table that might result from signalizing one or more intersections within a given interval were ignored. In addition, mode split was ignored.

Delay Calculations for the NSDP

Since both the objective function and the greedy function depend upon delay, the means by which it is calculated is critical to the results of the algorithms. If the selection of signals is to be reasonable, then the delay calculations must accurately predict the conditions on the network. Consequently, this research closely followed the operational analysis procedures in the HCM or adopted other similar quality methods.

Delay along uncontrolled portions of roads were calculated with the BPR travel-time/volume function with parameters recommended in NCHRP Report #365. Capacity along two-roads was adjusted to reflect opposing flow similarly to Chapter 8 of the 1997 HCM, as described by Horowitz (1991).

Stopped delay at signalized intersections was calculated with the operational analysis procedures from Chapter 9 of the 1997 HCM or Chapter 16 in the 2000 HCM, supplemented with the upstream filtering adjustment. These procedures included the extensive computation of saturation flow rate for permitted left turns at both multilane and single-lane approaches. The presence or absence of pedestrians was ignored. The peak hour factor was set to 1 (the maximum theoretical value), the time period length was set to 1 hour, and any queuing from prior periods was ignored. Stopped delay was taken to be 77% of control delay in the 1997 or 2000 HCM, consistent with both the 1985 and 1994 HCM editions.

Stopped delays at one-way or two-way, stop-controlled intersections were calculated using the operational analysis procedures from Chapter 17 of the 2000 HCM . The calculations included upstream filtering, but did not include two-stage gap acceptance for divided roads and did not consider flared right turns.

Stopped delay at all-way, stop-controlled intersections was calculated from an M/G/1 queuing model (Richardson 1987), which contained an iterative solution for service times (and therefore, capacity) that is similar to the 1997 HCM (Horowitz 1993). Delay was calculated from the theoretical queuing equation, rather than from the empirical equations found in either the 1994 or 1997 HCM. All parameters were adjusted to closely conform to the empirical all-way stop model from the 1994 HCM.

Acceleration delay for all three types of intersections was calculated from traffic speeds using elementary principles of kinematics. Thus, a total delay at an intersection approach consisted of the sum of stopped delay and acceleration delay.

Formal Description of the Greedy Function

The objective function of the network signal design problem, as illustrated here, is the total vehicle hours travel (VHT) for the entire network. However, it takes a prohibitive amount of computation to precisely determine the savings in the VHT over the entire network, if a given unsignalized intersection were to be signalized. As a proxy of the VHT objective function, a greedy function (GF) is used to calculate savings in VHT locally, with corrections based on information from earlier steps. Thus,

$$GF_i = \left(VHT_{ui} - VHT_{si}^{u}\right)\left(1+\varepsilon_i\right)+\eta_i$$

where

VHT_{ui} = unsignalized intersection *i* VHT,

VHT_{si}^{u} = signalized intersection *i* VHT, given the unsignalized intersection *i* volume,

ε_i = the arc elasticity of intersection *i*'s volume with respect to total intersection delay,

η_i = a correction factor for network-wide VHT, calculated from earlier trials, for intersection *i*.

It should be observed that two complete traffic assignments are required for computing each set of ε_i.and η_i, which is computationally intensive, at the beginning of each time period. These values can subsequently be updated for any intersection that fails to be admitted into the solution using the results of the traffic assignment that either verified or disqualified that intersection. The quality of the greedy function can be measured by the stability of the η_i, throughout a time period.

Developing a good greedy function for a given class of problem is a trial and error processes. Simpler formulations were tried and found to be wanting. We are not asserting that the GF is the best or most efficient, just that it tends to work well on a wide variety of traffic situations on the networks adopted for this research.

The two sets of constants, ε_i.and η_i, are updated infrequently and only when an intersection is rejected (indicating a mistake has been made in its selection), so they can become old and less accurate as the strategic problem progresses. Although the use of elasticity in the greedy function can be derived from fundamental principals, its use here should be considered an arbitrary choice in the design of the algorithm stemming from the selection of minimizing VHT as the objective. After an intersection is rejected the elasticity is calculated first and then η_i is calculated to absorb any remaining error; thus the greedy function is exact at this particular moment for this one intersection. At this point in the algorithm there are two complete, almost identical, traffic simulations, with and without a signal at one intersection. Each of the two simulations yields the total network VHT (*VHTN*), the VHT for the intersection by itself and the total volume for the intersection. The arc elasticity is calculated, as described in standard economics textbooks, from the pair of VHTs for the intersection and the pair of total volumes. Then η_i is simply:

$$\eta_i = (VHTN_{ui} - VHTN_{si}) - (VHT_{ui} - VHT_{si}^{u})(1 + \varepsilon_i)$$

where,

$VHTN_{ui}$ = network wide VHT when intersection i is unsignalized, and

$VHTN_{si}$ = network wide VHT when intersection i is signalized (network is unchanged otherwise).

The need to calculate a set of correction factors would apply to any objective, not just VHT used here.

Additional Notation

The discussion of the example NSDP runs presented later involves these variables.

S = the current set of signalized intersections,

r = the allowable consecutive number of failed intersections (a value prespecified by the user),

I = the current set of eligible unsignalized intersections.

Evaluation of the Strategic NSDP with a Small Test Network

The strategic NSDP is tested primarily for accuracy on a variation of the five-zone UTOWN network. The UTOWN network was selected because it was small and has served well in many earlier tests of traffic assignment algorithms. The UTOWN network was developed for demonstrating UTPS, an old-style forecasting package authored by the US Department of Transportation about 30 years ago. It was modified for this research by splitting freeway links (formerly two-way streets) into one-way pairs and adding on and off ramps. In addition, link and node attributes were expanded to accommodate data necessary to evaluate signal and stop sign delay. Demographic characteristics of zones were left unchanged.

Test Problem

The UTOWN network, as modified, has eight intersections that were stop controlled. All remaining intersections were signalized throughout the tests. Eight intersections were considered enough to properly evaluate the strategic NSDP without making the search for the true optimal solution impossibly difficult.

The eight stop-controlled intersections were purposefully concentrated near the center of the network where traffic volumes could become sufficiently large to justify a signal. Four of the stop-controlled intersections were at the end of freeway off-ramps and the remaining intersections were two-way stop-controlled intersections between surface arterials. These intersections are illustrated in Figure 4. Figure 4 also shows the intersection identification numbers, which will be used later to distinguish between solution sets of newly signalized intersections, *S*.

The original UTOWN network would be considered severely congested at many locations, but less congested at other locations. Three different levels of demand on the UTOWN network were tested: original; 80% of original; and 120% of original. As expected, each level of demand produced distinctly different solutions to the strategic NSDP.

Because of the need to compare the solution of the heuristic algorithm with a true optimal solution, it was necessary to exactly know the list of candidate intersections prior to the start of calculations. Thus, the checks of Warrant #3a and the lower-bound threshold of GF were bypassed for all UTOWN runs.

Obtaining True Optimal Solutions

The true optimal solution for the strategic NSDP was found by calculating the network-wide VHT for every one of the 256 possible combinations for the signalized set, *S*. The true optimal solutions are presented on Table 1.

It is interesting to note that there does not seem to be a simple relationship between the level of demand and the number of signalized intersections. Intersections 35 and 38 were part of the optimal set with the original demands, but were dropped from the optimal set when demands increased to 120%. The value of VHT in the 120% network with the solution from the original network was 10442.1 or about 27 hours worse than optimal. The smaller number of signalized intersections in the 120% optimization are paradoxical and stem from the simplistic nature of the UTOWN network and the choice of VHT as an objective rather than from any issues with the optimization algorithm itself. Similar counterintuitive behavior was not observed in the Cedar Rapid network (see later section).

Example Strategic NSDP

The most interesting of the UTOWN runs involved a network with 100% demand and $r = 2$, producing a tree of success and failures as shown in Figure 5. The algorithm starts with $I = \{8, 24, 25, 35, 37, 38, 71, 72\}$ and VHT = 8722.1. As shown in Figure 5, intersection 25 is signalized, followed by the failure of intersection 24. Then, intersection 37 is signalized followed by the failure of intersection 8. Before two intersections, 72 and 71, fail to be signalized in a row, intersections 35 and 38 are added. At this point branch 0 ends with VHT = 8633.4.

The branch 1 starts with $S = \{25, 37, 35\}$, $I = \{8, 24, 71, 72\}$, and VHT = 8638.0. Intersection 71 fails to be signalized followed by the signalization of intersection 72. Finally, intersections 71 and 24 fail in a row to be signalized. At this point branch 1 ends with $S = \{25, 37, 35, 72\}$ with VHT = 8635.2. Branch 2 starts with $S = \{25, 37\}$, $I = \{8, 24, 71, 72\}$, and VHT = 8647.5. Two intersections 72 and 71 fail to be signalized and branch 2 ends. Branch 0 gives the best solution with $S = \{25, 37, 35, 38\}$ and VHT = 8633.4, which is the true optimum (see Table 1).

Tests with 80% and 120% of original demand and with r = 2 also found true optimal solutions.

Evaluation of the Long-Range NSDP with a Full-Size Network

Although, the algorithm was demonstrated to be accurate on a small test network, it was still necessary to determine if the algorithm was practical for long-range analysis in a real city of meaningful size.

Cedar Rapids Case Study

Testing of the long-range NSDP was done on the full-sized planning network for Cedar Rapids, Iowa, and it provides considerable challenges for the NSDP. This is the same network used by Horowitz and Granato (2000). Initially, the network contained about 230 signalized intersections, about 80 all-way, stop-controlled intersections, and about 300 two-way, stop-controlled intersections. The network had almost 2600 nodes, about 3800 links (one-way or two-way), 482 internal zones, and 55 external stations.

The base year for the demands on the network was 1994. The final year was 2030, giving 10 analysis years (or time periods) separated by 4 years each. Except for the addition of signals, the network was held constant throughout all 10 time periods. Demands were linearly increased each analysis year, as described by Horowitz and Granato (2000).

A single MSA average on this network took about 1.5 minutes on a 500 Mhz personal computer. Given the number of time periods and the number of initial unsignalized intersections, it was determined that execution times could become unreasonably long if the number of MSA averages for each traffic assignment were set comfortably high for excellent precision of assigned volumes. Instead, each traffic assignment was limited to just 10 MSA averages consisting of 11 all-or-nothing traffic assignments. Even with this limit, execution time for each long-range NSDP was measured in *days*.

The large number of unsignalized intersections prohibited the calculation of a true optimal solution for any strategic subproblem. Unlike the UTOWN tests, Warrant #3a and the GF threshold checks were imposed.

Application of the Long-Range NSDP

Figure 6 shows the outcome of the execution of the long-range NSDP algorithm with r = 2 for the first three time intervals; namely, time intervals 0, 1 and 2.

Carrying the algorithm through all time period resulted in 40 new signalized intersections, as given by the set:

$$S = \left\{ \begin{array}{l} 23, 25, 110, 171, 206, 213, 341, 344, 428, 578, 664, 765, 833, 851, 891, 946, 1060, 1068, 1147, \\ 1165, 1172, 1237, 1351, 1377, 1632, 1789, 1847, 1849, 1861, 2003, 2005, 2046, 2055, 2064, \\ 2118, 2349, 2432, 2440, 2475, 2502 \end{array} \right\}.$$

The same algorithm was applied to Cedar Rapids network with $r = 3$. Figure 7 presents a comparison of cumulative number of intervals that were signalized when $r = 2$ and $r = 3$. As expected, each time interval had accumulated a greater number of signalized intersections for $r = 3$ than $r = 2$.

Table 2 presents a comparison of VHT for each time interval before and after the application of the strategic NSDP algorithm for $r = 2$ and $r = 3$. It is seen that an improvement of 106.6 hours is achieved by changing the value of r from 2 to 3.

The larger value of r increases the number of signals; however, the gain in VHT is marginal – about 18 hours for each signal. No conclusion can be drawn as to which value of r starts with a better VHT at the beginning of each time interval. For example, intervals 4 and 6 start and end with a better value of VHT for $r = 2$ than for $r = 3$. Intervals 5 and 8 start with a better value of VHT for $r = 3$ than for $r = 2$; however, they both end with a better value of VHT for $r = 2$ than that of $r = 3$.

When $r = 2$ the algorithm required 86 traffic simulations. The number of traffic simulations increased to 101 when $r = 3$.

A Brief Comparison to Simulation Results

Horowitz and Granto (2000) attempted to solve the same problem by a series of rules, with the one optimization element of adjusting cycle lengths. The network and demands were identical to this one, but delay estimation procedures had been taken primarily from the 1994 HCM (thereby yielding slightly different VHT estimates for the same signal configuration) and left turn bays were added along with any signal to any intersection that did not previously have them. In any given time interval intersections were considered for signals in the order that they most violated Warrant 3a, and an individual signal was always kept if total VHT for the network was lowered by its installation.

It is interesting to note that the number of signals found optimally, even with r=3, is far less than the 121 signals deployed in the study by Horowitz and Granto (2000). In the present study, VHT increased from 1994 to 2030 by just 141% (11517 to 27765) compared to 182% for the previous study. Thus, the optimization provided a substantially better level of VHT, in spite of the considerable advantages the rule-based method had by adding left turn lanes and by setting optimal cycle lengths.

Conclusions

The network signal design problem, which asks to find the optimal set of signalized intersections in a network, is so large and so complex that heuristic methods must be employed for any meaningfully-sized traffic network. Two variations of the problem, strategic and long-range, seek solutions that are valid over differing lengths of time. The long-range NSDP has become important to transportation planners because of recent advances in travel forecasting methods.

The *strategic* network signal design problem, where VHT is minimized within a single time period, can be approximately solved with a constructive greedy search, where the greedy function has terms for the expected improvement in VHT at the intersections with constant volumes, the elasticity of volume with respect to delay, and VHT effects on neighboring intersections. The selected algorithm found exact optimal solutions on a small test network with 8 candidate intersections and found at least good solutions on a full-sized network with approximately 380 candidate intersections.

It was necessary to modify a traditional constructive greedy search in order to overcome the possibility of certain intersections blocking other intersections from entering the solution set. This modification involved "branches" where the algorithm is restarted while prohibiting certain intersections (already considered acceptable) from entering the solution.

The cost of a single evaluation of the objective function is very high, so it is necessary to pay close attention to the amount of evaluations. The heuristic algorithm described requires a much smaller number of objective function evaluations than an exhaustive search.

The *long-range* network signal design problem is affected by the tradition of grandfathering signalized intersections into later time intervals. Tests of on both networks suggest that grandfathering produces VHT values that are slightly higher than would be obtained if all intersections competed fairly for a place in the solution set. However, grandfathering is realistic and greatly reduces the computational effort.

During early stages of testing the algorithm, a decision was made to keep the test problems as clean as possible without compromising realism. For example, it might have been possible to do any of the following: introduce turn lanes or make other geometric improvements when signalizing an intersection; recalculate the trip table prior to each traffic assignment; consider optimized signal coordination; introduce additional MUTCD warrants; better accounting for the effects of signal coordination; or use a more complex objective function (other than VHT) that better approximates the needs of a traffic agency. It is likely that the algorithm would have still performed well with these complications. However, additional tests would be necessary to fully understand the algorithms limits in terms of problem complexity and the effects on signal deployment from an increased richness in the available set of traffic engineering remedies.

The optimization method of this paper results in far fewer signalized intersections than the rule-based method of a previous study. Although the optimization method produces a much more efficient allocation of signals and a much lower value of VHT than rules-based method, it is unclear which method would be preferred for long-range travel forecasting.

References

1. Chen, Migyuan and Alfa, Attahiru Sule (1991). “Network Design Algorithm Using a Stochastic Incremental Traffic Assignment Approach.” *Transportation Science*, 25(August), 215-224.
2. *Highway Capacity Manual* (2000). Transportation Research Board, Washington, DC.
3. Horowitz, Alan J. (1989). "Convergence Properties of Some Iterative Traffic Assignment Algorithms." *Transportation Research Record*, No. 1220, pp. 21-27.
4. Horowitz, Alan J. (1991). "Integrated Traffic Flow and Region-Wide Travel Forecasting." Proceedings of the 61st Annual Meeting of the Institute of Transportation Engineers.
5. Horowitz, Alan J. (1992). "Implementing Travel Forecasting with Traffic Operational Strategies." *Transportation Research Record*, 1365, 54-61.
6. Horowitz, Alan J. (1993). "A Revised Queuing Model of Delay at All-Way Stop Controlled Intersections." *Transportation Research Record*, 1398, 49-53.
7. Horowitz, Alan J. (1997). “Intersection Delay in Regionwide Traffic Assignment: Implications of the 1994 Update of the Highway Capacity Manual.” *Transportation Research Record*, 1572, 1-8.
8. Horowitz, Alan J. and Granato, Sam. (2000). “Selection of a Traffic Control Strategy for a Long-Range Traffic Forecast.” *Transportation Research Record Journal*, 1706, 145-151.
9. *Manual on Uniform Traffic Control Devices* (2001), Millennium Edition, Federal Highway Administration and Institute of Transportation Engineers.
10. Martin, William and McGuckin, Nancy A. (1998). *Travel Estimation Techniques for Urban Planning*, NCHRP Report #365.
11. Meneguzzer, Claudio (1997). “Review of Models Combining Traffic Assignment and Signal Control.” (ASCE) *Journal of Transportation Engineering*, 123(2), 148-155.
12. Powell, W. B. and Sheffi, Y. (1982). “The Convergence of Equilibrium Algorithms and Predetermined Step Sizes.” *Transportation Science*, 16(1), 45-55.
13. Richardson, Anthony J. (1987). "A Delay Model for Multiway Stop-Sign Intersections." *Transportation Research Record*, 1112, 107-112.

14. Solanki, Rajendra S., Gorti, Jyothi K., and Southworth, Frank (1998). “Using Decomposition in Large-Scale Highway Network Design with Quasi-Optimization Heuristic.” *Transportation Research B*, 32B(February), 127-140.
15. Zanakis, S. H., Evans, J. R., and Vazacopoulos, A. A. (1989). "Heuristic Methods and Applications: A Categorized Survey." *European Journal of Operations Research*, 43, 88-110.

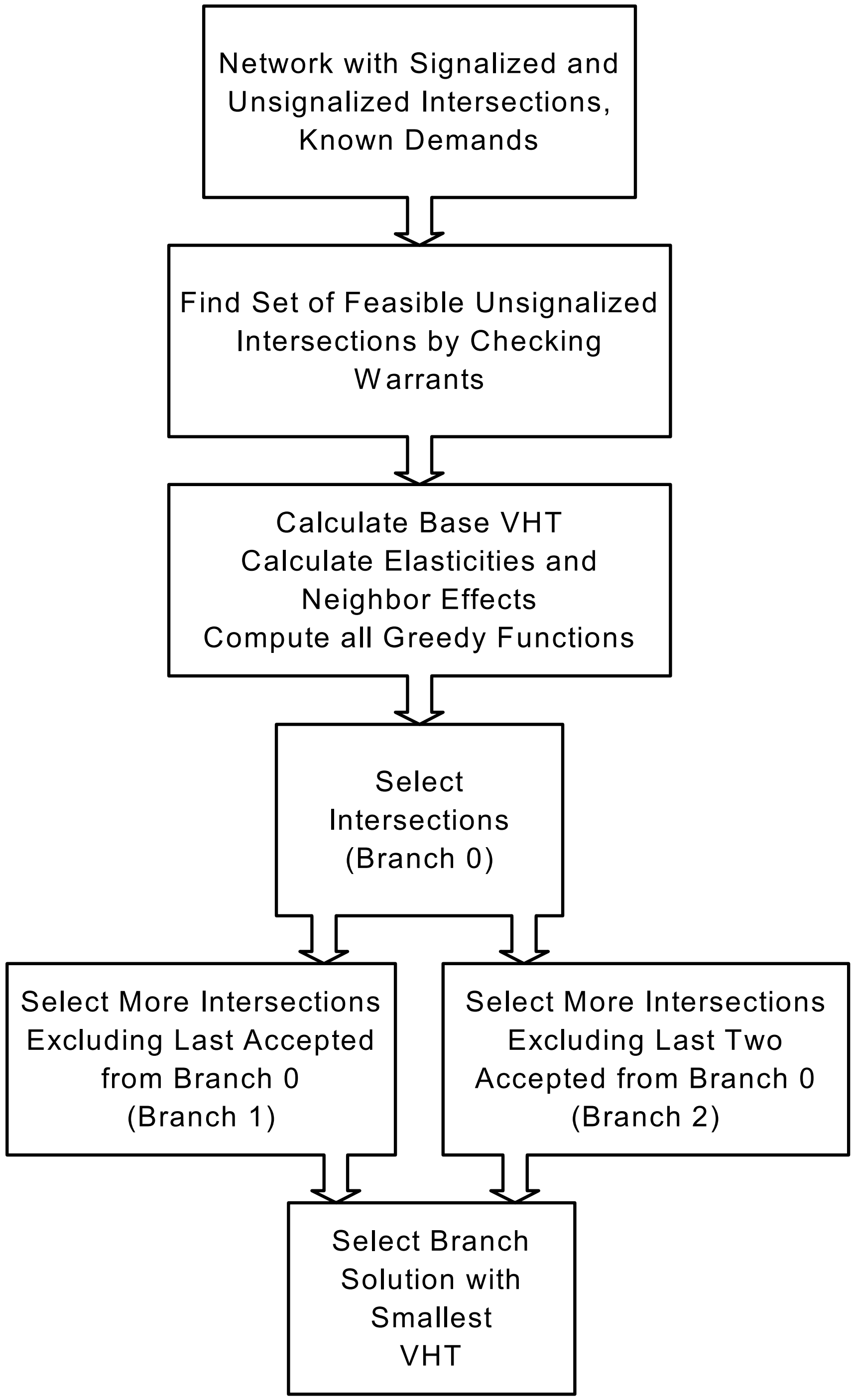

Figure 1. Overall Flow of the Greedy Search in the Strategic NSDP

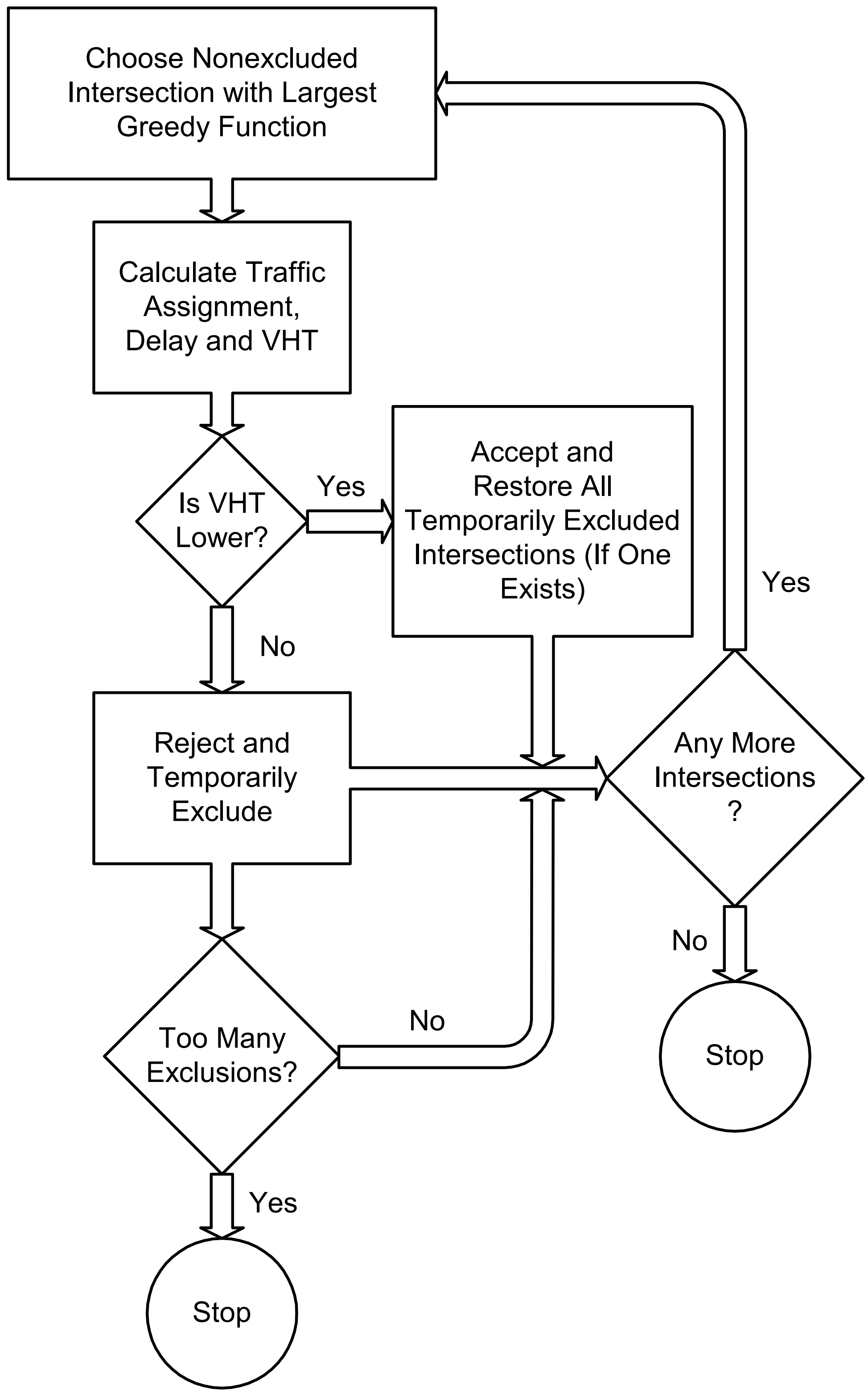

Figure 2. Selecting an Intersection

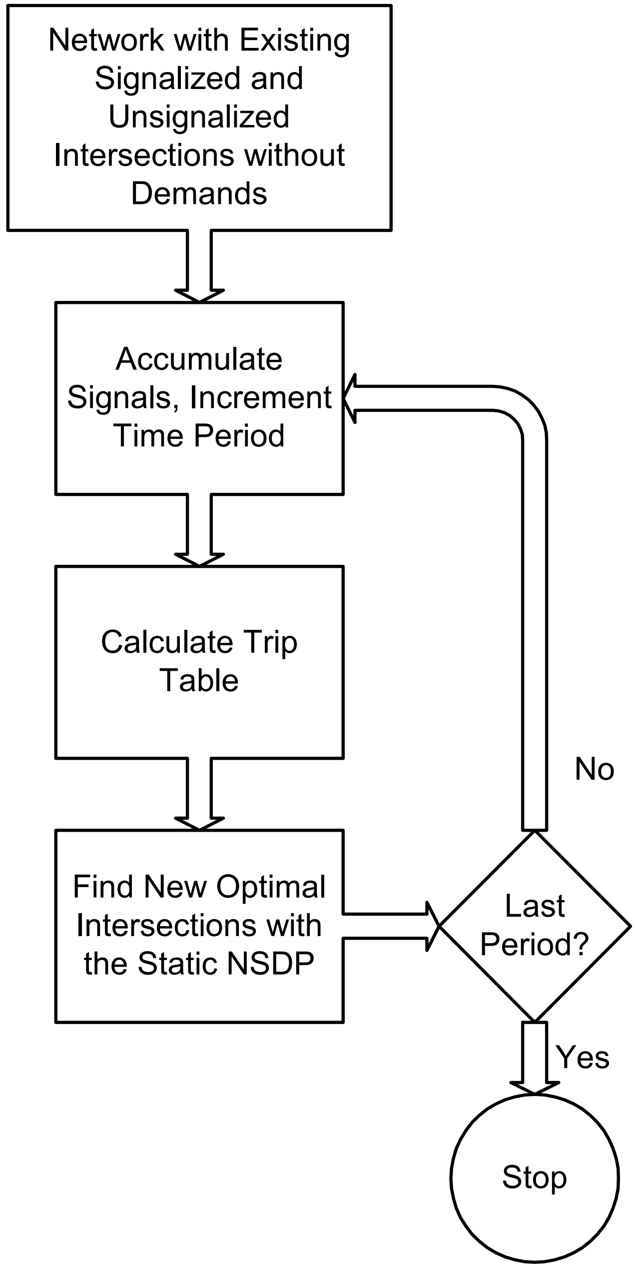

Figure 3. Heuristic Algorithm for the Long-Range NSDP

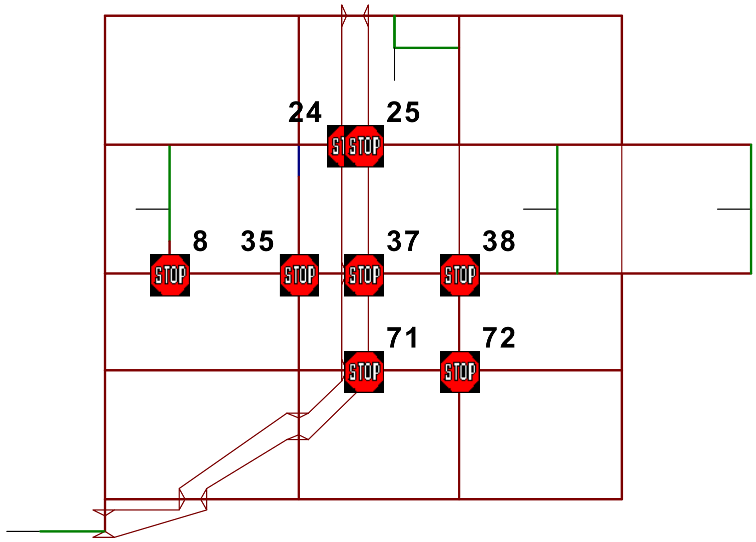

Figure 4. UTOWN Test Network and Stop-Controlled Intersections

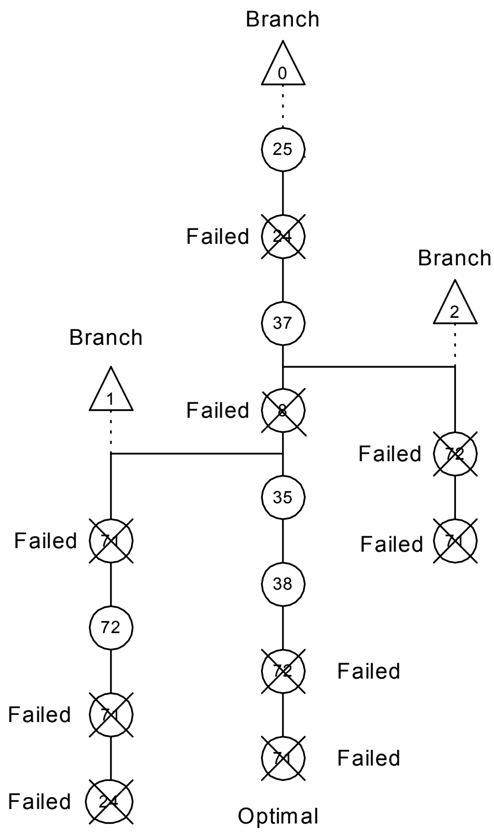

Figure 5. Strategic Algorithm Applied to UTOWN with 100% Demand

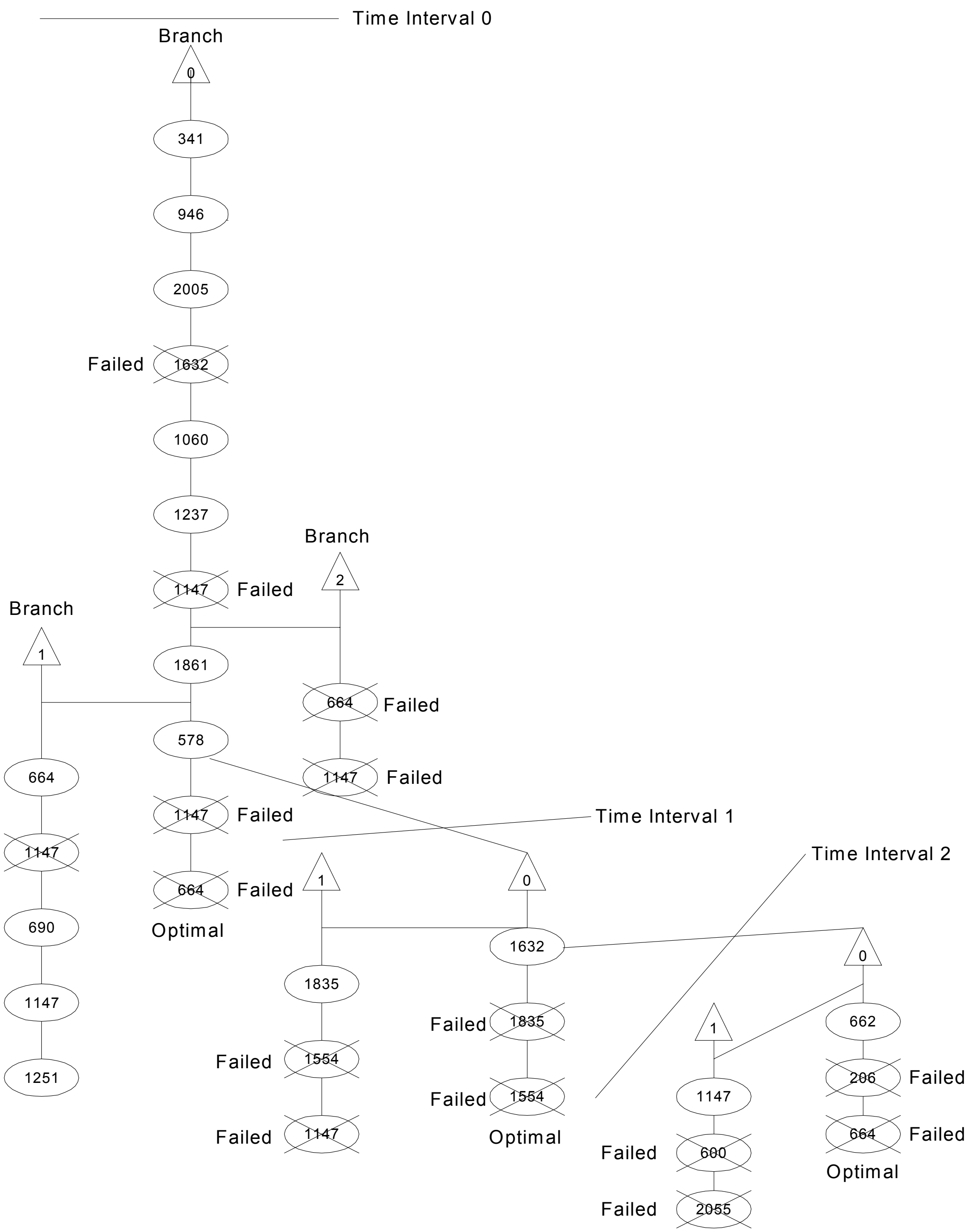

Figure 6. Partial Tree Depicting the Outcome of the Long-Range NSDP Algorithm Applied to the First Three Time Intervals

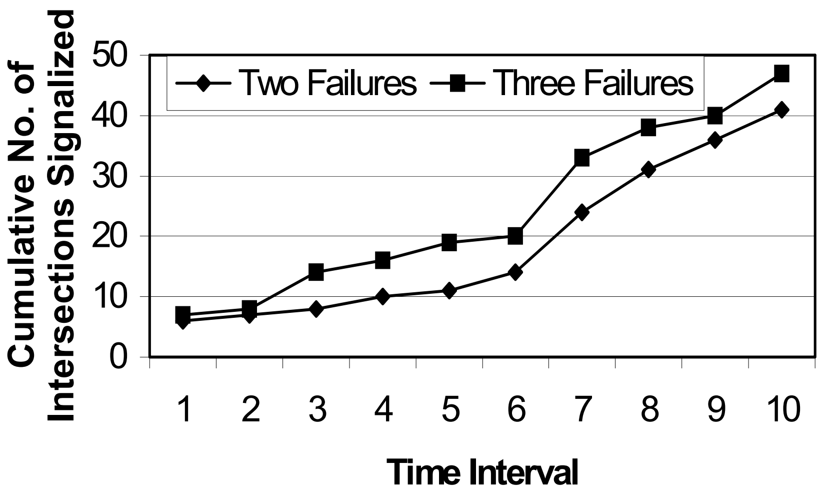

Figure 7. Comparison of Cumulative Number of Signalized Intersections for $r = 2$ and $r = 3$

Table 1. Optimal Solutions on the UTOWN Network with Three Demand Scenarios

Level of Demand	Optimal Set, S	VHT with $S = \phi$	Optimal VHT
80% of Original	35,37	7175.8	7165.5
Original	25,35,37,38	8722.1	8633.4
120% of Original	25,37	10537.9	10415.0

Table 2. Comparison of VHT Between r = 2 and r = 3

Time Interval		*r* = 2 VHT		*r* = 3 VHT	
		$S = \phi$	S^* (optimal)	$S = \phi$	S^* (optimal)
0	1994	11627.9	11525.7	11627.9	11516.9
1	1998	12953.7	12952.0	12949.3	12941.2
2	2002	14425.2	14403.4	14420.9	14381.5
3	2006	15973.9	15918.4	15952.6	15911.9
4	2010	17648.1	17608.2	17663.9	17627.1
5	2014	19502.8	19316.0	19475.4	19335.9
6	2018	21340.7	21134.7	21497.2	21166.3
7	2022	23561.5	23284.7	23310.0	23220.8
8	2026	25633.0	25460.4	25552.7	25517.8
9	2030	28173.7	27871.3	28910.2	27764.7