

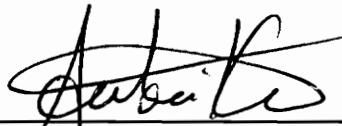
**Heuristic Network Generator - An Expert Systems Approach for
Selection of Alternate Routes during Incident Conditions**

by

Vijay Krishnaswamy

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Master of Science
in
Civil Engineering

Approved:



Dr. A.G. Hobeika, Chairman



Dr. Donald R. Drew



Dr. R. Sivanandan

June 23rd, 1994
Blacksburg, Virginia

C.2

LD
5055
V855
1994
K737
C.2

A HEURISTIC NETWORK GENERATOR - AN EXPERT SYSTEMS APPROACH FOR SELECTION OF DIVERSION ROUTES DURING INCIDENT CONDITIONS

by

Vijay Krishnaswamy

Dr. A.G. Hobeika

(Chairman)

(Abstract)

Congestion on the freeways of the U.S. has increased multifold over the past few years. A significant portion of this congestion is caused by non-recurring events such as incidents. Diversion has been accepted as a method that can reduce delays during incidents.

The process of diversion involves the selection of the alternate routes, which is currently done off-line and is not responsive to each incident case. The volumes on these preselected routes on that particular day are also ignored. The preselected routes, in most cases, serve only to bypass the link on which the incident occurs. Considering the volumes that flow on the freeways, this leads to considerable delays in terms of lost time and productivity. Another important issue that is currently neglected is user compliance.

The network generator is used to reduce the delays in selection of these alternate routes. It uses characteristics such as the congestion levels and available capacities in selection of alternate routes in real-time. Also, used in selecting alternate routes are feasibility criteria, that significantly affect the available capacities on the links. These include presence of trip generators (schools, offices, etc.) or safety factors (icy bridges, height restrictions, etc.). The model thus generates a reduced network and a set of alternate routes to divert the traffic upstream of the incident. Disutilities that drivers associate with route-choice, such as the number of left-turns and signals, the relative time spent on the freeway and arterials are attached to each route. The routes with the minimum disutilities are displayed to the user. A user-equilibrium assignment module to predict traffic flows in the future is also incorporated into the framework. As a precursor to the network generator, there is a module which calculates the clearance time for an incident. It uses other characteristics of

the incident such as the weather and time of occurrence in order to predict if the delays are significant to initiate diversion.

Numerous tests were conducted in order to validate the rules and functions developed. The tests were based on varying incident and traffic conditions.

The results showed that the model, was able to select better routes for off-peak conditions rather than peak conditions. There is a threshold value of the delay caused by the incident, beyond which the model is very effective.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. A.G. Hobeika, chairman of my committee, for his effort devoted to this research effort, and for his financial support, during my involvement in this research. I would also like to thank Dr. Donald Drew and Dr. R. Sivanandan for being my committee members.

In addition, Mr. Kaan Ozbay deserves special mention for his guidance and helpful comments in reviewing and refining this work. I would also like to thank a number of special friends and colleagues at the Center, who have helped me during my stay at Virginia Tech.

This thesis is incomplete without mention of my mother, Mrs. Mallika Krishnaswamy, who provided moral and intellectual support throughout this effort. She has always been the epitome of strength and courage and has set an example throughout my life. I would like to dedicate this thesis to her.

Table of Contents

1.0 INTRODUCTION	1
1.1 The Congestion Problem	1
1.2 Freeway Incident Management (FIM)	1
1.3 Background	2
1.3.1 Framework	4
1.4 Analysis Framework for Network Generator	5
1.5 Organization of Report	7
2.0 LITERATURE REVIEW	8
2.1 Introduction	8
2.2 Network Aggregation Models	8
2.3 Expert Systems for Route Guidance	14
2.4 Real-Time Diversion And Routing	18
2.5 Prediction Models	21
2.6 User Route Choice Models	24
2.7 Conclusions	26
3.0 DEVELOPMENT OF MODEL & ANALYTICAL FRAMEWORK	27
3.1 Introduction	27
3.2 Freeway Incident Management and Diversion Strategies	28
3.2.1 Problems with Current Practices in Diversion	28
3.3 Requirements for the Analytical Framework of the Network Generator	30
3.4 Knowledge Acquisition Process	31
3.5 Justification for Using a Knowledge Based System	33
3.5.1 Deficiencies of Analytical Models Used for Diversion	33
3.5.2 Advantages of Knowledge-Based Models	34

3.6 KAPPA-PC Expert System Shell	36
3.7 Problem Representation	38
3.8 Analytical Framework for the Model	40
3.9 Conclusions	44
4.0 IMPLEMENTATION DETAILS OF THE MODEL	45
4.1 Introduction	45
4.2 Knowledge Representation	45
4.2.1 Incident Types and Duration	46
4.2.1.1 Decision Tree for Collisions	47
4.2.1.2 Decision Tree for Fatalities	47
4.2.1.3 Special Events	48
4.2.1.4 Spilled Loads	48
4.2.1.4 Other Effects on Incident Duration	48
4.2.1.5 User Interface for Clearance Time Module	50
4.2.1.6 Example	50
4.2.2 Incident Delay Calculations	52
4.2.2.1 Capacity Reduction on Incident Link	52
4.2.2.2 Delay Computation based on Deterministic Queuing	53
4.2.2.3 Diversion Initiation	55
4.2.3 Network Generator	56
4.2.3.1 Link Elimination	56
4.2.3.2 Route Generation	58
4.2.3.3 Prioritization	63
4.2.3.4 CMS Update and Signal Timings	64
4.2.4 Assignment Model	64
4.2.4.1 Need for an Assignment Module	66
4.2.4.2 Linear Programming Formulation of the Algorithm	66
4.2.4.3 Algorithm for Computer Implementation	69
4.2.4.4 Bisection Algorithm	69
4.2.4.5 Shortest Path Algorithm	70
4.3 Conclusions	74
5.0 RESULTS AND ANALYSIS	75
5.1 Introduction	75
5.2 Network Description	75
5.3 Inputs to the model	76
5.4 Outputs from the Model	78

5.5 Test Cases	78
5.5.1 Scenario 1	79
5.5.2 Scenario 2	80
5.5.2.1 Case 1	80
5.5.2.2 Case 2	82
5.5.2.3 Assignment and Delay Calculations for Scenario 2	84
5.5.3 Scenario 3	85
5.5.3.1 Case 1	85
5.5.3.2 Case 2	86
5.5.3.3 Case 3	88
5.5.3.4 Assignment and Delay Calculations for Scenario 3	89
5.5.3.5 Case 4	90
5.6 Conclusions	93
6.0 CONCLUSIONS AND RECOMMENDATIONS	94
6.1 Conclusions	94
6.2 Recommendations	95
BIBLIOGRAPHY	97

List of Illustrations

Figure 1.1	Framework for Real-Time Diversion Strategies	3
Figure 3.1	Sample Map used for Diversion in NOVA	29
Figure 3.2	Analytical Framework for the Network Generator	41
Figure 4.1	Clearance Time for Spilled Loads	49
Figure 4.2	Incident Input Data Screen	51
Figure 4.3	Delay Calculations	54
Figure 4.4	Diversion Screen and Eliminated Links	59
Figure 4.5	The Search Algorithm	60
Figure 4.6	Route Generation and Display	62
Figure 4.7	CMS Display and Signal Retiming	65
Figure 4.8	Bisection Algorithm	71
Figure 4.9	Shortest Path Algorithm	73
Figure 5.1	Test Network	77
Figure 5.2	Delay Characteristics (Peak Hour)	91
Figure 5.3	Delay Characteristics (Non-Peak Hour)	92

List of Tables

Table 1.1	Sketch Planning vs. Regional Assignment	10
Table 3.1	List of Variables Associated with Route Choice & their Values	35
Table 4.1	Clearance Time based on Incident Type	47
Table 4.2	Fraction of Freeway section Capacity available under Incident Conditions	53
Table 5.1	Links that are not usable for Diversion (Scenario 2; Case-1)	81
Table 5.2	Links that are not usable for Diversion (Scenario 2; Case-2)	83
Table 5.3	Delay Values (in vehicle-hours) for Scenario 2	84
Table 5.4	Links that are not usable for Diversion (Scenario 3; Case-1, Period 1)	85
Table 5.5	Links that are not usable for Diversion (Scenario 2; Case-1, Period 2)	87
Table 5.6	Delay Values (in vehicle-hours) for Scenario 2 (Period 1)	89
Table 5.7	Delay Values (in vehicle-hours) for Scenario 2 (Period 2)	89

1.0 INTRODUCTION

1.1 The Congestion Problem

Urban freeway congestion is a serious and growing national problem; one that is receiving greater attention from transportation engineers, planners and local and state officials. The important reasons for this are:

- Increase in demand due to increase in number of cars,
- Falling vehicle occupancy,
- Increasing use of freeways over arterials.

This has resulted in two types of congestion:

- Recurrent, and
- Non-recurrent congestion.

Recurrent congestion is said to occur, when the normal demand exceeds the supply for the roadway system. This usually occurs during the peak hour, in all major urban areas. There are several ways for alleviating this, such as widening of roadway, ramp metering, high-occupancy vehicle facilities and low-cost geometric improvements.

Non-recurrent congestion occurs due to the demand exceeding the capacity, due to sudden disruptions in the traffic flow pattern. This could arise from incidents, freeway construction and maintenance activities, or due to natural disasters. The focus in this study has been on freeway incidents and their management.

1.2 Freeway Incident Management (FIM)

It is estimated that incidents on urban freeways cost about \$100 billion each year in direct costs. Approximately 60 billion dollars is lost due to congestion caused by incidents every year. This cost does not include the cost of excess fuel consumption and the damage to the environment. It is clear that incidents on the nation's freeways pose a significant

problem in maintaining a high level of mobility in urban areas. During peak hours, even small savings in incident clearance time can effect significant reductions in the costs of traffic congestion. Thus, the issue of managing these incidents deserves appropriate attention and this concern is reflected in the growing interest in Freeway Incident Management (**FIM**) nationwide. It is part of Advanced Traffic Management Systems (**ATMS**), an area of IVHS that has been ear-marked for early deployment.

Given a choice, one would like to prevent all incidents occurring on freeways. Unfortunately, this cannot be accomplished. However, their impact on traffic flow can be mitigated through proper incident management. Freeway management views the highway system as a resource which has to be utilized efficiently to maximize throughput. The following can be stated as the broad goals of incident management:

- reducing the time for incident detection and verification,
- reducing the response time,
- exercising proper on-scene management,
- providing to the public timely accurate information on traffic conditions and rerouting.

Thus, it can be seen that **Freeway traffic diversion** is considered an integral part of incident management, and is accepted as an effective method for alleviating delays to motorists during incidents.

1.3 Background

Driver Information Systems and diversion strategies act in tandem to reduce the strength of the shock wave that is propagated by the sudden decelerations or stops during incident conditions. It has been a topic of research at the Center for Transportation Research at Virginia Tech for the past four years. The research, being led by Dr. A. G. Hobeika (1989), is an effort to analyze incident conditions so as to develop accurate diversion and rerouting strategies. At present, an elaborate framework has been defined and several components have been developed (**Figure 1.1**).

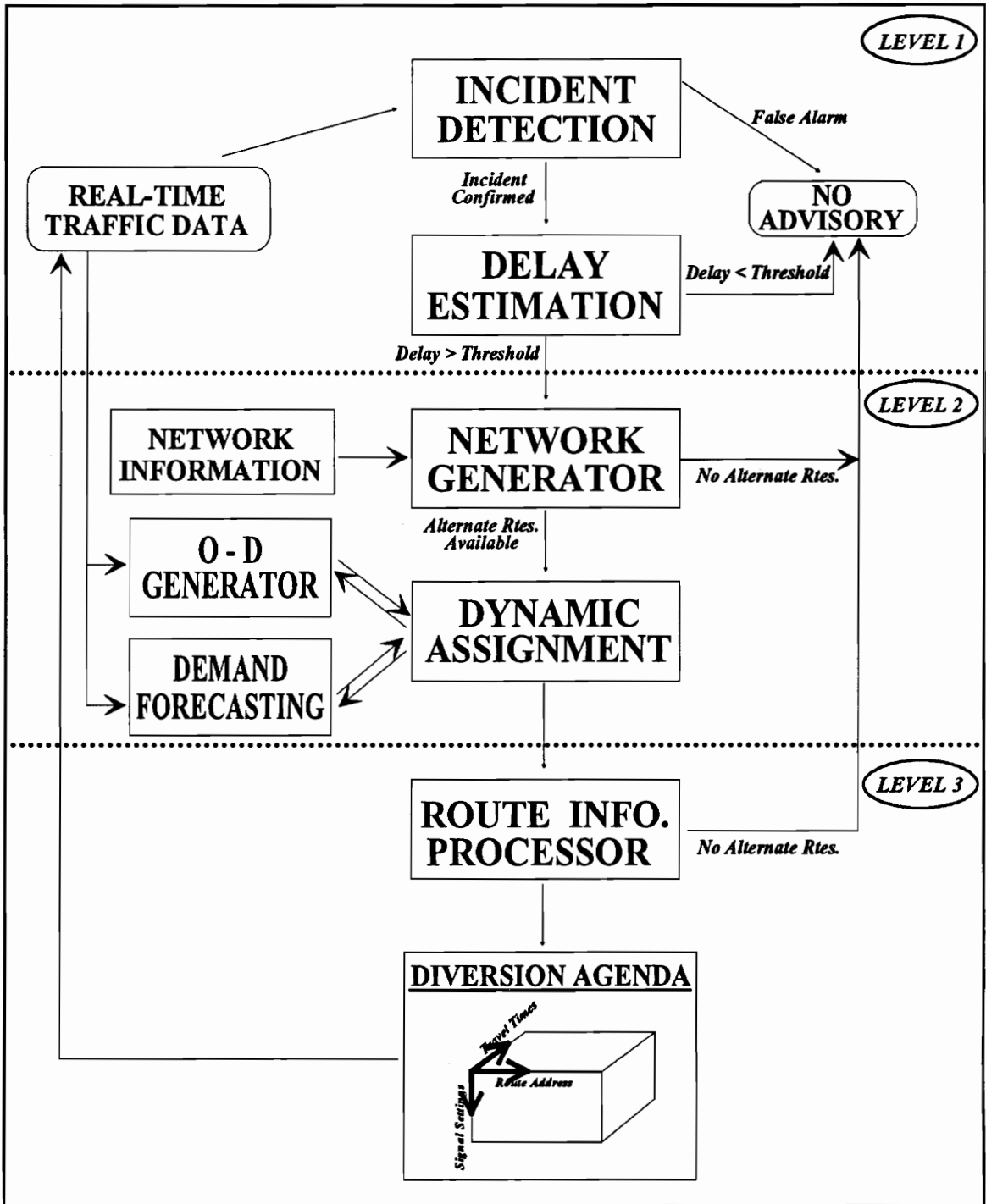


Fig. 1.1 A Multi-module Framework for Developing Real-Time Diversion Strategies

1.3.1 Overall Framework for Traffic Management

The framework consists of three main hierarchical levels. At the first level, the detection of the incident and subsequent verification together constitute the initiation protocol for the system. Subramaniam et. al. (1992) developed a model for the estimation of delays on freeways due to incidents. Depending on whether significant savings in delay time is expected, this model triggers the next stage of the route diversion process, which is to evaluate the alternate route options.

In the second level of hierarchical process, a **Dynamic Traffic Assignment (DTA)** module is employed to select time-dependent shortest paths. DTA is an improvement over static assignment techniques because it accounts for the fact that all vehicles assigned to a route are not present on the links at the same time. This module assigns drivers to the time-dependent shortest paths. Future travel times needed for identifying the time dependent shortest paths are determined by using a look ahead feature that is currently under development. The DTA module being computationally intensive, cannot consider all the possible routes from Origin to Destination. This implies that there should be a precursor to this module which would reduce the universal set of all possible routes to a smaller subset, which the DTA module would then use for its input. This process is known as the network abstraction process and the module which precedes the DTA is called the Network Generator. This module will reduce the computational complexity for the DTA, but will also introduce new problems with regards to connectivity and completeness of the reduced network.

Although, abstraction of the network is expected to be one of the main uses of the network generator in the future, it is not the focus of this thesis. The network generator can also be used to find a set of feasible routes for diversion (after network reduction) under incident conditions. This new context for the network generator is explained in detail in the next few sections. Also described below are the modules that are used in conjunction with the network generator. Any reference to the network generator in this thesis, is in association with selection of feasible diversion routes under incident conditions.

Lastly, the third level of the hierarchical process deals with the dissemination of information to motorists. Travel times on different routes at different times are provided by the second level. This travel time information is then used to devise a diversion agenda for motorists.

1.4 Analysis Framework for Network Generator

The proposed model incorporates feasibility criteria in conjunction with travel-time criteria, in reducing the network for the DTA model. For *e.g.* Some links may be congested, while some others may pose safety problems during hostile weather conditions or may lack CMS facilities. Such links are eliminated from further consideration and a reduced network is developed for the DTA. The network generator employs similar criteria to apply to the traffic during incident conditions in order to develop a reduced network. Subsequently, routes can be generated for diversion from point “a” to point “b” in order to dissipate the queues, that develop when an incident occurs.

There are a certain important considerations when designing the network generator:

1. **Selection of feasible alternate routes to divert traffic:** Different criteria based on which routes can be selected are:
 - **Operator Viewpoint:** The operator would ideally like to minimize the travel time network-wide. To achieve this goal he would be willing to penalize individual users (sometimes quite significantly). The various traffic conditions that are of interest to him would be the traffic flow and thereby the travel time on the links. The assignment of traffic onto these routes can then be used to study the delay characteristics. Since this methodology is quite time consuming, a significant amount of time needs to be saved so that the process has a high benefit-cost ratio.
 - **User (Driver) Viewpoint:** It is important to ensure user compliance when diverting traffic. If the network is subjected to haphazard flows because the driver rejects the set of routes selected for him to travel on, the network-wide delays could increase in certain cases. The literature on user-behavior has been studied extensively in the next chapter. It is found that drivers associate high disutilities with high travel times, left-turns and signalized intersections. Similarly, unsafe

neighborhoods are also a cause for great concern among drivers. The routes selected have to account for the user preferences in some manner. This will help assuage some of the driver's fears about diversion.

- **Prediction Model:** A prediction model is necessary so that the diversion process can look ahead in time to foresee any problems caused due to the diversion. This would in effect provide a self-correcting mechanism for the diversion module. Some of the prediction modules were studied in the literature review section and their applicability to the model currently under development was studied. Ideally, a DTA module can be used, but since one has not been developed yet, a static user-equilibrium formulation was finally used in order to predict the future traffic flows on the network. The convex-combination formulation was used in order to develop a quick response to the same.

Precursor Modules to the Network Generator: There are two important modules that precede the network generator:

- **Incident Duration Estimation:** is done based on decision trees that vary with the type of incident, the location, the time of occurrence and the weather conditions at the time of occurrence. The estimation is necessary, in order to determine the delays caused by the incident.
- **Incident Delay Estimation:** is done based on the incident duration (calculated from the previous module), number of lanes closed and the actual flows on the links of the network. Quantification of the delay is instrumental in determining if diversion is necessary for that particular incident. If diversion is found to be necessary, the network generator module described above is initiated. The model developed here uses a deterministic queuing algorithm to compute the delays.

Measures of Effectiveness (MOE's) to justify use of the Network Generator:

There is a need to prove that diversion strategies carried out based on the output of the network generator, helps in alleviating delays on the network. The user-equilibrium module helps in generating MOE's for the same. Therefore, results of this assignment should be compared to the case where the traffic simulation is done without any attempt at diversion. Also, the assignment technique can compare the results with the case where travelers have 100% information about the incident and the network conditions at the beginning of their trip.

1.5 Organization of Thesis

This chapter has discussed briefly the urban congestion problem and the approach to the solution at Virginia Tech. The next chapter is the Literature review which discusses the topics that have been identified as relevant to the network generator. The conclusions drawn from the literature are also discussed at the end of the chapter. The third chapter justifies and explains the development of a real-time expert system to divert the traffic, and also deals with the different issues involved in the model development. The fourth chapter discusses the implementation details of the model that has been developed based on the explanations in the previous chapter. Chapter 5 presents the results of the different analyses carried out on a test network, based on this model. Chapter 6 discusses the conclusions and recommendations for future research.

2.0 LITERATURE REVIEW

2.1 Introduction

Owing to the hybrid nature of the solution technique, the literature review involves papers from different areas, each of which pertains to a different aspect of the problem. The different areas are discussed in the order listed below:

- Network aggregation models,
- Expert systems for route guidance,
- Real-time diversion and routing
- Prediction models, and
- User route choice models.

2.2 Network Aggregation Models

The problem is first studied as one dealing with network aggregation, in which the aim is to reduce the size of the network to reduce computation costs. Incident conditions can then be incorporated as part of the network characteristics. Assignment can then be performed on this reduced network and the delays can be studied using the new computed flows. A few papers have been written on the network aggregation problem. Network aggregation is the art of condensing a given network into one that can be managed efficiently as well as preserving the desired characteristics of the original network. There are two main types of aggregation:

- **Network Element Extraction (NEE)**: This is the process of removing elements that have been identified as being insignificant, according to some pre-specified criterion, from the network. This method has the disadvantage of causing network disconnection.
- **Network Element Abstraction (NEA)**: This method collapses the insignificant network elements into pseudo or dummy elements. This is very difficult to perform. Also, the results are not easily transferable to the original network.

The first paper presented here is by Haghani et. al., (1983). Based on the reasons mentioned above, the authors decided to build a NEE model. The algorithm is based on the distribution of traffic on the network after an assignment process. There are a few links which do not carry a significant amount of traffic. These links are identified and extracted from the network and the performance of the remaining portion of the network is then studied. A link is deemed to be insignificant if it carries an equilibrium flow below a fraction α of the maximum equilibrium flow in the network. The algorithm consists of the following steps:

- Specify α . Solve the equilibrium flow problem. Let X_i be the flow on link i .
- Identify the unextracted link k , with the minimum flow and find

$$\alpha_k = X_k / \text{Max}(X_i)$$

If this is greater than α stop. Otherwise disaggregate the flow on link k , by specifying the origin destination of all the flow on link k .

- Discard link k . Declare the forenode of link k as a destination and the backnode of the link k as an origin (if they are already not).
- Update the O-D matrix.

Some of the disadvantages of this model are:

- If the number of links extracted increases, the model may produce a set of disconnected networks which adversely affects the assignment process,
- Possibility of the number of origins and destinations increasing because of the model is significant. This could increase the computation complexity of the algorithm, and
- The model is incapable of handling real-time conditions of traffic. It was developed for use in project-planning applications, where different scenarios can be studied without significant computational effort.

Eash et. al., (1983) presented a NEA methodology for the Northern Illinois network and also compared regional planning to the sketch planning approach. The characteristics used for the comparison are the total vehicle-miles, vehicle hours, and average speeds. A user equilibrium algorithm was used to compare the results of the two. One of the

obvious impacts of the abstraction was the immediate increase in the number of intra-zonal trips for the sketch planning case (because the zones were bigger). A methodology is described here to account for this increase, in which the vehicle miles for each of the two cases is studied and results of the regional assignment is adjusted according to the difference. This provides a basis for comparing the results from the two techniques, the adjusted values of which are presented below. The additional intra-zonal trips are assigned on to the same minimum time paths, in the same proportions used in the regional traffic assignment. A few of the results of the comparison are presented below:

Table 1.1: Sketch Planning vs. Regional Assignment (Source: Eash et. al., 1983)

No.	Item	Sketch Planning Assignment	Regional Assignment
1)	Network Nodes	820	12040
2)	Network Links (One-way)	2422	37065
3)	Computing Time (CPU)	3 min. 45 sec.	163 min. 7 sec.
4)	Freeway Average Speed	33.1 mph	36.6 mph
5)	Arterial Average Speed	26.8 mph	24.5 mph
6)a	Veh.-hours: Freeway	104,962 (29%)	81,446 (22%)
6)b	Arterial	261,048 (71%)	286,664 (78%)
7)a	Veh.-miles: Freeway	3,475,759 (33%)	2,981,913 (30%)
7)b	Arterial	7,000,609 (67%)	7,025,788 (70%)

Some of the general conclusions that can be drawn are:

- Different intra-zonal trips in the two assignments did not significantly affect the results.
- Assignment of traffic is more seriously affected by the coding of the arterial network.
- The method used for aggregation is not necessarily applicable in other regions of the US.
- However, the overall results compared quite well. Therefore, sketch planning assignment is probably adequate for estimating most highway travel characteristics, including operating costs, emissions and gasoline consumption.

The model is a good one and proves that network aggregation in this region of the country does not affect the assignment process significantly. The methodology used for

aggregating the network itself seems dependent heavily on the local conditions of the network. This however, has to be proved for other areas of the country.

In a similar study carried out by Bovy and Jansen (1983), the network is aggregated and the effects on the car traffic assignment module is empirically investigated. Three network models were developed for the road network of Eindhoven: a fine, medium and a coarse model. The paper presents results of the all-or-nothing and equilibrium assignments based on the sensitivity of link load estimates to the different cases.

The road system was simplified using the reduction method. This implies that the network links were selected directly from the actual road network and that the selection was based on the functional class of links. The three levels were hierarchical which means that all links present in a lower level network is also present in all higher levels. Also, care was taken to maintain the medium level at equal "distances" from the fine and coarse levels. Link congestion functions were developed (the travel time-volume relationships for the network model links) so as to maintain consistency within the network.

The Measures of Effectiveness (MOE's) compared were the load kilometers and the load-hours. It was found that the load-kilometers estimated by the two cases did not vary significantly, but the load-hours for the coarse-level estimate was substantially higher (50%). This is explained by the increase in route lengths and the rather drastic speed changes due to the removal of Class II roads for the coarse network. An important component of the 50% increase was the fictitious congestion caused because of the volume tending towards the capacity in the BPR functions.

The important conclusions drawn from the research are:

- Extreme reductions in network size leads to significant errors in the estimations. On the other hand, a medium level network, consisting of all arterials and collectors appears to give results which can hardly be improved upon.
- Even in only slightly congested networks, an equilibrium analysis seems to give better results than the all-or-nothing model.

Hearn (1984), evaluated each of the models discussed above. In addition, a transfer decomposition algorithm is developed. The salient feature of this algorithm is the division of the given assignment problem into two problems: a Transfer Decomposition Master problem (TDM) and a Transfer Decomposition Sub-problem (TDS).

Although there are no requirements for the sub-problem, the network chosen does not contain any origins or destinations in the sub-problem for reasons of simplicity. The sub-problem is chosen so that it is the complement of the master problem. The master network is completed using artificial links in the gap left by the sub-problem. The sub-problem is then solved using the Frank-Wolfe algorithm and the attributes (like travel times) for the fictitious links are fixed. Conversely, the volumes traveled on the artificial links of the master problem serve as the demands for the sub-problem. The significance of this framework is that theoretically any algorithm designed to solve the standard assignment problem can be used to solve both the TDM and the TDS. Thus it is possible to use a generalized method (like the Frank-Wolfe algorithm) for the TDM and a more complicated algorithm (involving higher order derivatives) for the TDS.

The main drawback of the model is that if the sub-problem is very big, then the Frank-Wolfe algorithm which involves a line search and computation of the objective value for the sub-problem, becomes highly computation intensive. However, if the sub-problem is chosen to be simple and feasible then the solution could make a significant gain computationally.

This method presented by Hearn is a realistic model and serves all the major criteria for a good aggregation algorithm. It helps save computing costs and is also suitable for any assignment technique. The method does not deal with diversion in particular and therefore does not deal with real-time information. Also, the networks analyzed are symmetrical and small. The methodology needs sufficient changes in order to be applicable to the problem of incident diversion.

An important difference between network aggregation models presented here and the network generator is that aggregation models emphasize the planner's perspective whereas the model under development emphasizes the real-time characteristics. This is reflected in the fact that the aggregation models are keen on retaining the original characteristics of the network. This is an important characteristic because for a planner any new modifications (like new construction etc.) always occurs on the original network. Saving computation time is only a secondary consideration here. For diversion, one is interested in developing a set of routes for a short period (in most cases) in real-time, for which the original characteristics of the network are not significant. Therefore, instead of considering volume-capacity ratios alone, other factors are incorporated into the model as expert knowledge.

The network aggregation models suffer from the following drawbacks:

1. NEE models remove links in an arbitrary fashion thereby losing the network structure and affecting the volumes on the links quite drastically. The models described here are also computation intensive.
2. NEA models shown here are all experimental ones and are parochial to the area in question. An analytical methodology to study the network seems to be quite complex.
3. The method presented by Hearn does not identify the network for the sub-problem. Although the sub-problem does not need to have any specific characteristics, it is important for diversion under incident conditions that the sub-problem be clearly defined.

From this it is quite clear that the problem is in identifying a set of conditions to choose the network for the sub-problem. The sub-problem for this case, can be thought of as selecting a network that satisfies the various criteria developed for diversion. For diversion purposes, the network to be selected has to consider the existing traffic conditions and generate a set of feasible routes, based on expert knowledge. An expert system has been developed to generate such a set of routes. Based on these set of routes a suitable assignment technique has to be developed to study if diversion actually assists in reducing delays. To ensure compliance the driver's point of view should also be incorporated into the model for the choice of routes.

2.3 Expert Systems for Route Guidance

The task of choosing the network for the sub-problem does not involve the shortest paths alone, but other real-time occurrences as well (as explained above). This is a powerful argument to use a tool, that will be able to handle such real-time data in a very efficient manner. Also important is the ability to explain to the users the thought process, which is the logic behind using an expert system for route guidance.

Taylor (1990) in his paper on specifications for a route guidance Knowledge Based System (KBS) indicates some of the salient features. For example, the problem data is seen in two parts; the first of which is the basic trip data about the origin, destination, time of day etc. The second part involves the information the system may need about the characteristics of the traveler. The contention is that in a road network with many alternatives available, the best decision (the preferred route) largely depends on the attitudes and preferred behavior patterns of the driver. An unfamiliar driver might want to drive only on the major arterials, whereas a local may be prepared to accept a route that passes through the local street system. Depending on the trip type and the driver, sometimes a route which offers a guaranteed travel time may be more appealing. Some drivers might prefer to make turns at a traffic signals, while others do not. Truck drivers would dislike slowing down at intersections. They would rather travel on a route which involved a longer time. Thus, the success of the system depends entirely upon its ability to provide satisfactory advice to its users. Also, the paper states that the fifth-generation languages like Prolog and Lisp are very conducive to develop routing algorithms.

Ritchie et. al. (1990), discuss a knowledge-based decision support architecture for advanced traffic management. The important characteristics for a real-time system are defined here. The MOE's used to judge a real-time expert system are:

1. **Truth maintenance:** As data enters the system the validity of a number of facts may change with time. The true state of the system needs to be monitored continuously. A change in the truth value of a fact higher in the hierarchy, affects the truth value of other events.
2. **High Performance:** Very short response times are usually required in the face of rapidly changing data.

3. ***Asynchronous events***: Unscheduled events should be capable of interrupting the process, according to their importance levels.
4. ***External and Sensor Interface***: Real-time systems need to be capable of gathering data for many thousands of variables. Any obvious failure (or partial failure) should be accepted "gracefully" and one should not see a rapid deterioration in performance.

In a subsequent paper, Ritchie et. al. (1993) describe a real-time decision support system for freeway incident management and control, in which, the vast amount of information flowing in to the traffic control center is stated as an important reason for automation. The freeway real-time expert-system demonstration (FRED) places very high emphasis on incident detection and verification. The FRED system currently operates on a simulated freeway network in Orange county, California. For collection of incident data, the model uses inductive loop detectors, CCTV's, Changeable Message Signs, and ramp meters. The model is next used for developing incident response. FRED has the capability to handle multiple incidents. It prioritizes the various incidents based on their characteristics (hazardous material or accident or stalled vehicle etc.). If this is not enough it studies the delay causing potential of the incidents to prioritize them. Having determined their importance, FRED makes the following decisions:

1. Determines if a management team needs to be present at the incident site,
2. It helps providing CMS responses on which it also posts a static alternate route
3. It helps decide which ramps to close down based on the demands and the incident conditions

It does not develop diversionary routes based on real-time traffic conditions. Also, it does not study the environmental impacts of the incident. The system provides a framework for incident management in the California region. The characteristics of an expert system shell called G2 are also discussed here. The salient features are listed below:

- All data entities are represented by objects. It is capable of representing the basic data types like integer, floats, characters and logical variables. It can create objects for transient events. The relation between these objects is represented by setting up a binding between two objects.

- Knowledge representation is in the form of "if...then" rules and procedural functions. It can support both backward and forward chaining. Some of the knowledge that requires sequential manipulations of the data are served by developing functions in high-level languages such as Pascal.
- G2 also provides a data interface to external programs. A simple bridge has been written to traffic simulation software.

It is important to note here that KAPPA-PC which has been used in the development of the network generator provides all of these capabilities. It has additional capabilities like links to database software and spreadsheets. There are a few problems associated with the development version of KAPPA-PC which are discussed in Chapter 6.

There are other expert systems developed to assist the teams in incident management. Gupta et. al. (1992), aims to develop an expert system for Freeway Incident Management on the Massachusetts Turnpike. The expert system identifies the incident type and location and the material being carried by the vehicles that are involved in suggesting the response strategies. Based on the position of the incident, the program develops displays of pre-planned diversionary routes. These routes are checked for available capacities in real-time. In contrast, the model that has been built, builds the routes based on real-time traffic conditions instead of using pre-planned diversionary routes.

The KBES developed for incident management here consists of four parts:

1. ***Incident Detection and Verification:*** This requires a freeway surveillance system. In the system used here, passing motorists or police patrol are used for surveillance. Verification again is done using police patrol.
2. ***Classification of the Incident:*** Information is then provided by the police patrol on incident characteristics, such as time of occurrence, location and the severity of the incident.
3. ***Notification of the Incident:*** This consists of notifying all the agencies required to clear and manage the incident site, after the verification of the incident.
4. ***Diversion Module:*** This module determines if diversion is necessary and checks if the pre-planned diversion routes are available for diversion (i.e. if the volume to capacity ratio along this alternate route is less than 0.6)

This paper explains the basic system needed for incident management. The model built is highly superficial. The different incident types considered is not complete. There are spatial and location factors, that are neglected which could have a significant impact on the delay characteristics of each incident. The process described here is also highly static. The routes for diversion cannot be selected anew if the pre-planned routes are congested for some reason. Thus the diversion process is constrained by the model developed.

The next paper studies an expert system model for post-incident traffic control, that was developed by Zoe Ketselidou (1989) at the University of Massachusetts at Amherst. The model developed is based on predetermined weights assigned to links based on the time of the day and the historical traffic volumes. Points at which diversion can be initiated and the potential destinations for particular links are also determined beforehand. The search process then carries out an exhaustive search based on these preset thresholds. The different criteria that were ideally used for selection of alternate routes are described here:

1. Proximity to main corridor routes,
2. Usefulness for access to ultimate destination,
3. Driving quality on route,
4. Impact on adjoining land-use,
5. Jurisdictional problems.

The system compared the results obtained from the expert system to a simulation package called TRAFLO, which was developed by the FHWA, and found that the results of this model were fairly consistent with the results of the simulation runs. The network used for testing the system is in Long Island, New York. This represents a freeway corridor, which offers great potential for alternate routes assessment, because there a number of freeways and associated arterials in the network, running parallel to each other. The model has the following modules in it:

1. **Data Inputs:** Incident characteristics and traffic flow characteristics are obtained as the major inputs
2. **Capacity reduction module:** This estimates the reduction in capacity of the incident link, based on the incident duration and number of lanes blocked, and decides if diversion is necessary.

3. **Search Algorithm:** The search algorithm selects the best route based on the preset link weights, and assigns a portion of the volume on to that route. Selection of the best route could be a time-consuming process. This is repeated till the entire volume is diverted.

The alternate route pre-selection process used in the above mentioned models will have to be carried out repeatedly if:

1. New roads are constructed in the current network,
2. The network itself is changed,
3. The pre-selected alternate routes are unusable due to certain combination of real-time factors.

The model is not responsive to sudden changes in the traffic flow patterns as it uses only the historical data for analysis. Certain safety considerations or occurrence of secondary incidents can seriously impact the validity of the solution provided by this model. Also, user compliance has not been explicitly addressed in the development of the diversion strategy. Also, the calculation of delays network wide has not been considered. The network generator model aims at improving these features.

2.4 Real-Time Diversion And Routing

This portion of the literature review deals with the need for real-time dynamic routing of vehicles during incident conditions and with the importance of quantifying the delays for different assignment techniques. Also, emphasis has been laid on predicting traffic flows to act as a feedback mechanism for the diversion algorithm.

In real-time, the diversion strategy must be responsive to changes in demand as well as to changes in capacity. According to Ketselidou et. al., the best diversion strategy is the one that balances the unused capacity according to three rules listed in order of decreasing priority:

1. Wherever possible, the link demand will be kept below the capacity for all links in the system.

2. Whenever the unused capacity is less than a certain threshold, control will be exercised to balance the unused capacities on an adjacent pair of links provided Rule 1 is not violated.
3. Diversion will be terminated when neither of the above rules is violated.

This was demonstrated by simulation on the New Jersey Turnpike by the Sperry Systems Management group.

An alternative to diversion of traffic is the flexible choice of departure times during the most congested hours. Since the diversion of traffic is most critical during the peak hours, it is necessary to compare the two techniques. Numerous papers have been written on departure time flexibility. Ben-Akiva et. al., (1984) have developed a dynamic model of traffic congestion, that considers a limited number of bottlenecks. Choice of departure times is based on the trade-off between travel time and schedule delay, which is the difference between the actual and desired arrival times. The model which was developed based on a nested-logit approach, found that the departure time flexibility can result in shifted peaks and also reduce their amplitude.

A paper on similar lines was written by Chris Hendrickson et. al., (1983). They developed a logit model for work trips and the results of the analysis are presented below:

1. Departure time flexibility offers a latent capacity in transportation facilities and are exceedingly useful when short-term disruptions (transit strikes, reconstruction of major facilities, etc.) occur. Prior knowledge of these short-term disruptions is necessary,
2. Time dependent tolls are likely to be useful tools in transportation system management, and
3. Congestion relief due to transit service improvements is likely to be over-estimated because reduction in congestion induces shifts to the most congested period.

Estimation of delays was done using a simulation package by Garrison and Mannering (1990). They, too, have tried to estimate the impacts of incidents on the freeway level of service. This study was conducted in the Seattle area, which was divided into six zones with 1,071 links. Of these 387 were physical highway links, and 684 were zonal access

links. The package used was XXEXQ, in which the route performance (i.e., travel time versus traffic-flow relationship) is defined by modified Bureau of Public Roads (BPR) functions.

Traffic is assigned using an user-equilibrium algorithm. However, instead of assigning traffic to the network with a single peak-hour origin/destination matrix, the model assigns traffic sequentially, in 10 minute intervals, with a series of O-D matrices. This is called the sequential user equilibrium assignment process. This framework allows the issue of driver information to be addressed.

The conclusion reached was that if all the drivers were informed about the traffic conditions, then the system optimal (in terms of minimizing delays) does not occur. The system optimal occurs when roughly 50% of the drivers have full information. This situation is a result of the sacrifice a certain portion of the system users have to make in order to reduce the system-wide delay. Furthermore, when they assumed that around 60% of all drivers had complete information (60 was chosen to realistically represent the travel conditions in Seattle), every minute of an incident costs around 2,000 dollars in terms of lost time and fuel. This amount was estimated based on a conservative estimate of the value of time in Seattle as \$8/hr. This cost is a reflection on the importance of an appropriate response to an incident. . This study clearly shows the need for a cohesive incident management team that can provide a complete response to any incident.

The model fails to implement any diversion strategy which is an important aspect of the incident response process. It is passive in the sense that although it studies the flow on the links, it does not attempt to change the flows on any of the links. The process of diversion of traffic could help in saving significant amounts of time and money. Another benefit of diversion would be the prevention of the occurrence of secondary incidents, which is also a cause for great concern in congested traffic networks.

2.5 Prediction Models

For any diversion strategy it is necessary to have a prediction model that will predict future traffic conditions in real-time. The models will shed some light on the traffic conditions in the immediate future. This will help the diversion algorithm identify problem areas and links that are going to be congested. The algorithm can then take necessary steps to alleviate this problem. A few of the prediction models are studied below and their applicability to the model developed is discussed briefly.

The paper by Stamatiadis et. al., (1994) examines the quality of real-time travel time predictions, obtained with a recursive identification algorithm. Two elements are considered:

1. the structure of the model
2. the estimation of the parameters of the model.

Historical data is used as input to the model, and this includes travel time information from the upstream and downstream links. This accounts for any traffic waves that appear in the model. A traffic simulation package is used to examine the performance of the model under two different sets of traffic conditions, i.e. normal and congested (non-recurrent) traffic conditions.

The structure used by the model assumes that if the travel time of link l , at time t is linear in its parameters, then it can be written as:

$$A(q^i).T_l(t) = \sum_{k \in I} B_k(q^i).T_k(t) + \sum_{j \in O} C_j(q^i).T_j(t) + d.T_l(t) + \varepsilon(t)$$

where I is the set of links ending at node i ,

O is the set of links starting at node j ,

where i and j are the start and the end nodes of link l respectively,

$T_l(t)$ is the average travel time of the link l for the time interval t , obtained from previous days, and

$\epsilon(t)$ term is the noise which is assumed to be a series of independent random variables with mean value equal to zero.

Based on this the prediction algorithm is developed which gives us a one-step prediction for the travel time of link l . The n step ahead predictions are obtained by successively using one step ahead predictions. The estimation of travel times will have to be performed on line, which implies that the model uses the most recently observed data.

The results that have been observed for normal conditions:

1. The prediction errors resulting from these models are smaller than the ones if no predictions are made.
2. When a diurnal term is introduced into the model the performance improves perceptibly.

During incident conditions:

1. The one step predictions are worse than the five or ten step prediction process, and they are all better than the no prediction case.
2. When the diurnal term is included, predictions are similar to the case with just the autoregressive component.

The model works well for prediction during normal conditions. The model framework is not conducive to account for abrupt changes in the traffic flow- which implies that the model will not perform well for incident conditions. Also, the historical and current flow traffic data is not substantial to accurately predict the travel times. Traffic volumes can change significantly from the historical volumes, because of extraneous factors. It might be necessary to incorporate more number of variables.

In another study by Stephanedes et. al., (1990) to predict traffic volume for the short-term, two models employing the Kalman Filtering theory are proposed. In these models the prediction parameters are improved using the most recent prediction error and better volume prediction of up to 80% is achieved compared to a package UTCS-2. The disadvantages of the package are as follows:

1. The package is highly dependent on historical data: Traffic volume can vary substantially because of external factors (e.g. weather, incidents, special events etc.).
2. The package is not transferable across systems.
3. The package requires a massive data base which has to be updated off line periodically.

The Kalman filter employs a matrix which reflects the traffic conditions based on a number of variables. These are associated with a parameter matrix to predict the traffic volumes τ periods ahead as follows:

$$z(\tau + k) = H_0(\tau).x(\tau) + H_1(\tau).x(\tau-1) + \dots + H_r(\tau-r).x(\tau-r) + w(\tau)$$

where,

$z(\tau+k)$ = the traffic volume k intervals ahead of time τ ,

$H_j(\tau)$ ($j=0,1,2,\dots,r$) = parameter matrix of dimensions ($m \times n$)

$w(\tau)$ = a noise vector of dimension m

The procedure then iterates to further predict the volumes based on the current prediction. The models have been tested for a 4 link network in Nagoya city, Japan, to incorporate the effect of the presence of other links. This Kalman filter is then tied to a logit type model for optimal freeway ramp-metering to prevent congestion during incident conditions. The Kalman filter effectively serves to adjust the model parameters of the logit -type model to allow it to monitor the current state of the system.

The results from the model are as follows:

1. At all times the new prediction models perform substantially better than UTCS-2
2. When prediction is performed with smoothed rather than raw data, UTCS-2 performance deteriorates, whereas the model developed achieves its best performance.

The model develops a sound procedure for predicting the traffic flows, although further improvements are possible. An important aspect is with respect to the effect of the traffic flow on other links on the links for which the flow is being predicted. Also the network used is quite a simplistic one. As the model is quite computation intensive it would be

difficult to implement it over a wide network in real-time. The model needs to be tested over a variety of networks before it can be directly applicable.

Yet another model to predict destination specific traffic densities on urban freeways was built by Gary Davis et. al., (1994). The model is a continuous time Markov compartment model of freeway traffic flow. The study uses the method of "*large population approximation*" to approximate the underlying stochastic process by the sum of a non-linear deterministic process and a linear time varying Gaussian stochastic process. This approximation is then used by a Kalman Filter to track a freeway section's density broken down by destination.

The model was able to predict destination specific densities for freeway segments better than for a road network. The model needs to be improved to update the parameters involved in the process real-time. Then the process could be made recursive.

Although an assignment technique was used finally, in order to predict future traffic flows, modifying one of the models mentioned above, to be used along with the network generator, will prove to be an important contribution. This could prove to be an important area for future research.

2.6 User Route Choice Models

User compliance is another important issue with respect to diversion. Although the routes selected to divert the driver on, may be very good, the diversion process may fail if the confidence level of the driver is not high. When the network generator selects the routes in real-time, it is important to model user behavior and ensure that the final routes do consider the user view-point. Therefore, modeling user behavior is considered very important.

In a recent study in Texas A&M Conrad L. Dudek et. al., (1994) it was found that in order that results of diversion be predictable, it is very essential to model individual user's preferences and thresholds. The study concludes that drivers based their time-saved thresholds on how much they disliked one of the recommended routes (and not what they preferred most). The effect of the individuals preferences on the traffic flow is significant

enough to negate the gains of diversion. Therefore, to maintain the validity of the selected routes, it was necessary to model the user's viewpoint.

The propensity for drivers to divert is also investigated in this section. Researchers have found that traffic information influences diversion. Heathington et. al., (1971) found that driving patterns of frequent diverters were influenced slightly more by traffic reports than by visual observation. Mahmassani et. al., (1990) found that drivers who listened to radio traffic reports have a "greater propensity to switch routes". Daniels et. al., (1976) found that the average frequency of diversion for Chicago drivers was 16-27% for observed congestion and 17-35% for traffic reports. Also most of the benefits of route are expected to ensue from savings in travel time and fuel, and wear and tear of the vehicle.

Drivers diversion propensity was studied extensively by Khattak et. al., (1992). The model developed was an empirical one that focused on a survey of downtown Chicago automobile commuters. The stated preference approach was used to study drivers diversion propensity. A survey focusing on several aspects of driver behavior was handed out to drivers at parking lots in Chicago. The multivariate models of diversion propensity developed indicate the relative importance of each variable in determining the diversion route. The results of the model showed that drivers expressed a greater willingness to divert during incident conditions rather than for recurrent congestion. The second important result was that commuters were more willing to divert to an alternate route that passed through safe neighborhoods, was familiar and had no traffic stops. Furthermore, it was found that drivers of higher income were more willing to divert possibly due to their higher value of time.

Cascetta et. al., (1992) studied the route choice characteristics of the users. A statistical descriptive analysis was carried out on the data from a road network in Torino. The survey was carried out on workers from the same plant in Torino, thereby resulting in a single destination. The results of the analysis showed that the significant variables in an individuals route choice are:

1. The travel time on primary roads,
2. The travel time on secondary roads,
3. The total length of the path,

4. The number of left turns, and
5. The number of signalized intersections.

This result is also supported by Ben-Akiva and Lerman in their book on discrete choice analysis.

2.7 Conclusions

The basic task is to identify a set of routes, based on certain feasibility and user behavior criteria. This is done by developing an expert system that identifies such routes, by extracting links from the original network. This would represent an important improvement over currently existing decision support systems, because these systems work only after alternate diversion routes are manually input into the system.

The model developed here attempts to automate the process of reducing the network and selecting a set of alternate routes. This reduced set of routes does not necessarily have to retain characteristics of the original network. A static assignment technique is to be used on this set of routes so that one can see how the delays vary and if diversion really helps alleviating the problem of non-recurrent congestion. The approach and the methodology for developing the different models for each of these aspects is described in the following chapters.

3.0 DEVELOPMENT OF MODEL & ANALYTICAL FRAMEWORK

3.1 Introduction

This chapter deals with the model development process for the network generator. The model represents a proof-of-concept demonstration for selection of alternate routes in real-time during freeway incidents. In order to validate the model, a study of the delay characteristics (during the diversion period) has been initiated within the analytical framework of the model. The model aims to reduce delays on the network by selecting an appropriate set of feasible diversion routes. A generalized framework applicable to most networks is described in the latter part of this chapter, although the model may require calibration of thresholds for the networks on an individual basis.

The next section of this chapter discusses the current diversion practices and identifies areas for improvement. Before this is done, it is important to compare existing network aggregation models with the analytical framework developed for the network generator, so that the function of the network generator can be clearly defined. Aggregation models involve an assignment technique within their framework in the link elimination stage, i.e. in order to identify links that carry low volumes and eliminate them. They also involve issues of connectivity and completeness in the network. In the network generator, assignment is an external module and is used only to forecast future traffic conditions and to develop Measures of Effectiveness for comparison. In order to reduce a network so that alternate routes for diversion can be found, it is not necessary to retain completeness or connectivity of the network. Links may be removed without consideration of the effects on the rest of the network. This is because, the model is intended to solve a problem that is transient in nature. Diversion during an incident does not extend for more than two hours, in most cases.

When the network generator is extended, to be used in association with dynamic traffic assignment, it will be necessary to maintain the completeness of the network, i.e. all the origin destination flows must find routes to travel on and the reduced network should represent the original one.

3.2 Freeway Incident Management and Diversion Strategies

There is a growing need for Freeway Incident Management (FIM) and several FIM teams are being established all around the country. Although the FIM teams have been highly successful in dealing with incidents, there is still room for considerable improvement, in carrying FIM practices to the 21st century. Automation of incident detection, real-time coordinated response of different agencies involved in FIM, more efficient allocation of resources and real-time implementation of diversion strategies are among the possible improvements. The implementation of real-time diversion strategies is the single most important concern in this research effort.

3.2.1 Deficiencies of Current Practices in Diversion

The state-of-the-art in diversion during incidents clearly demonstrates the need for research in this area. Some of the important features of the diversion techniques currently used around the country for IM are:

- Diversion is not attempted for any incident with a duration of less than two hours. This is because the time taken in setting up alternate routes is high; for e.g.: the selection of the alternate route is arbitrary, all the members of the team are unaware of the details of the entire route, etc.
- Routes for diversion are selected from a book of maps (Washington D.C., Massachusetts etc.) that have been prepared in anticipation of an incident on every freeway link. A sample map is shown in **figure 3.1**. Many of these maps are based only on the shortest feasible distance that ensures access back to the freeway. They do not consider the incident severity or other real-time traffic conditions such as the traffic restrictions, school openings and closings, etc.. As a result, secondary incidents occur in many of these side roads which do not have adequate capacity to carry traffic volumes diverted from the freeway.
- Day-to-day variations in the traffic are not accounted for by the diversion maps. Some of these preselected routes may be unavailable for diversion purposes at that time because of high congestion levels. This may be a result of factors like the weather or ball games, or due to maintenance/construction work on a link, etc.

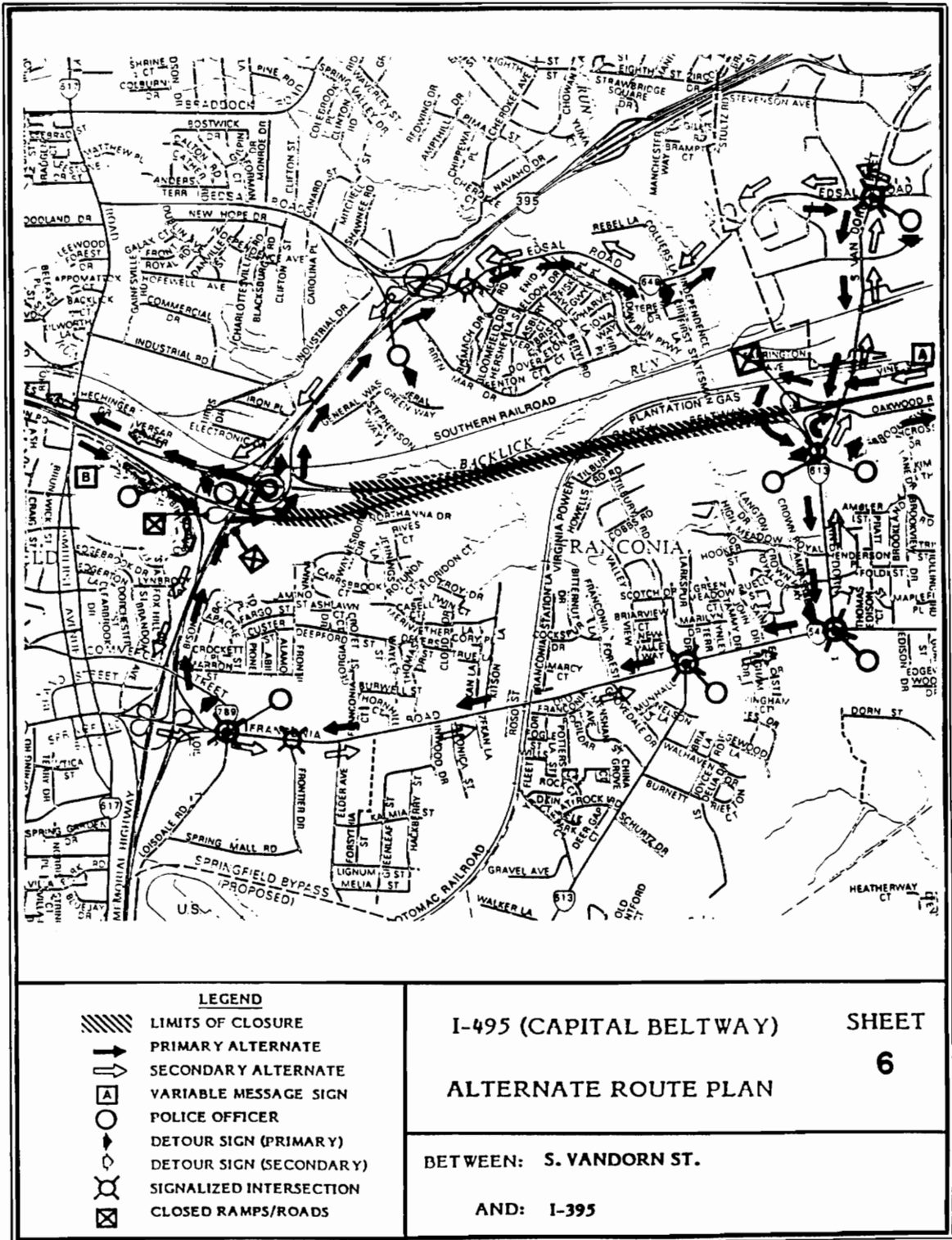


Fig. 3.1 Diversion Map from NOVA

The diversion maps serve only to bypass the incident. The diversionary routes pass in the vicinity of the incident link itself. Diverting traffic upstream of the incident may be more beneficial because the incident zone would be relieved of a fraction of the incoming traffic.

- The driver's behavior has been neglected completely in the selection of alternate routes which could result in poor user compliance. This phenomenon has been observed in different areas of the country.
- There has been no quantification of the delay based on the implementation of diversion strategies. Some efforts to this end have been expended in Seattle and Long Island (New York), but they still need improvement.

However, with the advent of Intelligent Vehicle-Highway System (IVHS) technologies, off-line selection of diversion routes for non-recurrent congestion will no longer be adequate. It is assumed that IVHS in later stages, will help in the collection and dissemination of network wide information with great accuracy and in real time. These are fairly reasonable assumptions, within the context of an Advanced Traffic Management System (ATMS). This information can then be processed and used to develop and evaluate time dependent decisions, i.e. the most suitable decision for each period. Under such a system feasible alternate routes must be determined in order to perform the benefit/cost analysis for each decision before implementation. Such a model does not exist at present. The network generator is aimed at satisfying these needs.

3.3 Requirements for the Analytical Framework of the Network Generator

Based on the problems identified with current diversion practices, the important functions of a diversion model within an ATMS framework, are listed below. These can be considered to be the objectives for the model:

1. Decisions on diversion needs for that incident should be made. If the delay caused by the incident is not significant then a decision not to divert could be taken.
2. The model should then select a set of feasible routes upstream of the incident, on to which traffic can be diverted. This is an elimination process.
3. Routes selected by the model should be meaningful, i.e. the routes should not be circular or too long. They should have enough capacity available to carry the

diverted traffic. As far as possible they should also not have a set of links in common (unless the volume-capacity ratio is very low). These are examples of the heuristic knowledge applied by experts in determining alternate routes. This should be used in collaboration with available knowledge on future traffic conditions to select links suitable for diversion. Further explanation of the criteria for route and link selection are discussed in Chapter 4.

4. Driver compliance is a problem that plagues diversion strategies all around the country. The road user's viewpoint (comfort and safety) should be accommodated, so that user compliance can be ensured. Building the confidence of drivers to the level that they select one of the alternatives suggested to them, is very important and is not possible without modeling their behavior.
5. Delays caused by the incident are to be quantified. The traffic flow on adjacent links should also be collected, in order to study the effects of the incident. This is carried out by using a static assignment technique. The effects of these assignment techniques needs to be studied, in order to predict the traffic flows for the diversion routes. This prediction can then be used to compare the delays with and without the network generator.
6. The model should be efficient and be capable of producing results in real-time. Also, the results produced by the model should be easily understood by the user. It can then be used as a training tool for new members of the FIM team. It will also help in identifying areas for improvement in the functioning of the team. This is an important consideration for deciding the usefulness of the model.

Based on these specifications for the model, it was decided to seek out experts and question them regarding the validity of the steps. This is described in the next section.

3.4 Knowledge Acquisition Process

This step represents a very important component in the development of any knowledge-based decision support system. The aim of the knowledge acquisition process was to research the expertise in diversion during incidents, from various sources and classify them in such a manner that it can be used in the expert system. As stated earlier, the knowledge is then codified as "if-then" rules which can then obtain the necessary output from the model. The different sources for the knowledge are:

1. **Experts:** Major Ronald Miner, who served as the head of the Fairfax County police department, was one of the most important sources of expertise. He was also the head of the Freeway Incident Management (FIM) team in Northern Virginia (NOVA). His intimate knowledge of the traffic characteristics and the expectations of the public in the NOVA area, greatly helped in the development of the system.
2. **Diversion Maps:** As stated earlier, in NOVA, the FIM team uses a set of static diversion maps that are available at the incident site for various links on the freeway. These maps were prepared to cover most of the locations where diversion may be necessary and are based on a certain set of geometric and network data. Routes have been chosen after being assessed as the best possible routes over which to divert traffic, if that should be necessary. These maps were studied extensively to inculcate the knowledge used in their selection into the model.
3. **Literature:** The literature on diversion during incident conditions helped identify the defects in existing models and also define the functions the current model should have in order to make a significant impact on the diversion process. They also help define the important characteristics of a real-time system for traffic management.

There are two important aspects to the problem on which the knowledge was sought:

- **Operator Viewpoint:** The operator aims at optimizing the system resources. In this case, the system resources are the capacities available on the road network. The optimal use of the road networks is of interest here and penalizing individuals to improve the overall solution would seem lucrative to the operator. The safety of the drivers on the alternate routes selected is also a prime concern of the operator. This is because, unsafe routes may lead to secondary incidents, that further disrupt traffic flow. The knowledge was mainly obtained from the experts and the diversion maps developed in the Northern Virginia Freeway Management Team Operating Manual. This includes congestion levels, future traffic predictions, traffic restrictions such as HOV lanes, icy bridges, height restrictions, school/office zone locations, etc.
- **User Viewpoint:** The user's are usually very interested in the nature of the alternate routes. Their concerns include the relative time spent on the freeway and the arterial, the number of left turns and signals on the routes, unsafe neighborhoods, etc. If these are not addressed or accounted for, the driver does not feel comfortable changing the route he commutes by. In other words, it can be said that the driver selects the route

which has the least disutility associated with these variables. Also, there is a maximum threshold involved and if the minimum disutility is greater than this threshold, he may not divert.

The knowledge acquisition process is one that continues through a software's lifetime. This is especially true of knowledge-based artificial intelligence models. The model can be updated periodically when new knowledge is acquired from the above mentioned sources. The OOP paradigm used, facilitates the development of such a system.

3.5 Justification for Using a Knowledge Based System

Developing an analytical model to solve the complex problem described previously has several disadvantages. Analytical models do not truly represent the original problem. If it attempts to represent the original problem in a satisfactory manner, the model becomes highly computation intensive and the emphasis shifts from realistic representation of the problem to the development of good solution techniques. Moreover, the knowledge and the data for this problem involve a significant amount of heuristics. This cannot be easily represented in an analytical framework. Also, analytical abstraction models discussed above have all been developed for planning purposes and do not try to incorporate features for a real-time response.

3.5.1 Deficiencies of Analytical Models Used for Diversion

Although many analytical models have been used to reduce computation times, they suffer from the following deficiencies:

1. Savings in computation time for many of the sample networks used is not significant. For most models the savings is between 25 and 30% for the tested networks. So, the real-time capability, which is an important requirement for the incident management process being developed here is not satisfied. Many of them have not been validated for bigger networks.
2. Selection of the alternate routes, is based only on the travel time on the routes. This leads to a mis-representation of the original problem. The network generator model

on the other hand, is able to represent apart from the travel-time other criteria such as the incident severity, weather, safety, driver perception and network geometry.

3. Some of the drivers are diverted to areas which are not directly on their perceived path to their destinations. Also, the models do not incorporate any of the disutilities associated with the user's point of view in deciding the routes. Therefore, although these models help in reducing the network for the assignment process, driver compliance is neglected.
4. The models themselves sometimes might increase the complexity of the assignment by introducing new origins and destinations (Haghani et. al., 1982). Moreover, most of the algorithms have been developed specifically for planning purposes and not for real-time applications.

3.5.2 Advantages of Knowledge-Based Models

The highly heuristic nature and the large volume of the information involved points to use of paradigms other than conventional algorithmic procedures. It is important to examine the possibility of using Artificial Intelligence (AI) procedures for path processing and routing. AI tries to replicate intelligent human behavior on computers. This capability is especially suited to the route selection problem, since it involves predicting human choices. **Table 3.1** presents a sample list of five variables associated with route choice decisions, that are used commonly in models for diversion and incident management. Possible values of these variables are also shown.

Table 3.1 List of Variables Associated with Route-Choice and their Values

Variable Name	Allowable Values
Link Type	Freeway Signalized Arterial Arterial
Day	Weekday Weekend
Time	Morning Peak Evening Peak Off-Peak
Weather	Sunny Foggy Icy/Foggy Windy Cold
No. of Lanes Blocked	1 2 3 or greater

For a given incident, the 5 variables from this partial list could occur in as many as 300 different combinations, thereby making the number of possible scenarios very large. Thus, it is obvious that the problem is very complex. In comparison, most of the analytical models for rerouting, aggregate all the variables as stated earlier into a single Measure of Effectiveness (MOE) such as the travel time. In this process, most analytical models make simplifications that reduce the complexity of the problem.

In comparison, AI models use an expert's knowledge, which helps them to reason, as the task is being performed. An expert (in this case freeway incident management personnel or a traffic management center operator) would typically assign different weights to each of these factors and decide to divert traffic on routes which satisfy most of the constraints. Knowledge is used in a parallel fashion as opposed to a serial use of knowledge by analytical algorithms. This allows them to search the solution space in a swift and efficient manner. As a result, although the speed of procedural languages (like 'C') is greater than fifth generation languages (like Prolog), AI methods often solve the modeled problem in a shorter duration of time.

There are three basic approaches to AI (15):

- a) Artificial Neural Networks,
- b) Artificial Evolution, and
- c) Knowledge Based Systems (Heuristic Programming).

Although the first two approaches have been used extensively in other fields of engineering with a fair amount of success, transportation engineers depend heavily on heuristic programming or on neural network technology, because of the human factors involved. Also, since there are a large number of parameters, each involving some uncertainty and unpredictability, that aid in reaching the decision, it becomes necessary to develop a model that will incorporate some amount of fuzziness in it. Another important capability of the Expert Systems (ES) methodology, which is very difficult to implement in procedural programming is that, it makes the reasoning process transparent to the user. These requirements confirm the need for a heuristic knowledge-based system.

The first task in any expert system development process is the knowledge acquisition from experts and other sources such as the literature and simulation processes. Once this is achieved, the system has to be modeled based on this information such that it replicates the thought process of the expert. Also, since a real-time response is required for the network generator, advice needs to be provided within several seconds of the input. Based on these and other considerations the model was developed on an expert system shell called KAPPA-PC. This is discussed in great detail in the following section.

3.6 KAPPA-PC Expert System Shell

The development version of KAPPA-PC (Intellicorp) was used. It is a software that can be used in applications requiring intelligent monitoring and control. The version used here, runs on an 80386 or an 80486 machine with Windows 3.1 programming environment. There are several features of KAPPA-PC which have facilitated the development of this decision support system. These characteristics are discussed below.

KAPPA-PC stores all the data in the form of classes, objects (known as instances) and properties of instances. The objects can inherit properties selectively from their parents

and can also inherit from different classes. These objects can be synthesized dynamically and their properties can still be manipulated. The shell supports the OOP paradigm which is discussed in detail below and uses rules, functions and methods to manipulate the data. This allows data-hiding. i.e. A certain set of functions and methods can manipulate properties only within a certain section of the framework. Other sections are shielded from the effect of their manipulations. This helps preserve the integrity of the data. Another implication of data-hiding is that new branches can be attached to the existing framework, without having to change the current system extensively.

Rules, functions and methods are the three constructs used to manipulate the data. Rules are simple "if.....then" constructs which encapsulate the knowledge acquired, during the acquisition process. The shell supports both forward chaining and backward chaining. Forward chaining assigns values to the variables in the antecedent part (the "if" part) of a rule. Backward chaining starts from a goal and searches for other rules that provide values to the antecedent of the current rule. The functions and methods are written in a language called Kappa Application Language (KAL) which resembles "C" to a great extent. It is highly structured and has built in functions for various purposes such as list manipulations, string manipulations, math functions, logical operations, etc. This allows the data to be used in a number of different ways.

There is an additional feature called the rule trace, which displays the rules as they are being executed. The inference browser and rule-relations capabilities allow the developer to step through the rules one at a time and select the rules to be watched. This helps significantly in debugging the code. The functions and methods editor also include a debugger. Rules can be executed in groups with priorities assigned within a group. This allows control over the direction the knowledge should progress in. These features are extremely useful in designing and debugging the system. Development of the system has been greatly aided, by the presence of other features like the dynamic creation of objects and the graphical user interface to the shell.

The expert system also allows access to files in a database format which includes access to software like dBase, Excel, etc. This allows a significant amount of data to reside outside the main program, to be accessed when the need arises. Maintenance of an incident-log

and summary of actions undertaken during the course of the management process is also possible using this tool.

The shell also allows the opening and closing of ASCII text files, which can store formatted data. Data can also be input directly by the user within a framework known as the "**consultation metaphor**" in expert systems parlance. For problems involving repetitive input, this mode of input is cumbersome and should be used selectively.

The shell also has various display capabilities such as buttons, meters, single and multiple list-boxes, bitmap images and text boxes which can show the manipulated data to the user efficiently. The display capability extends to the opening and closing of new windows for display of only the relevant material at every stage. The graphical user interface for the shell is custom-made for quick development of applications.

3.7 Problem Representation

KAPPA-PC is a tool that uses the Object Oriented Programming (OOP) paradigm in the representation of the knowledge and data. The paradigm involves defining abstract data types (ADT's), which represent complex real-world or abstract objects, and organizing the program around these ADT's with an eye toward exploiting their common features. An ADT refers to a programmer-defined data type together with a set of operations that can be performed on that data. OOP is a method of designing and implementing software and the user of the program is transparent to the use of such techniques in the code. The programmer gains significantly in many cases from the use of OOP techniques. Although it is easier to implement object-oriented techniques, in a language that supports it, OOP is also independent of the programming language. There are three basic concepts underlying in this paradigm:

1. **Data Abstraction:** is the process of defining a data type, using data hiding. This definition of an ADT not only specifies its internal representation, but also the functions that other program modules use to manipulate it. Data hiding ensures that one can alter the internal structure of the ADT without breaking the programs that call functions operating on that ADT. For e.g. Both arterials and freeways have similar

properties like length, backnode and forenode, but only arterials have the property of signals associated with them.

2. **Inheritance:** Each object is related to one or more other objects. This is the notion of defining a new object in terms of the old ones, because many objects have certain characteristics in common. For e.g. The time of occurrence and the weather during clearance are variables associated with all the incident types (spilled load, vehicle collision, fatalities, disabled vehicles etc.) and they inherit this from their parent class called "incidents". Even new incidents possess these properties.
3. **Polymorphism:** In the context of OOP, polymorphism means that a single operation can behave differently in different objects. For example: In KAPPA-PC there are a variety of different objects that can be represented on the screen. A single function called "ResetImage" will help display updated information on different objects. Obviously the actual mechanism of resetting the image differs from one object to another.

All of these concepts have been used in the development of the framework for the network generator. An ADT is a template from which specific instances of objects can be created as needed. The term class is used synonymously with this definition of an ADT. The network is represented by means of links, nodes and routes for this problem. The different classes used in this framework are:

- **Links:** These are uni-directional pathways between intersections. There are a number of variables associated with this class such as the backnode, forenode, length, capacity, speed, number of lanes, special characteristics (bridge, tunnel, overpass, presence of school or stadiums, etc.).
- **Nodes:** This is another important class which consists of intersections and decision points for route change. The variables associated with this are the ground coordinates, screen coordinates (for graphical representation) and the presence of signal lights on this node. These are usually static information, that is made available to the model.
- **Paths:** These are also created in real-time and their important characteristics are the list of links and nodes. Each path also has a disutility value associated with it. This is based on the number of left turns and signal lights the user is required to pass through, the time spent on the freeway as opposed to arterials and secondary roads

and the length of the entire route. Three paths with the least disutilities are displayed finally to the user. This ensures that the driver has been considered while making the decisions, and will help him/her feel more comfortable.

- **Incidents:** Although, this class exists, in most cases it is necessary to create objects in this class dynamically. i.e. Incidents are objects that are created dynamically. The properties of an incident also vary depending on the incident type, location, etc. As an example consider a spilled load incident and a traffic collision. These have certain common characteristics such as the properties of time of occurrence, location, etc. The spilled load has certain characteristics, for e.g. if the material is on fire, whether the material is easily identified, etc. In a traffic collision the important characteristics are the number of vehicles involved and whether any personal injuries are involved. These variables need to be created dynamically, based on the incident type. This is done by functions associated with the object itself. The functions that operate on particular objects are known as methods in OOP paradigm. They define the behavior of an object.

The other ways in which the data can be manipulated in the OOP paradigm is by the use of methods to pass messages to other objects. The changes in the values of variables in other classes causes the activation of other methods in that class. Thus a new set of methods can be activated. Since these methods are constantly waiting for changes in the values of certain variables, they are also known as demons.

The expert system shell allows the use of functions and rules to change the data. The system developed here, uses the forward-chaining approach for this problem. For the purpose of diversion since the number of goals are too many, it was decided to use the forward chaining mechanism.

3.8 Analytical Framework for the Model

Based on the knowledge acquired and the functional requirements expected from the system, a framework for selection of diversion routes has been developed for the model and is shown in **figure 3.2**. It involves seven major steps:

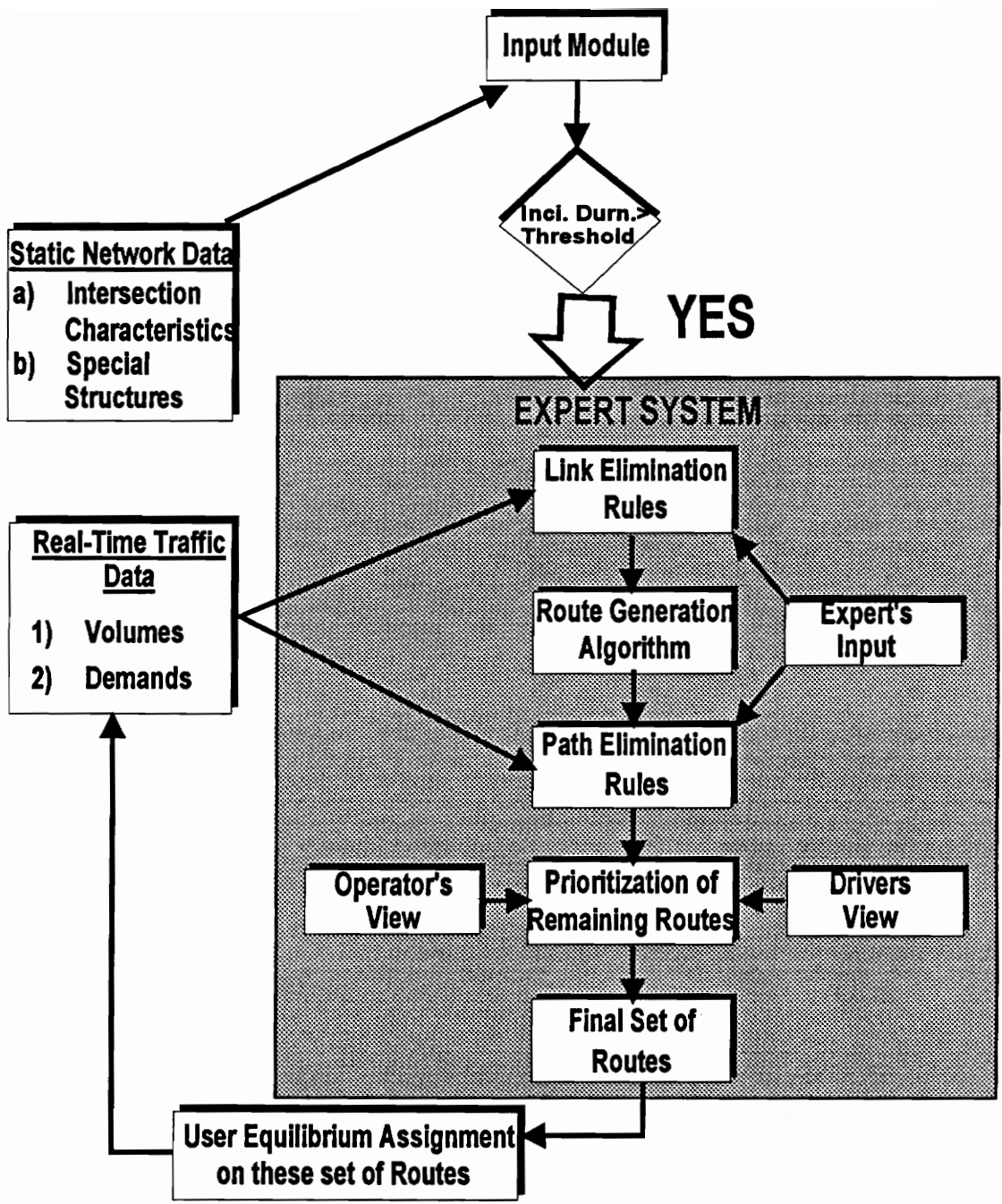


Fig. 3.2 Analytical Framework for the Network Generator

1. The static network data is read. The current traffic flows are calculated based on a simple assignment of the O-D trip tables for the network. This involves a one-time computation cost. This represents the flow for a "normal" day.
2. Duration of the incident is calculated based on the different characteristics of the incident. This involves details like the type of incident, the day and time of occurrence, the weather conditions, the location on the freeway, etc. This is a knowledge base developed from the literature on the subject of incident management.
3. The delay in vehicle hours is calculated for the incident using a queuing model. A deterministic model is used in order to model the road conditions realistically. The model uses data about the incoming volumes and the number of lanes blocked as input. The data for the incoming flow of traffic is procured from an one-time assignment procedure, which is assumed as the current normal flow on each of the links. The stochasticity in the arrivals and departures is lost during congested conditions. The need to divert is decided based on the calculated delay.
4. The next step is the link elimination that is essential in reducing the search space within the network. The elimination process should be reasonable and efficient so that the model is very efficient. It is acceptable for the model to reduce the network to a great extent because, only a select few routes are needed as output from it. The model uses several heuristics in eliminating the links. A few of them are presented below:
 - a. During extremely bad weather, drivers may be unwilling to divert. Therefore, wide-area diversion may be infeasible.
 - b. On some routes, bridges during icy weather can be very dangerous. If more traffic is diverted to such a bridge, adverse circumstances may arise.
 - c. Some of the routes may have very low capacities or extremely low speed limits, and it will not be prudent to divert freeway traffic onto it.
 - d. Some of the links may not possess Changeable Message Signs (CMS), because of which the drivers cannot be informed of the details of the incident and the diversion routes. This could lead to drivers losing their way in traffic and spending more time on the network, thereby exacerbating the congestion problem.
 - e. In most cases when the incident occurs during the peak hours, diversion volumes cannot be carried by other freeways nearby. This is because of the

fact that the other freeways are themselves congested. These are coded as rules in the expert system, and a reduced network is obtained.

5. The route generation process occurs next. This involves finding the routes on this reduced network from an origin to a destination. The algorithm for performing this function, simply enumerates all possible routes from O to D. It does not concern itself with the shortest paths, because the aim of the network generator is to just select a set of feasible routes. Some of the enumerated routes are then eliminated, because of their length or because of some other unwanted characteristics.
6. This task is to assign priorities to the selected routes based on the different characteristics of the generated routes, which are of critical importance to the user. This helps in providing the user with a set of routes, that addresses his concerns. It is hoped that this would lead to a higher user compliance rate.
7. The final step in the process is the assignment of the volumes entering the links to the routes generated in the previous step. The delays network-wide can then be studied based on this technique. It was decided to start with the use of the user-equilibrium method to study the delay patterns and then study the effects of different assignment techniques on the calculated delays. This leads to new volumes on the links and the traffic conditions network-wide changes. The process can be repeated once more using the new network conditions, as calculated by the assignment technique.

Another important aspect of the problem is the occurrence of concurrent incidents. Although this situation is rare, the impact on normal traffic flow is so high, that it warrants special mention. This case is usually handled differently by each incident management program, in the country. In NOVA, one important strategy used is that the incident with the highest impact on traffic is given preference over the others. The impact of the less important incident is considered only from the point of view of reduction of the feasible links to divert traffic.

Also important is the fact that of the above-mentioned steps only part of step 3 and 4, 5 and 6 is considered the network generator. The rest of the steps are necessary in order to study the effects of the network generator, on the delay caused to the entire network. The assignment of traffic is more to predict future flows than to actually, reduce the network. Steps 1 and 2 may be considered as precursors to the model, which are necessary to initiate the network generator.

3.9 Conclusions

It is clear that the current practices in diversion are in need of change. It should be made possible within the framework of IVHS and the growths in the area of information transfer. The expert system methodology could serve as an important medium for the implementation of a real-time diversion strategy. Based on the analytical framework developed, the network generator will be able to:

1. Generate a usable set of routes for diversion during single and multiple (concurrent) incidents,
2. Save computation costs for computation-intensive assignment algorithms,
3. Reduce the cognitive overload on the TMC operator, by providing him with a standard procedure to follow and anticipating potentially dangerous situations,
4. Ensure user compliance as they are considered in the decision-making process,
5. Constantly monitor the situation and develop alternate routing plans (whenever possible) due to changes in the traffic situation,
6. Serve as a training tool for inexperienced incident management personnel, by simulating different incidents and providing explanations for different actions.

It is important that the volumes are estimated for the different links at all times. This is one of the important requirements for any real-time system. The model also has potential to be expanded for other uses such as area-wide signal retiming and linking with a simulation model.

The implementation of the system involves knowledge engineering and problem representation. This is explained in detail in the next chapter.

4.0 IMPLEMENTATION DETAILS OF THE MODEL

4.1 Introduction

The previous chapter presented an overview of the methodology used in the process of developing a decision-support system. This chapter discusses the knowledge representation process and aspects of the user interface in greater detail. The user interface for each module is discussed along with that module itself. Also a test case is described along with each of the modules in this chapter. These are discussed in association with the analytical framework developed in the previous chapter.

4.2 Knowledge Representation

The next section deals entirely with the knowledge engineering process and knowledge representation. This is the stage where the knowledge acquired is transformed into a meaningful process, to help reduce the data-processing time for the operator. Seven main processes are involved in representing the framework for selection of alternate routes:

1. Duration estimation
2. Delay calculations
3. Link elimination
4. Route generation
5. Prioritization based on user choice
6. Equilibrium assignment
7. User interface

Each of these steps further involves child processes, which is either a set of rules, methods and functions or a combination of these. One of the important gains of using OOP technology in the representation, is the isolation of these processes. They are used only when necessary in the framework and their effects need not be felt in other frames. The reasoning process thus follows a context-based philosophy, i.e., only a certain set of rules is fired for a particular incident type and certain hidden images in the user interface become transparent to the user based on the stage of processing. This allows the

reasoning process to be more structured and explanation of the steps involved is made easier.

The first step as the model is initiated is the loading of the network data and initializing the different variables in the model. The network data is stored in dBase files, which can be easily accessed by KAPPA-PC. Any network can be input to the model, as long as the files follow a certain sequence of fields as requested by the model. Sample files are provided in the model. The links' data-base file has to have the following fields in order to read the input data:

- Back node
- Fore node
- Length
- Capacity
- Freeway
- Actual Volumes
- CMS
- Number of Lanes
- Special Flag (Bridge, Tunnel, etc.)

There are no restrictions on the size of the network for data input. The links are given names attached with numerals, (for *e.g.*, link_13) because these are all objects created dynamically and in real-time. Most of the variables store values in terms of lists. In this aspect KAPPA-PC is similar to Lisp.

The next step is the initialization where all the temporary variables are removed and counters are reset. This process is comparable to the "garbage collection" process that occurs in Lisp or C++. Once this is completed the operator is prompted to click on a box to initiate the process of monitoring an incident. An attempt has been made to allow the operator to point and click to navigate the program. This facilitates faster processing and reduces the work load on the operator. The links in the network can be displayed to the operator at the click of a button. This allows the operator to study the environments of the incidents and also understand the incident conditions better.

4.2.1 Incident Types and Duration

Seven different traffic incident types are considered in these analyses. Each of these has a clearance time associated with it. These are meant to be an estimate, based on previous

incidents of the same type. Small adjustments in the clearance times are made based on the location and the weather conditions during the occurrence of the incident. This module assumes the availability of stable information regarding the incident, based on verification by an agency in the FIM team. The clearance times based on the type of the incident is shown in the table below. These are based on the literature(18-22) and expert knowledge collected during the knowledge acquisition process. Some of these incidents follow a decision tree to determine duration. These are explained later in this section.

Table 4.1: Clearance Time Based on Incident type

Incident Type	Clearance Times Estimate (in hours)
Stalled Vehicle	0.5
Overtuned truck	2
Jack Knifed truck	1.5
Vehicle Fire	1
Traffic Collision	based on decision tree
Spilled Load	based on decision tree
Fatalities	based on decision tree
Special Event (Game, maintenance, etc.)	based on decision tree

4.2.1.1 Decision Tree for Collisions

The clearance time for traffic collisions depends to a large extent on whether personal injuries are involved or not and also the number of vehicles involved. The decision tree for collisions¹⁹ is shown below:

no. of vehicles involved < 3 and personal injuries: 1 hour

no. of vehicles involved < 3 and no personal injuries: 1/2 hour

no. of vehicles involved >= 3 and personal injuries: 1 hour

no. of vehicles involved >= 3 and no personal injuries: 1 and 1/2 hours

4.2.1.2 Decision Tree for Fatalities

Fatalities are incidents that involve considerable effort in the area of evidence collection. The responsibility for this rests on the police department. This is due to the fact that

incidents resulting in fatalities usually involve court cases, in which the police department is required to testify. This increases the clearance time of the incident to about 2 hours. This is further increased to three hours if the deceased is an employee of the incident management team.

4.2.1.3 Special Events

Special events are of two types:

1. *Planned lane closure*: because of construction or maintenance work, games, etc. For these incident types, there is evidence in literature²⁴ to show that staggering of work arrival times (Flex-Time) is a better option than diversion of traffic. The plans for Flex-time that have been prepared beforehand should be used if possible.
2. *Natural Disasters*: These include earthquakes, floods, etc., due to which the clearance time cannot be quantified¹⁹.

4.2.1.4 Spilled Loads

The important factors affecting the clearance time of a spilled load are whether the material is on fire, whether the material is hazardous and if the material can be easily identified¹⁹. These are important incidents and special care is needed to clear these incidents in a safe manner. This is because spilled loads can cause damages in numerous ways. They may be inflammable or hazardous (and in some cases even poisonous). In such cases a full-scale evacuation from nearby residential areas may be required. Also, these loads could seep into the ground and damage water-pipes or cause ground water damage. If such incidents occur in tunnels or bridges, extraordinary measures may be necessary before the link can be reopened to traffic. The decision tree for these incidents is shown in **figure 4.1**.

4.2.1.4 Other Effects on Incident Duration

Once the duration is estimated based on the incident type, corrections are made based on the location and the weather conditions as stated above. Even for a simple vehicle disablement (stalled vehicle) the clearance time could be higher by as much as half hour if

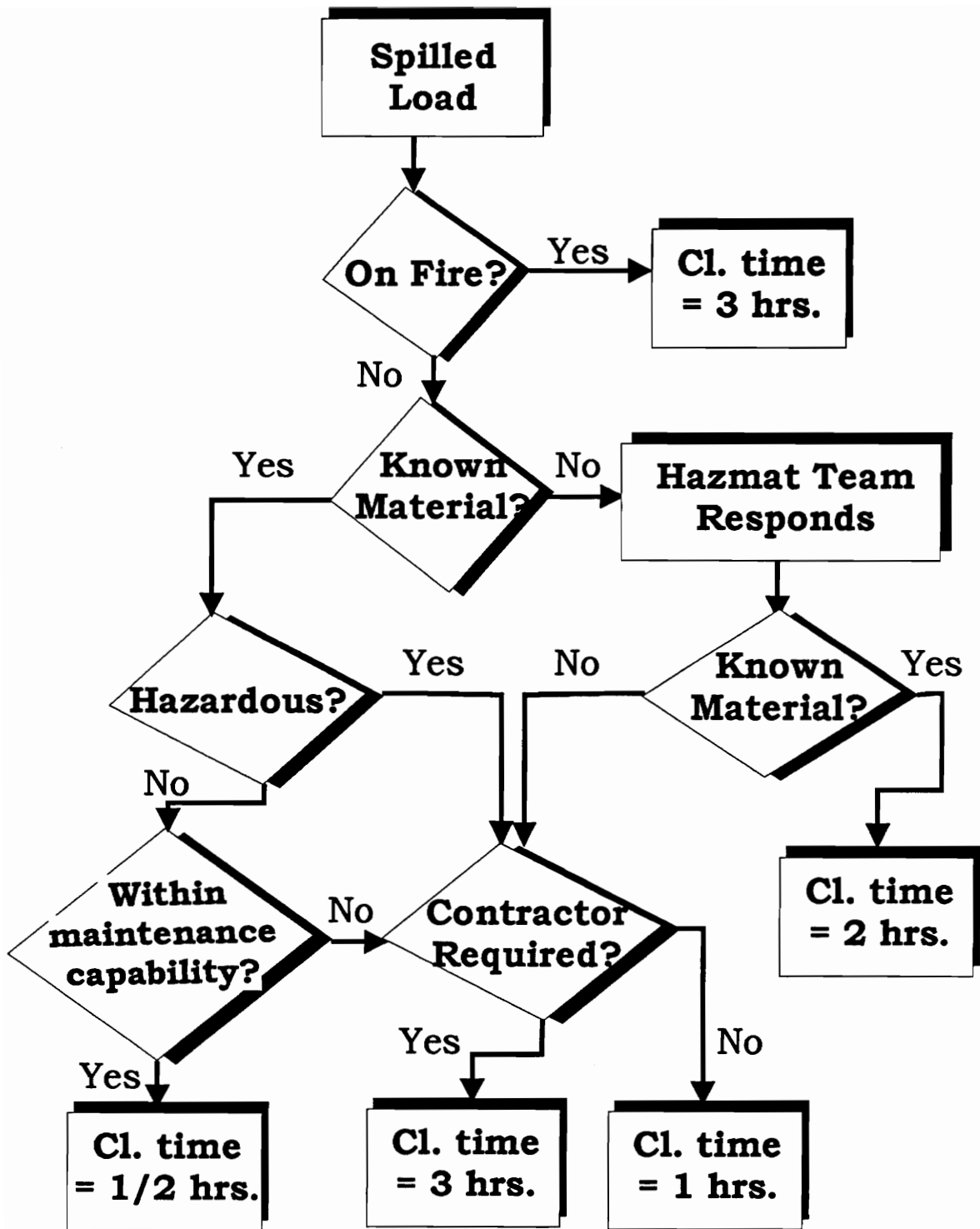


Fig. 4.1 CLEARANCE TIME for SPILLED LOADS

it occurs on a bridge. Similarly, a spilled load within a tunnel could necessitate a check of the air quality and a fire on a bridge could cause structural damage to the same. These cases need special consideration and rule antecedents ensure that appropriate adjustments are made.

Weather conditions also affect incident duration. The effect of adverse weather conditions is only to hinder the incident management team in carrying out its functions efficiently. Rules have been incorporated that account for the effect of weather on the incident duration. For *e.g.*, During foggy and icy conditions it is not advisable to divert, as the safety of the travelers is seriously compromised on unfamiliar routes.

4.2.1.5 User Interface for Clearance Time Module

The input associated with the clearance time calculations is called the Incident Input Data screen. It consists of a number of incident characteristics to which the operator can point and click to. The important characteristics for an incident are the incident type, link on which it occurs, weather conditions, time of occurrence and location. Most of the data input have fixed options and have built-in checks. For example: If the time of the incident is input as non-peak, but it is found that during the course of the incident peak-period flow will be affected, the time of occurrence is automatically changed to either the morning or evening peak. These are typically the duties of methods (demons), in the expert system. Once the incident data is input the user is prompted to calculate the incident duration by clicking on a button. Once the incident duration, is calculated it is displayed to the operator. The operator then has opportunity to override the software based upon other knowledge that he might possess. The option to override the software is provided at every important decision point. This helps the operator fine-tune the decisions. This screen is shown in **figure 4.2**.

4.2.1.6 Example

After loading the test network and initializing the different variables, the incident data is input into the computer. The incident shown on this screen is of spilled load type, and blocks two lanes on link_13 (freeway link) of the test network. It occurs during the morning peak hour on a weekday. The weather conditions are normal at the time of the

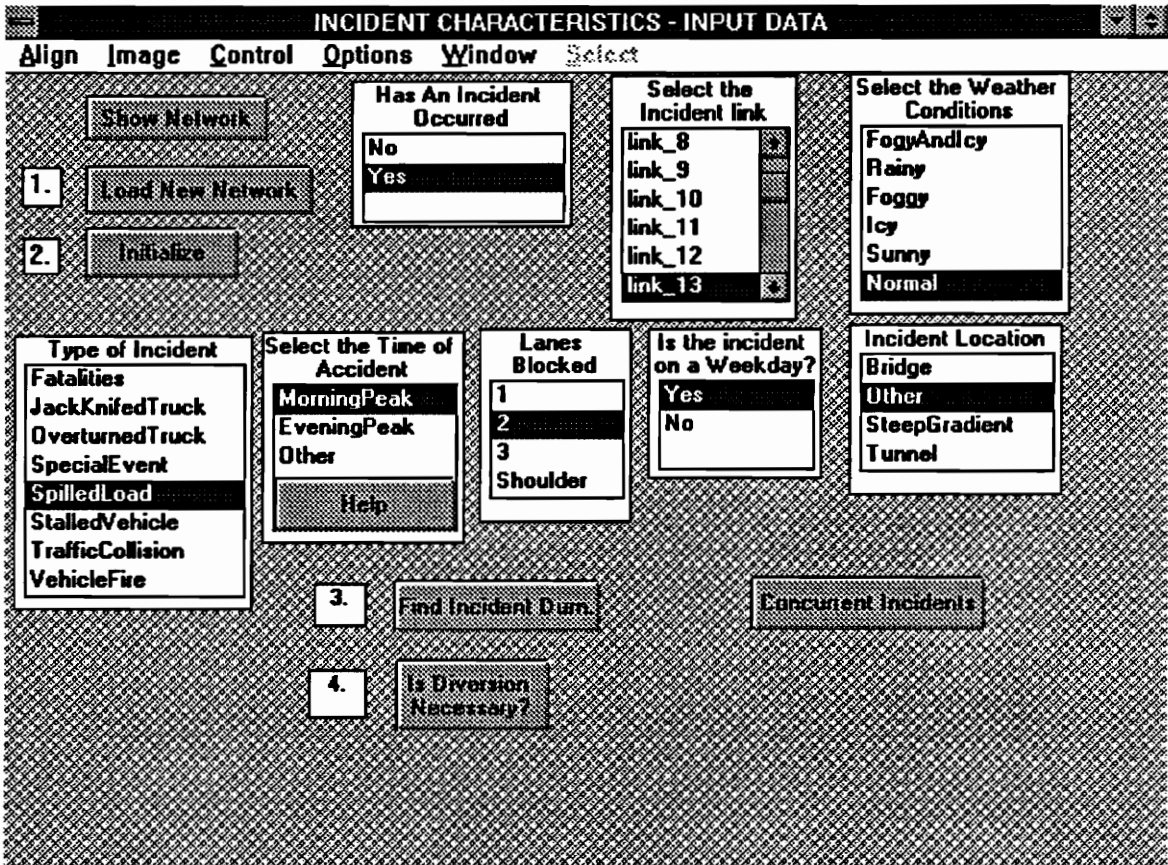


Fig. 4.2 Incident Input Data Screen

incident and the incident occurred on a normal stretch of roadway. The incident duration for this case was found to be three hours, based on the fact that the spilled material could not be identified by the incident management personnel and the hazardous material team had to respond. This will qualify as a very serious incident because it can delay two lanes of peak-period flow on a freeway for more than three hours. The model estimates that the delay caused by this incident requires the diversion process to be initiated.

4.4.2 Incident Delay Calculations

The next step is the quantification of the delays due to the incident and making the decision to start diversion. On the basis of the calculated incident duration and number of lanes closed for the incident clearance process, this module will calculate the volume of traffic that needs to be diverted and the delay caused in vehicle-hours.

Based on the duration of the incident and the time of occurrence, the first factor to be determined is whether the incident will affect either the morning or evening peak flow. An impact on peak period flow increases the costs associated with the incident considerably. This directly affects the need for proper management and diversion. The change in volumes in the network is handled by the use of "demons" that ascertain if the peak-direction flow is affected.

4.4.2.1 Capacity Reduction on Incident Link

Subsequently, the reduction in capacity of the link on which the incident occurs is estimated. This reduction is determined using a table developed by Jeff Lindley. The values are expressed either as a percentage of the total capacity remaining during an incident. The values for shoulder disablements, two and three-lane incidents and freeway cross sections for up to 16 lanes have been developed. The figures are shown in the table below:

Table 4.2: Fraction of Freeway Section Capacity Available Under Incident Conditions
(Source: Jeff A. Lindley, Transportation Research Record 1132)

# of freeway lanes in each direction	Shoulder Disablement	Shoulder Accident	Lanes Blocked		
			One	Two	Three
2	0.95	0.81	0.35	0	0
3	0.99	0.83	0.49	0.17	0
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.57	0.36
8	0.99	0.93	0.78	0.63	0.41

Locating the value of the reduction in capacity from the table and based on the actual flow in the link on which the incident occurred, an estimate of the volume to be diverted can be made. If there is no diversion process initiated all of this traffic backs up on the freeway, till either of the following happens:

1. The demand reduces because of the end of the peak period,
2. The incident is cleared and the capacity is back to normal,
3. Both of the above occur.

4.2.2.2 Delay Computation based on Deterministic Queuing

The sequence of events as shown in **figure 4.3** is described here. The initial capacity of the link initially is C_1 . The initial demand rate on the facility is D_1 . The capacity of the link drops to C_2 when the incident occurs. This is also known as the initial bottleneck flow rate. This activates congestion detection algorithms and other sources such as highway patrol, and surveillance elements into detecting an incident. Also, in response to congestion on the accident link, the demand on the link decreases to D_2 . This is called the revised demand flow rate. Then there is a "response" phase in which the efforts are concentrated in restoration of full freeway capacity. Till the end of the incident response phase, there is a discrepancy in the volume flowing in and the capacity available. After incident clearance, the capacity increases to the getaway flow that is the same as C_1 . The difference in time between this point and the point where the capacity finally equals

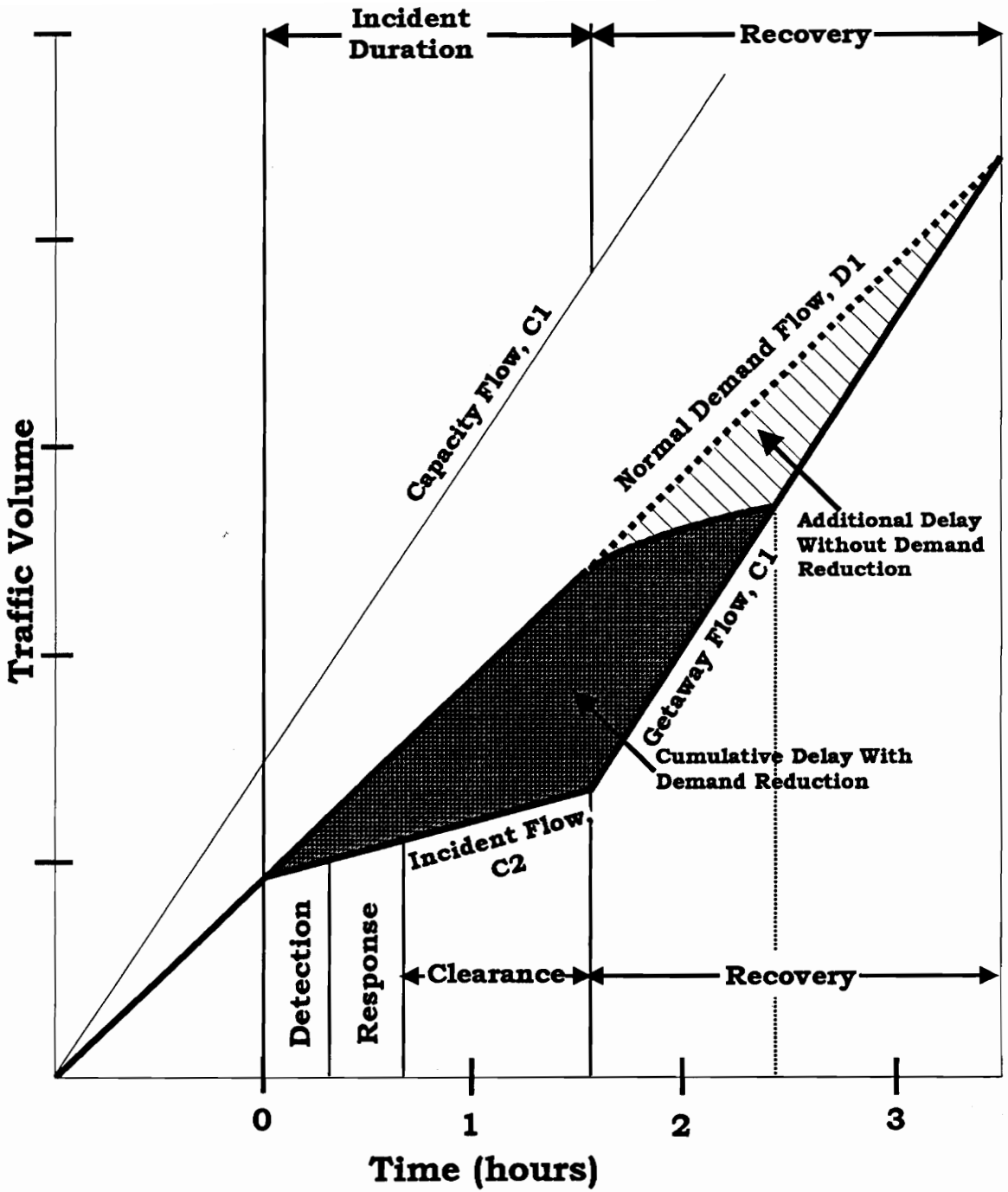


Fig. 4.3 DELAY CALCULATIONS

the normal demand on the freeway is called the "recovery" phase. The difference in area under the demand for traffic flow and the capacity of the link provides us with an estimate of the delay caused during the incident. This is the area shown shaded in the figure 4.3. Then based on simple calculations, the time to normal flow and the delay are calculated for the incident. Clearly to calculate the delays, freeway and bottleneck flow rates must be known and the revised demand rate must be estimated.

The use of the remaining capacity on the accident link cannot be overemphasized according to the experts. Since the delay for drivers who do not divert is usually greater than for diverted drivers, there should be incentive offered to drivers who use the incident link. The drivers who need to exit from the freeway, very close to the incident link are obvious choices to use this link.

4.2.2.3 Diversion Initiation

Once the delay is calculated, it is compared to the recurrent congestion on the network. If the delay is comparable to the recurrent congestion on the network, then the delay needs to be mitigated. The step of diversion during the peak hours is initiated based on the following rule:

If (delay caused by the incident) > 0.2 *(recurrent congestion on the network), **then** start the diversion process.

The recurrent congestion on the network can be estimated by the user-equilibrium assignment module. This module generates the volumes on the network based on the O-D tables. The volumes represent the network conditions on a normal day. The delays on each of these links can be added up to provide an estimate of the recurrent congestion network-wide.

Example

For the incident that was described in the previous section, the impact on the traffic is serious enough to start diversion. During the off-peak period it is implicit that diversion needs to be carried out. This is because off-peak periods typically are periods of low

traffic volumes on most other routes. Therefore, diversion would have an immediate positive impact. If diversion is started early enough, it might result in negligible delays on the network. The diversion is initiated based on the delays calculated in the previous step.

4.2.3 Network Generator

The need for the network generator has been discussed in detail in chapter 3. This module, which is aimed at reducing the computational difficulty, is discussed in this section. The different modules in the network generator are:

4.2.3.1 Link Elimination

In order to expedite the working of the traffic assignment process, it was decided to incorporate a module before the traffic assignment called the link elimination process. This uses real-time traffic conditions and the congestion levels on the different links in the network to eliminate certain links from further consideration. The contention is that links that are unusable from a diversion point of view need not be considered at all for the assignment step. This would help save computation times significantly.

Links are deemed as unsuitable for use in diversion based on different criteria²⁰. The most important among them being:

1. The links should have enough available capacity so that the level of service does not deteriorate significantly during the diversion process. The important measure of effectiveness used to monitor this is the volume-capacity (v/c) ratio
2. The links should also have a minimum absolute capacity. In some cases although links have low volume-capacity ratios, they are unable to carry any significant volumes because of low capacities,
3. In all cases, the volume on the link should not exceed the capacity, because of the volumes diverted on to the link. If the volumes exceed the capacities for all the links, the diversion process should be stopped. The process also ends if all the volumes to be diverted are carried by alternate routes.

The above mentioned constructs are treated as the constraints on diversion that can never be violated and are also called "**hard**" constraints. These heuristics are transformed into rules, based on the current flow patterns of the network. There are other criteria that are used for the elimination of the links:

1. In some cases there might be heavy traffic anticipated during the period of diversion, from an origin in the network. Examples of this are schools, offices, stadiums, ball-parks, etc. Traffic diverted to these links will impede the flow of future traffic. Therefore, diversion to these links may be detrimental unless information can be passed on to these trip generators. In Fairfax county in NOVA, the police department has access to many office buildings and can request them to stagger the work-hours of their employees.
2. There are safety considerations for the travelers such as the presence of bridges on a diversion route during icy conditions, the route passing through an unsafe neighborhood, etc. These are considerations that seriously impact user compliance and affects the diversion process.
3. The availability of the proper infrastructure in order to provide information to the travelers all along the diversion route must be guaranteed. For example: it may be impossible to provide Changeable Message Signs (CMS) to a certain link either due to a lack of infrastructure or lack of accessibility from the location of a resource.
4. The length of the route and the speed limit on the route should also be reasonable. Truck restrictions or high occupancy vehicle restrictions may be in force during the diversion process. Maintenance and construction activity may also hamper diversion.

The above-mentioned conditions are termed to be "**soft**" constraints because some or all of them may be violated if no diversion route can be found. Care should be taken in order to confirm that the operator knows the consequences of eliminating the effect of "soft" constraints. This is done in the expert system by providing the operator with warning messages, as the effects of the rules are nullified. In this manner these set of constraints are unlike the previous three listed.

If these set of rules are executed for all the links on the network, some of the links are eliminated from further consideration.

User Interface and Example for Diversion

For the user interface, the operator is shown a third window and is requested to select the origin and destination nodes for diversion. There are buttons available to eliminate routes and clicking on this will result in the set of rules for link elimination to be executed. Also available to the operator is a display of links that have been eliminated. These are classified according to whether they have been eliminated by hard or soft constraints. The operator can also browse through these links and determine why a link is unusable. The links that have been eliminated by hard constraints are colored in bright red, links eliminated by soft constraints are represented by a light red, while the links eligible for diversion are colored green. This can be seen in the **figure 4.4**.

4.2.3.2 Route Generation

A set of routes is generated from the origin to the destination based on the reduced network obtained from the previous step. This is done by developing a recursive search algorithm. The search algorithm is implemented by means of four functions. These represent the procedural programming part of the Expert System. An important factor to be considered is that, for any route from an origin to destination, the same node should not be passed more than once. The recursive algorithm for the search is shown in **figure 4.5**. The function can be explained as follows:

Let the Origin be 'O' and the destination be 'D'. The first step is to find all the links connected to 'O'. Next, each of these links is checked to see if they are eligible for diversion (this is already done by the rules). If the link is eligible to enter the diversion search space the algorithm shifts focus to the other node of this link, called the forward node here. If this node is flagged (i.e., has been visited before), the next eligible link is chosen. If the forward node is not flagged, it is checked to find if it is the destination node or not. If it is the destination node, the function has completed its task and the list of nodes and links along this route are stored. If this forward node is not the destination, the forward node is passed on as an argument to the same function till it is the destination node.

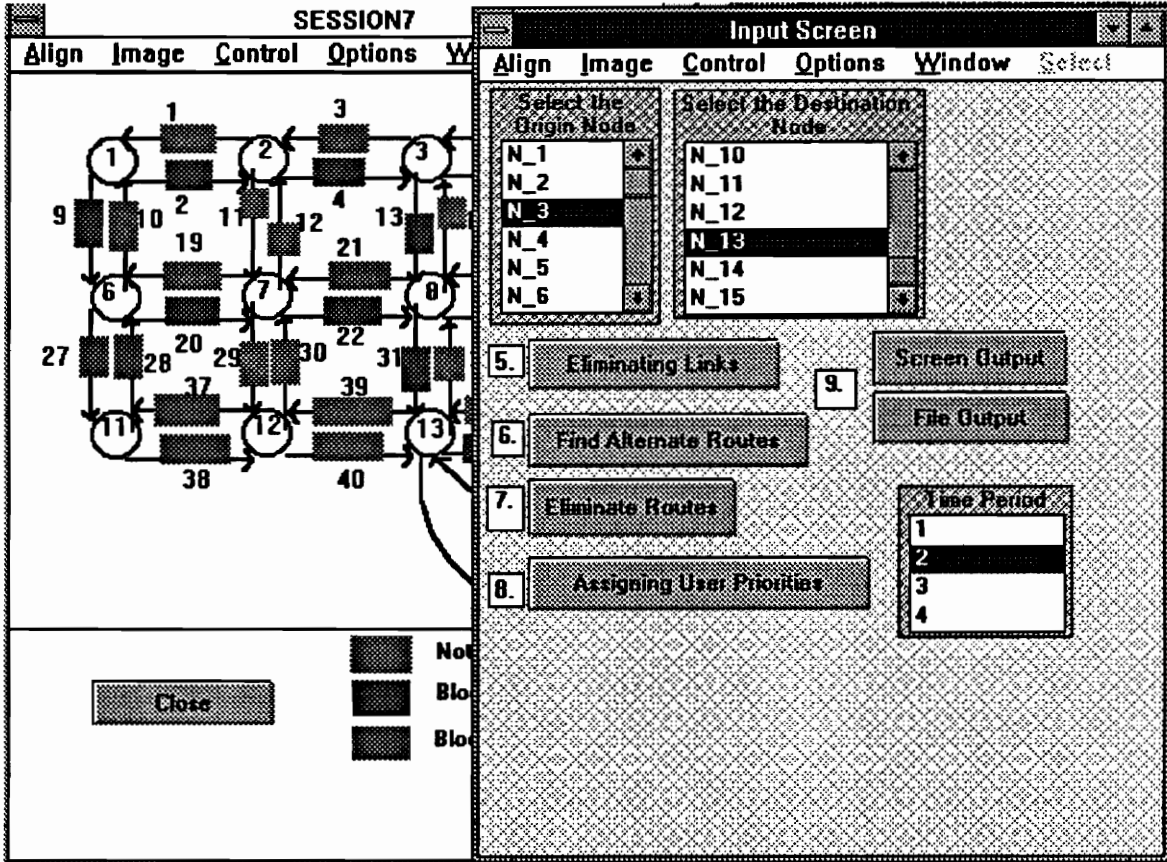


Fig. 4.4 Diversion Screen and Eliminated Links

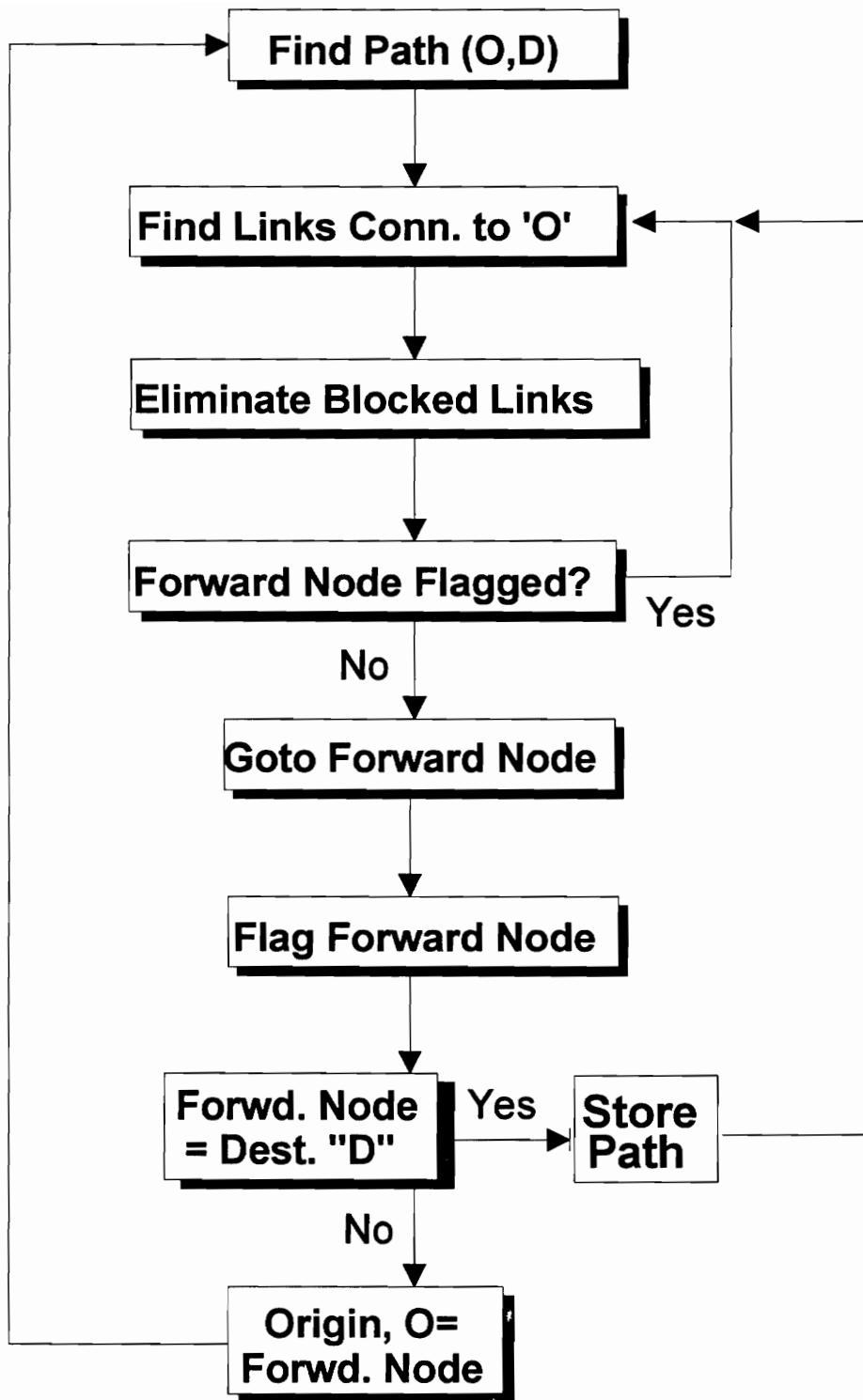


Fig 4.5 THE SEARCH ALGORITHM

This function eliminates many of the discrepancies in route generation (such as revisiting a node), by flagging the nodes that have been visited in each pass. There are other criteria that gain importance at this stage. One of the concerns is the non-availability of routes between the origin and destination nodes. The operator is given the option of browsing through the characteristics of each link in the network before it is decided that a certain set of links is unusable. Information about why a particular link was eliminated from further consideration is also provided to the operator. Thus the override capability is provided. If routes are still not found, the operator is given the option of eliminating the effects of the soft constraints. Ample warning regarding its effects is provided. For example: If the operator removes a school zone link from the blocked list, then a warning message that he should inform the school about his decision is provided.

In most cases, these steps result in routes becoming available if the search algorithm is executed again. If the network is highly congested there may still be no routes available for diversion. In such a case the operator is provided the option of "softening" the "hard" constraints. This refers to the reduction in the values of the thresholds for the hard constraints. The link elimination rules are then re-run to find the updated list of blocked links and the new routes found.

User Interface and Example for Route Generation

Due to the recurrent congestion problems in some networks, there is a possibility that routes cannot be generated at all between a particular origin-destination pair. The option to change the origin or the destination should be provided to the operator. A button appears on the screen, that allows the operator to change the origin and destination for the same incident. This is important in order to separate the effects of recurrent congestion from the effects of the incident itself. This also provides an opportunity to the operator to learn about the traffic characteristics of the network. The training of new personnel can also be done with these incidents as a base. This is an important element of any expert system tool and an area where it out-performs many analytical models. Also, the button is displayed only when there is a need for it, thereby reducing the load on the operator.

Nodes N_3 and N_15 are chosen as the origin and the destination respectively. For the incident case being described in this chapter, there are no routes available, even when the

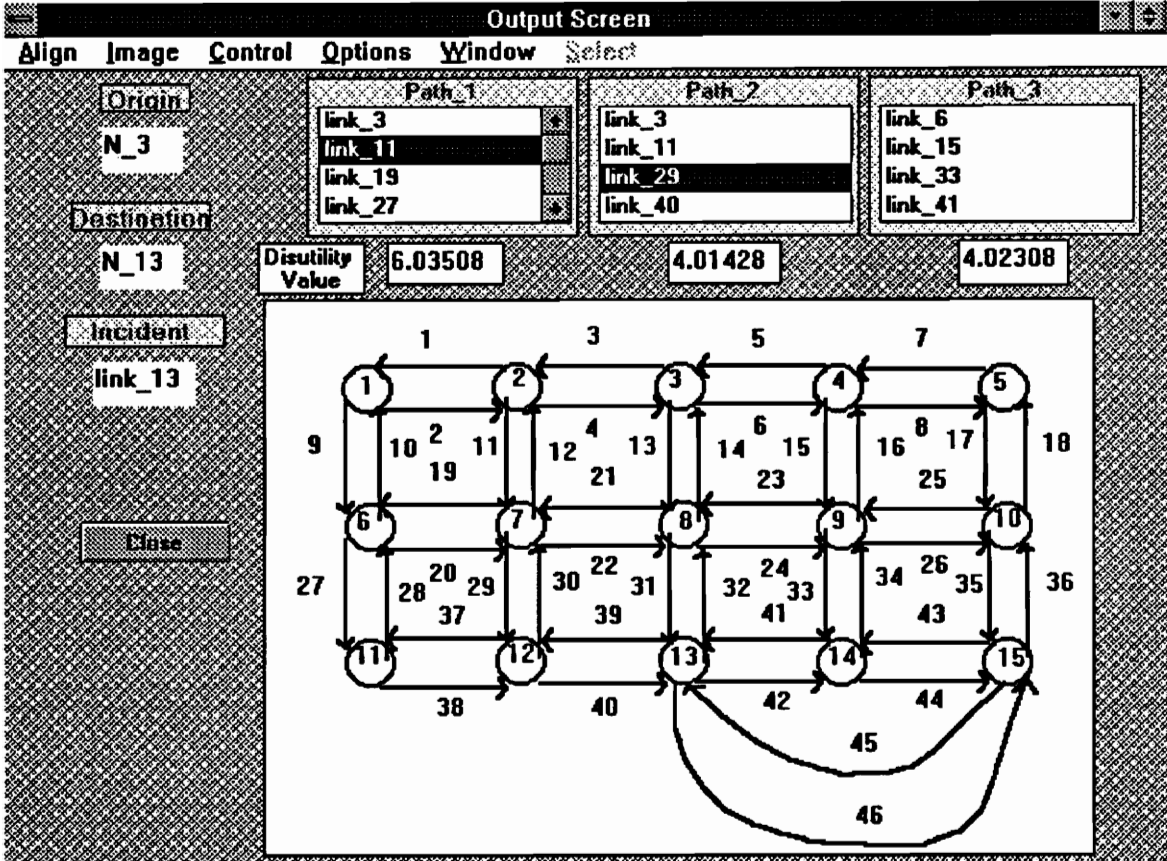


Fig. 4.6 Route Generation and Display

threshold values are changed. The main reason for this is the congestion that is typical of the links that enter N_15, during the morning peak hours. Therefore, it is necessary to change the origin destination pair. It was decided to change the destination node to N_13 instead. Consequently, there are three routes generated as shown in **figure 4.6**.

Diversion may continue for two or three hours, in most cases. If the severity of the incident is high, diversion routes are set up for several hours. The volumes diverted in the previous period directly affect the congestion levels on the links of the network. Therefore, it is necessary to specify which period of diversion it is, before eliminating links. Depending on the period the real-time characteristics of the network are updated. The volumes are predicted for each period based on the assignment module, which is to be described in the next few pages. The next step involves the prioritization of the selected routes.

4.2.3.3 Prioritization

The prioritization module is based on the user's point of view. The driver has disutilities attached to various characteristics of a route. The characteristics of alternate routes that are important to the user have been identified as:

1. Time spent on the freeway as compared to the time spent on arterial,
2. Length of the route,
3. Number of signals along the route, and
4. Number of left turns along the route.

The final disutility value attached to each route is a weighted average of all these factors on each route. The time spent on the freeway and the arterial are normalized values obtained by dividing these values with the total travel time along those routes. A weighted average of these values is then derived, with the time on the arterial being penalized around four times more than the travel time on the freeway. The length of the route is also normalized with the straight line distance between the origin and destination as the normalizing factor.

If during the course of prioritization, it is found that the number of selected routes is less than three, and there are some links eliminated due to soft constraints, the operator is given the option of removing the block on these links only for peak-periods. The route generation algorithm is executed once again in order to generate more alternate routes. Once the final set of routes is selected, the routes with the three lowest disutilities are presented to the user. The list of links for each of these three routes is presented to the user in a different screen. This screen also has the network displayed in it. The disutility associated to each of the routes are also displayed along with the routes (figure 4.6).

4.2.3.4 CMS Update and Signal Timings

A function has been developed that advises the operator regarding the links on which Changeable Message Signs need to be posted, in order to inform drivers about the incident. These signs are to be displayed at least on the links listed in this box. There may be other links on the network where the message needs to be displayed.

Since the lists of links along the selected routes are known, it is possible to check the signals along those particular routes, and change the timing plans according to new traffic volume along that link. If the signal timings are not changed to adapt to the new traffic volumes flowing on these links, a serious congestion problem might arise. Since large volumes are diverted on to these links, it could mean that this traffic might have significantly larger cycle lengths. These signals and link names for the CMS are displayed on a window within the system, as shown in **figure 4.7**.

4.2.4 Assignment Model

Although the network abstraction process has been done using expert systems, the problem of forecasting volumes on the selected diversion routes is done using a user equilibrium assignment. The incident delay characteristics, the enumeration of shortest paths from an origin to a destination and the assignment of traffic onto the selected routes are sub-problems best solved using procedural programming techniques. The Frank-Wolfe algorithm itself was chosen for the assignment module because of the ease of implementation and the rapid initial convergence of the algorithm.

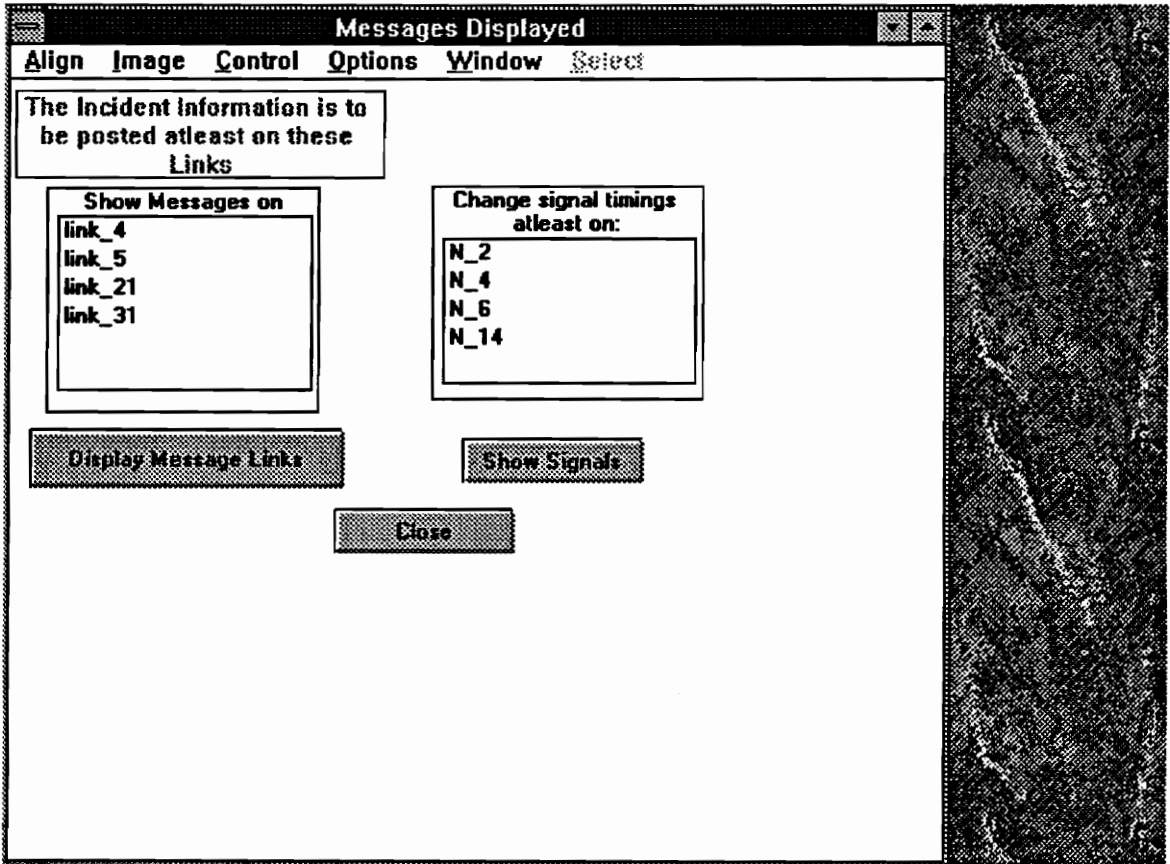


Fig. 4.7 CMS Display and Signal Retiming

4.2.4.1 Need for an Assignment Module

The volume to be diverted has already been estimated by the delay estimation function. These volumes are then redistributed over the selected routes using the Frank-Wolfe algorithm. The network wide delay is calculated and compared with the case when no diversion has been carried out. The volumes on the network should be known at any given time. This is because:

1. If along the diversion route congestion is very high and traffic on some other links is relatively less congested, the diversion routes can be changed. This module thus also serves the purpose of forecasting of future traffic conditions.
2. If concurrent incidents occur, an estimate of the current link volumes (after diversion) is necessary in order to develop diversion routes for the other incidents.
3. It would also provide us with an estimate of the actual delays experienced on the network with and without diversion and will help quantify the savings, resulting from use of the model.

It is assumed that in the future actual link volumes will be available in the network, but till then it will be necessary to simulate those conditions. The factors mentioned above imply that an assignment technique is necessary. The formulation of the user-equilibrium assignment module is discussed in greater detail below.

4.2.4.2 Linear Programming Formulation of the Algorithm

The convex combination algorithm was originally suggested by Frank and Wolfe, as a procedure for solving quadratic programming problems with linear constraints and is also known as the Frank-Wolfe (FW) algorithm. This method is especially useful for determining the equilibrium flows for transportation networks (40).

The FW algorithm is a feasible direction method. Unlike the general procedure for feasible direction methods, the bounding of the move size does not require a separate step. The bounding is accomplished as an integral part of the choice of descent direction. Consider the convex program:

Min $z(x)$

Subject to:

$$\sum_i h_{ij} X_i \geq b_j \quad \forall j \in J$$

The direction of the descent is determined not only on the basis of how steep each candidate direction is in the vicinity of the current feasible solution, but also according to how far it is possible to move along this direction. If, for example, only a small move is feasible in a certain direction, effort might be wasted in actually performing an iteration based on such a move. Even though the direction may be associated with a steep local decrease in the objective function, the overall reduction in the objective function from one feasible extreme point solution to another, may not be significant. On the other hand, a direction of change in which the local rate of improvement is modest but where the feasible region allows a considerable movement may achieve a larger reduction. The algorithm uses the direction that maximizes the drop.

To find a descent direction the algorithm looks at the entire feasible region for an auxiliary feasible solution, $\mathbf{y}^n = (y_1^n, \dots, y_n^n)$, such that the direction from the current solution \mathbf{x}^n to \mathbf{y}^n provides the maximum drop. The drop in the objective function is obtained by multiplying the slope of the objective function at the current feasible solution, multiplied by the distance between \mathbf{x}^n and \mathbf{y} . This is represented as:

$$-\nabla z(x^n) \cdot (y - x^n)^T / \|y - x^n\|$$

This drop is to be maximized. Alternatively, the expression can be multiplied by (-1) and minimized, resulting in the program

$$\min \nabla z(x^n) \cdot (y - x^n)^T / \|y - x^n\| = \sum_i (\partial z(x^n) / \partial x_i) \cdot (y_i - x_i^n) \quad \mathbf{[a]}$$

Subject to:

$$\sum_i h_{ij} X_i \geq b_j, \forall j \in J$$

Finding the descent direction, then involves a minimization of a linear program. This approach is based on finding a descent direction by minimizing a linear approximation to the function (instead of the function itself) at the current solution point. Minimizing this linearized function subject to a linear constraint set is a linear programming problem that has its solution at a corner of the feasible region.

To determine the step size, it is noted here that the new solution must lie between \mathbf{x}^n and \mathbf{y}^n (since \mathbf{y}^n being a solution of a linear program, lies at the boundary of the feasible region. The search for a descent direction automatically generates a bound for the line search by accounting for all the constraints (not only the binding ones) when the descent direction is determined. Since the search interval is bracketed, then, any one of the interval reduction methods would be suitable for the minimization of $z(\mathbf{x})$ along \mathbf{d}^n . The aim here is to solve:

$$\begin{aligned} \min z[\mathbf{x}^n + \alpha(\mathbf{y}^n - \mathbf{x}^n)] \\ \text{S.T. } 0 \leq \alpha \leq 1 \end{aligned} \quad [\text{b}]$$

Once the optimal solution of this line search is found, the next point can be generated with the step:

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \alpha_n (\mathbf{y}^n - \mathbf{x}^n) \quad [\text{c}]$$

The new solution is thus a convex combination (or weighted average) of \mathbf{x}^n and \mathbf{y}^n . The convergence criterion can be based on the similarity of two successive solutions or the reduction of the objective function values between successive iterations.

4.2.4.3 Algorithm for Computer Implementation

Given a current feasible solution, \mathbf{x}^n the n th iteration of the convex-combination method for finding equilibrium flows in a network can be summarized as follows:

Step 0: Initialization. Perform all-or nothing assignment based on $t_a = t_a(0)$, $\forall a$. This yields $\{x_a^1\}$. Set counter $n=1$.

Step 1: Update. Set $t_a^n = t_a(x_a^n)$, $\forall a$.

Step 2: Direction finding. Find \mathbf{y}^n that solves the linear program [a].

Step 2: Step-size determination. Find α_n that solves [b]

Step 3: Move. Set $\mathbf{x}^{n+1} = \mathbf{x}^n + \alpha_n (\mathbf{y}^n - \mathbf{x}^n)$

Step 4: Convergence Test. If $z(\mathbf{x}^n) - z(\mathbf{x}^{n+1}) \leq \kappa$, stop. Otherwise, let $n = n+1$ and goto Step 1

Starting with a feasible solution, \mathbf{x}^0 , the algorithm will converge after a finite number of iterations. The rate of convergence is very high for the first few iterations, but the rate drops as the number of iterations increases. This also implies that a very large number of iterations may be needed, in order to obtain a very accurate solution.

The code for the user-equilibrium assignment has been developed in "C". This is because the language KAL that is supported by KAPPA-PC does not have the capability to handle complicated algorithms such as assignment with high speeds. Some of the details of the computer implementation is presented here. After the initialization phase (step 0) in which an all-or-nothing assignment is performed and updating the travel times on the network, there is the task of direction finding for the next step. It is known as the bisection algorithm.

4.2.4.4 Bisection Algorithm

This algorithm is an interval reduction method and involves iterative procedures, in which each iteration is focused on a current interval (40). At each iteration this interval is examined and divided into two parts:

- one part where the minimum cannot lie and
- the current interval for the next iteration

The method of interval bisection exploits the fact that a ditonic function is monotonic on either side of the minimum. I.e., the derivative of the function to be minimized $dz(x)/dx$ is negative for $x < x^*$ and positive for $x > x^*$. The algorithm computes the derivative of $z(x)$ at the midpoint of the current interval, $[a^n, b^n]$. If the derivative at the midpoint < 0 , then the interval $[a^n, \text{midpoint}]$ can be discarded. The next current interval will be $[\text{midpoint}, b^n]$. If the derivative of the function at the midpoint > 0 , then the search is focused in the interval $[a^n, \text{midpoint}]$. The bisection method requires that the derivative of the function $z(x)$ be evaluated in every iteration. This may not be easy in some cases, and therefore this method should be used only if the calculation of the derivative is not much more difficult than the calculating the function itself. Interval reduction algorithms are typically terminated when the size of the interval of interest is less than a predetermined constant and this is true of the bisection algorithm also. This value of the line search is used in updating the volumes on the links of the network. The flowchart for this method (also known as the Bolzano search) is presented in **figure 4.8**.

Once these volumes are updated, it is checked to see if the convergence criterion for equilibrium flows is achieved. If it is achieved, the set of equilibrium flows are output, otherwise an all-or-nothing assignment is repeated. As part of this assignment the shortest path from the origin to the destination needs to be estimated.

4.2.4.5 Shortest Path Algorithm

The algorithm used here is a label-correcting method (40). It finds the shortest path from a given origin (root) node to all other nodes in the network. Accordingly, it has to be used for each origin in turn for a complete all-or-nothing assignment to be performed. The algorithm essentially scans the network nodes in an iterative manner. At each iteration the algorithm tries to find a path from the root to the node being scanned that is shorter than the current path. The algorithm terminates when no better path can be found from the root to any of the other nodes in the network.

The links are stored as a list of links identified by their end nodes. A travel time t_{ij} is associated with each link ij . In addition, two pieces of information are stored for every node i in the network:

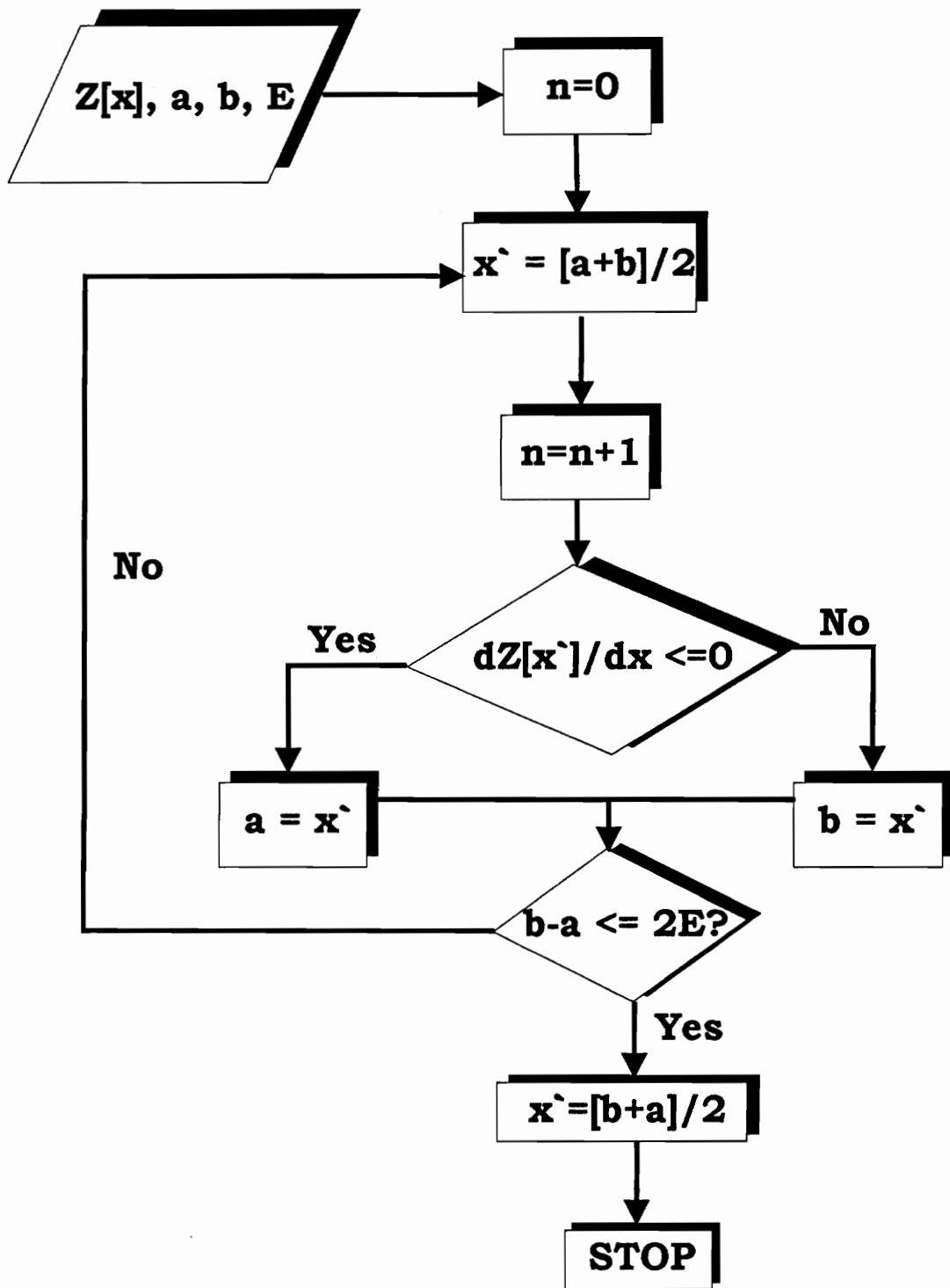


Fig. 4.8 BISECTION ALGORITHM

1. the (current) label of this node, l_j , and
2. its (current) predecessor node, p_j .

the label of node i is the distance from the root node i along the (current) shortest path and the predecessor of node i is the node just preceding node i along the (current) shortest path. A list of predecessor nodes is constantly updated so that the minimum paths can be traced once the algorithm is terminated. To help manage and keep track of the nodes, the algorithm uses an additional list called the sequence list. The list includes all the nodes that have yet to be examined as well as the nodes that require further examination.

The algorithm is initialized by setting all labels (which are arranged in a label list) to a very large number, setting all predecessor nodes (in the predecessor list) to zero, and placing the origin node r on the sequence list with label $l_r = 0$.

Each iteration starts with the selection of a node i , from the sequence list for examination. All nodes j that can be reached from i by traversing only a single link are tested in the examination process. If the minimum path to j through i is shorter than the previous path to j , then l_j is updated. This means that, if:

$$l_i + t_{ij} < l_j \quad (\mathbf{a})$$

then the current shortest path from the root node to j can be improved by going through node i . To reflect this change, the label list is updated by $l_j = l_i + t_{ij}$, the predecessor list is updated by setting $p_j = i$, and the sequence list is updated by adding j to it. Once all the nodes j that can be reached from i are tested, the examination of node i is complete and it is deleted from the sequence list. The algorithm terminates when the sequence list is empty. At this point, the shortest path from the root to any other can be found by tracing the predecessor list back to the root node. The flowchart of this algorithm is found in **figure 4.9**. The algorithm first places all the nodes that can be reached from the origin by traversing only a single link on the sequence list, and eliminates the origin itself from the list. The algorithm then keeps adding nodes to the list if the label test is met and deleting nodes that have already been examined. A node eliminated from the sequence list at an earlier stage can reappear later.

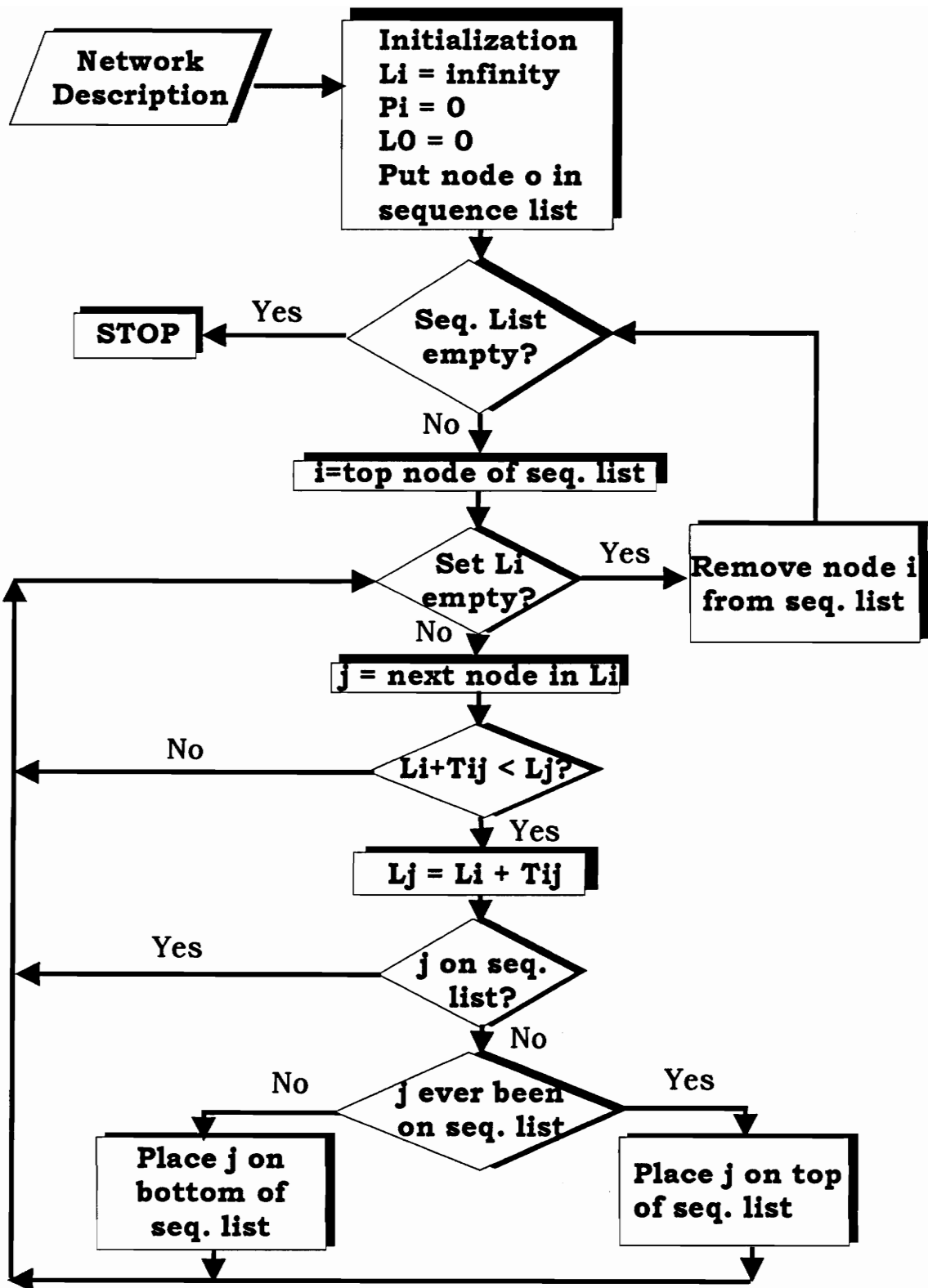


Fig. 4.9 SHORTEST PATH ALGORITHM

This assignment module provides the operator with the updated volumes on the links and also a prediction of the problems that can occur during the course of the diversion. These set of updated volumes are used for the next period of diversion (if found necessary). Links are eliminated on by the above mentioned rules based on the new congestion levels in the network. Some links that were previously congested may not be congested during this period. Therefore, the diversion routes may differ significantly from the routes in the first period. The module thus selects a set of feasible routes that also possess beneficial characteristics for diversion during each time period.

4.3 Conclusions

The analytical framework developed in the previous chapter has been transformed so that it can be used in an automated process as a computer program. The different aspects of each step and their implementation have been described in extensive detail. The next step is the validation of the model. Various case studies were used to test and validate the model.

5.0 RESULTS AND ANALYSIS

5.1 Introduction

The network generator module described in the previous chapters was applied to different incident cases that differed widely in their impact on the traffic. The network used and the list of inputs and outputs for the model is described in the initial section of this chapter. Further, the test cases are explained and the analysis of the results obtained.

5.2 Network Description

There were a number of small networks used to validate the rules and functions for the model. Care was taken to ensure that there were no repetitions or superfluous conditions in the knowledge base of the decision-support system. A more comprehensive network is needed to analyze the functioning of the entire model. The network used is described below.

The network consists of fifteen nodes and forty-six links. There are two freeways and an arterial network connecting the two freeways. Each of the links is one-way and has an average length of 1 mile. For real networks, it would be necessary to divide a link, which is longer than a mile into two links by introducing a dummy node. Although the results of the model are independent of the location of trip origins and trip destinations on the network, node 1 was chosen as the residential zone and node 15 was chosen as the Central Business District (CBD). This implies that the origin for trips during the morning peak-period is node 1 and the destination is node 15. The roles are exactly reversed for the evening peak-period. The highways have a speed limit of 65 miles per hour whereas the rest of the network has speed limits ranging from 25-45 miles per hour.

For this model, there are also other characteristics of the network that are of interest. Some of the nodes in the network have signals on them while some of the links have certain special conditions (a bridge, tunnel or height restrictions). The presence of trip producers/attractors (for e.g., schools, stadiums or malls) are also important factors for

the model. These factors are brought into focus under different conditions on the network.

This network is shown in **figure 5.1**. Although the size of the network is not very large, various insights into the working of the model can be obtained and the behavior of the model under extreme conditions (with respect to diversion) studied.

5.3 Inputs to the model

There are three types of inputs to the model:

- 1. Incident Data:** This includes information about:
 - a. Type of incident
 - b. Time of incident
 - c. Location
 - d. Weather conditions
 - e. Number of lanes blocked

- 2. Network Data:** This consists of:
 - a. Link Data: Forenode, Backnode, Capacity, Number of lanes and Shoulders, free-flow travel time, special features such as school zone, bridge, height and HOV restrictions, etc.
 - b. Node Data: Coordinates on the ground, signals.

- 3. Traffic Data:** Flow on different links, available capacities, maintenance and construction schedules, speed limits on different sections, etc.

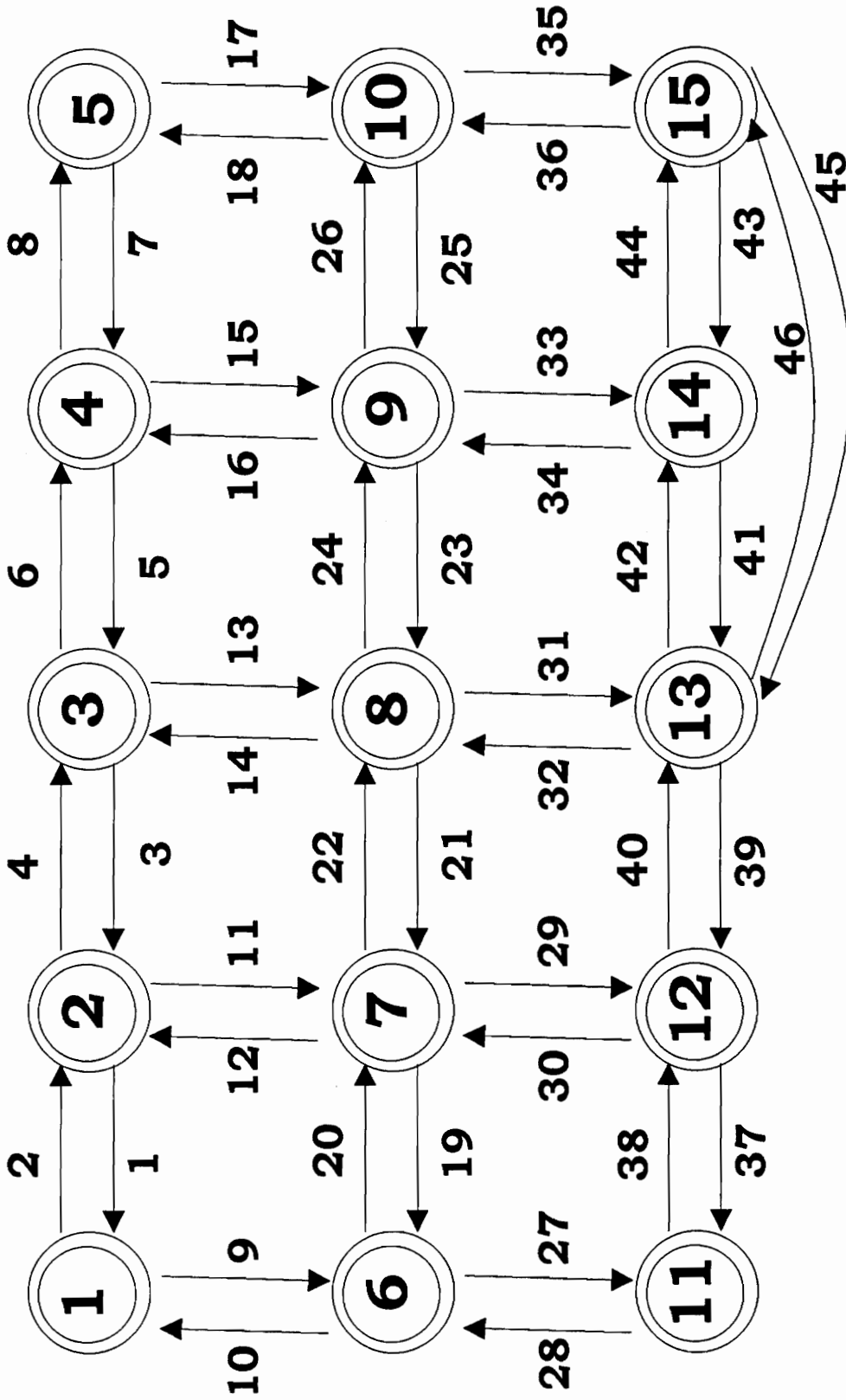


Fig. 5.1 TEST NETWORK

5.4 Outputs from the Model

The following are the outputs from the model:

1. Incident Duration Estimation
2. Delay Estimation
3. Diversion Volume Estimation
4. Alternate Routes selection and display
5. Warning operator of environmental hazards and problems during evacuation
6. List of links needing Changeable Message Sign messages
7. List of Signals along alternate routes requiring changes in timing plans
8. Prioritization of simultaneously occurring incidents
9. Maintenance of Incident log
10. Explanation facility provided to operator

The production and appropriate display of these results will help reduce the load on the operator and would significantly aid his decision-making process.

5.5 Test Cases

The test cases cover a wide range of conditions. Some of them are listed below:

- Incident types,
- Incident severity,
- Time of day,
- Location, and
- Weather.

During the morning peak there is an origin-destination flow of 7000 vehicle-trips from home to work (N_1 to N_15) and during the evening peak there are 6000 vehicle-trips (from N_15 to N_1).

The variation in the incidents' duration and the effect of the incident conditions in the selection of alternate routes is studied in detail here. The different cases for which the model was tested are:

5.5.1 Scenario 1

This involves a vehicle disablement on link_24 during the morning peak period on a weekday. Since it was on the right most lane the vehicle has managed to move to the shoulder. The weather conditions are normal.

Starting with this input, the model instantiates the rules for incident duration. The estimated duration for the incident clearance process is half an hour. The next step is to calculate the capacity reduction for the freeway link, which has two lanes in either direction. The reduction in flow is found to be five percent. Based on the actual link volumes that flow on the network (particularly on the incident link) it is necessary to calculate the delay caused by the incident.

The volumes flowing on the links are calculated based on origin-destination flows on the network. The user equilibrium assignment is used to calculate the actual flows. This entails a one-time cost. This is similar to using historic data in the analysis, but till such time that real-time information regarding traffic flows is made available to the model, they will serve the purpose. The link data is obtained by accessing the relevant database for the morning-peak period.

The delay caused by this shoulder disablement as calculated by the deterministic queuing model (87.5 vehicle-hours), was not found to be significant to warrant diversion. The model further suggests to the operator that the Changeable Message Signs should carry the duration for clearance of the incident and also inform the drivers that diversion may not be necessary for this incident. Also, message signs should request the drivers to refrain from “gawking”. The model suggests that this message should be displayed on links eleven, twenty and thirty. There may be other links where the operator deems fit to display the messages. There is no relevance of changing signal timing plans for this problem since no alternate routes are necessary.

From this example it can be seen that some incidents may not require diversion. This fact is even more relevant during peak periods in which most links are congested, and have low available capacities. Therefore, before the process of diversion is initiated, care should be taken to ensure that diversion will help. It may help to only issue advisory signals, when the impact on traffic is not significant.

5.5.2 Scenario 2

5.5.2.1 Case 1

The second incident case was on link_23 of the network on a weekday, during the evening peak-period. It involved an accident in which cars collided on the freeway link_23. The weather was rainy, and both lanes of the incident link were blocked by this incident.

The model requires further details about the incident before it can estimate the time for clearance of the incident. Other details regarding the number of vehicles involved in the collision and whether any of the occupants of the vehicles is injured, are requested from the user. For this incident there were two vehicles involved, and there were injuries to the drivers. Having obtained this information the duration for clearance of this incident is calculated as one hour. It is noteworthy that both lanes of the traffic were blocked because personal injury was involved in the incident. This is due to the safety personnel, who are trained to park their vehicles on the freeway in a manner that the safety of their workplace can be ensured.

Since this is the evening peak period and link_23 is in the peak flow direction, the volume flowing on this link is very high. The delay calculations on the link show that large volumes are backed up. The delay due to the traffic queuing on the link is obviously considerably high to initiate diversion. The volume to be diverted on the alternate routes is calculated as 2865 vehicles, based on the actual flow on the link.

This opens the second screen for the model. It deals with selection of the origin and destination for diversion. For this incident node N_9 is designated as the origin and node N_7 is the destination. The logic behind treating Node 9 as the origin is to start diversion upstream of the incident link. This has been identified as an effective diversion strategy.

The next step is the elimination of a set of links so that a reduced network can be generated. The rules eliminated the following links:

Table 5.1: Links that are not usable for diversion (Scenario 2; Case-1)

Elimination By	Eliminated Links
“Hard” Constraints	1, 3, 10, 14, 19, 21, 23, 25, 32, 36, and 45
“Soft” Constraints	7, 8, 15, 16, 17, 18, 27, 28, 37, 38

The hard constraints usually indicate the congestion levels on the links or the presence of better opportunities, whereas the soft constraints eliminate links due to a variety of reasons. Links eliminated by hard constraints cannot be used for that period, while links eliminated by soft constraints can be worked on. In many cases, even soft constraints may be binding, because traffic patterns associated with them are unalterable. As an example: Link_38 is expected to be a busy traffic zone in the near future, because of the presence of a school on the link. This generates trips to the home zone during the evening peak. The model recognizes that using Link_38 for diversion could result in a serious congestion problem. Therefore, this link is eliminated from the list of links available for diversion. If the operator is able to stagger the departure patterns for the school, by advising the concerned drivers that they would spend more time on the network by leaving immediately, than if they waited half an hour, this will be an alterable soft constraint. The operator thus accounts for the elimination of the effect of the soft constraint. Coordination, between the FIM team and offices & schools, is encouraged in NOVA. If this cannot be done, this soft constraint will be binding for the current time period.

The operator is provided with a facility to browse through the eliminated links and find out why a link has been eliminated. On the basis of further information that he might have, the operator may decide to override the software. For e.g., Since link_15 and link_16 do not have changeable message signs they are not considered any further. In this case, the operator is confident that a movable CMS (which is made available to the operators from the state’s Department of Transportation) can be sent to link_16. Thus, drivers on this link can receive updated information and the link is removed from the list of blocked links.

After this, the set of routes is to be generated. Using the path-finding algorithm, the model finds a set of routes on the reduced network, and displays them to the user. The set of routes that were generated for this case is listed below:

Route 1: Link_16, Link_5, Link_13, Link_31, Link_39, Link_30.

Route 2: Link_33, Link_41, Link_39, Link_30.

The two routes that were finally chosen have two links in common. This means that all the volume to be diverted will be finally carried by these two links. It may not always be possible to find routes that are completely independent of each other, especially during peak periods. Fortunately for this case, the current volume on those two links is relatively low. Also, due to this incident one freeway link carrying peak period flow is completely blocked. This results in considerable delays on the network, thereby justifying the use of alternate routes with common links.

The next step is route elimination, but in this case there were no routes eliminated. Finally, the routes are to be prioritized and displayed to the user. Since the number of routes generated is less than three, the operator is given the option of eliminating the effect of the softer constraints. This means that more routes can be generated for diversion. This option may be very useful during peak-period diversion, when not many links are available for diversion. If this option is utilized, the operator has to ensure that the soft constraints he overrides are accounted for.

There were two new routes generated based on the updated data for the current case. These new routes had significantly higher disutilities than the two presented currently to the user. Therefore, only the above mentioned routes were selected for diversion.

5.5.2.2 Case 2

It is important to analyze incidents during the off-peak, because the literature suggests that most incidents occur during the off-peak period. It is expected that savings in delay during this period, can reduce the overall delay caused by non-recurring congestion considerably. The same incident was simulated with all conditions other than the time of occurrence remaining a constant. The time of occurrence was changed to night-time. The

volumes on the network now are less than the evening peak. Therefore, the number of links that were eliminated from the network differs from the previous run. The links that were eliminated are:

Table 5.2: Links that are not usable for diversion (Scenario 2; Case-2)

Elimination By	Eliminated Links
“Hard” Constraints	3, 14, 23 and 32
“Soft” Constraints	7, 8, 15, 16, 17 and 18

The volume that needs to be diverted is also comparatively less. The best routes selected are quite different from the evening peak case:

Route 1: Link_16, Link_5, Link_13, Link_21.

Route 2: Link_33, Link_41, Link_39, Link_30.

Clearly the routes that have been generated are completely independent of one another. This will help in the quick dissemination of the queue. The disutilities attached to each of the routes is calculated and the best two routes have been chosen. These results suggest that diversion during off-peak periods could result in minimal delays in the network. Therefore, it may be very important to start the diversion process immediately after the occurrence of the incident, during the off-peak periods.

During the off-peak it is not necessary to find more than two routes to divert traffic on, because the diversion volumes are low. Therefore, the option to override soft constraints is unavailable to the operator. The delays on the network can be studied for both these cases. An interesting aspect of the problem is the difference in the delay characteristics between the same incident during the evening peak and the off-peak. This is discussed in the next section.

5.5.2.3 Assignment and Delay Calculations for Scenario 2

User-equilibrium of the diversion volumes was then carried out for the set of generated routes in each case. This provides information regarding the volumes on the different links, once diversion has been carried out. The delay caused by the diversion process can thus be estimated and compared to the case where diversion was not carried out at all. The results are also compared to the case in which it was assumed that all the drivers in the network had prior knowledge that two lanes of link_23 were blocked. This case is equivalent to applying the user-equilibrium technique on the network that has two lanes of link_23 blocked. There were significant differences observed between the cases. The results are as follows:

Table 5.3: Delay Values (in vehicle-hours) for Scenario 2

Time of Day	Recurrent Congestion	100% Information	Network Generator
Peak Hour	292	293	354
Non-Peak	6.6	34	16.7

Since the delay caused by the incident is very high during the peak period, the impact of the model is higher because the value of time saved during the peak period is very high. The difference in delay between the peak and non-peak periods is high. The Bureau of Public Roads (BPR) function used to calculate delays, is responsible to a significant extent, for this steep difference in delay values. This difference is because the BPR function employs a fourth power of the volume-capacity ratios. Therefore, they cease to represent the true conditions on the network.

In spite of this “fake congestion”, the delay caused by the incident has been contained considerably, by the diversion process. This is evident from the fact that the delay without diversion would have been 2000 vehicle-hours. Another indicator of the same fact is that the delays are also quite comparable to the case where the drivers have 100% percent information about the network when they start their trips. If all drivers knew all characteristics of possible alternate routes on the network, (which is almost impossible to achieve) they would have saved 20% more of the delay as compared to the network

generator. Thus, the network generator has contributed significantly in the delay reduction on the network.

5.5.3 Scenario 3

5.5.3.1 Case 1

The next incident case was on link_13 of the network on a weekday, during the morning peak period. It involved a spilled load on the freeway link. The weather was normal, and both lanes of the incident link were blocked by this incident.

An incident that involves a spilled load is treated with extreme caution by incident management teams, as stated in Chapter 4. Following the decision tree discussed therein, the user is queried as to whether the material is on fire and if the material was identified by the incident management team. Using the answers to these questions the duration for clearance of the incident is estimated to be two hours.

The next step is to estimate the reduction in capacity. Since the 2 lane freeway is completely blocked, there is no available capacity on the link. The diversion volume is estimated at 2450 vehicles for the first hour (period).

The second screen allows the user to input the origin and destination for the diversion of traffic. Initially N_3 was chosen as the origin and N_15 was chosen as the destination. In the link elimination stage the following links were found to be eliminated from further consideration.

Table 5.4: Links that are not usable for diversion (Scenario 3; Case-1, Period 1)

Elimination By	Eliminated Links
“Hard” Constraints	2, 4, 9, 13, 20, 22, 24, 26, 31, 35, 42, 44, and 46
“Soft” Constraints	7, 8, 15, 16, 17, 18, 27, 28, 37, and 38

The morning peak finds a number of links extremely congested. In spite of removing some links from the list of blocked links, it is impossible to find routes from N_3 to N_15.

The model executes each of the steps, i.e., the process of negating the effects of soft constraints, the changes in the threshold values, etc. No routes can be found even at this stage. This is because of the recurrent congestion problem on links 42, 44 and 46. Since these links are very highly congested, (with volume-capacity ratios near 1) there are no routes available. This emphasizes the fact that the decision support system developed here cannot completely automate the process of generating alternate routes for diversion. Some of the network specific heuristic data cannot be automated in the form of rules. This would be an area where the model currently developed, could be improved upon.

The option to change the origin and destination is provided to the operator if the model is unable to generate any routes for the current O-D pair. The origin is still retained but the destination is changed to node 13 for Case 2 of the network.

5.5.3.2 Case 2

As stated above, the same incident is used with a different O-D pair. The link elimination routine is rerun, after which the generation of alternate routes takes place. The path-finding algorithm selects routes (from this list of links) for possible use in diversion. After this stage is the route elimination process, in which excessively lengthy routes and the unsafe routes are eliminated. The prioritization of the remaining routes is attempted next and the model displays the following three routes for diversion. These routes are selected because they are assigned the least disutilities based on the drivers route choice algorithm.

Route 1: Link_3, Link_11, Link_29, Link_40.

Route 2: Link_6, Link_15, Link_33, Link_41.

Route 3: Link_3, Link_11, Link_19, Link_27, Link_38, Link_40.

The routes generated have very few links in common. These three routes are needed in this case, because the levels of congestion on the network are higher than for the evening peak. This is the time when the network experiences maximum congestion. Consequently each link has lesser available capacity.

Another portion of the output is the changes in signal timings and the CMS messages. The messages are to be displayed on links 4, 5, and 22. The signal timing plans for the signals on nodes 4, 6, 12 and 14 will need to be changed.

Then a user-equilibrium assignment is performed on these set of routes. This provides the information on the updated volumes on the different links on the network. Diversion during the second period of the incident clearance process, can use these updated volumes and the process of selection of alternate routes can be repeated. The link elimination routine is also rerun. The new set of links eliminated are:

Table 5.5: Links that are not usable for diversion (Scenario 3; Case-1, Period 2)

Elimination By	Eliminated Links
“Hard” Constraints	2, 4, 6, 9, 11, 13, 20, 22, 24, 26, 29, 31, 33, 35, 40, 41, 42, 44, and 46
“Soft” Constraints	7, 8, 15, 16, 17, 18, 27, 28, 37, and 38

It is seen that the number of links eliminated because of congestion levels increasing beyond a threshold is increasing. This is one of the effects of diversion. The diverted volumes from the first period are now carried by links that are forced to carry more than normal volumes. It is again attempted to find alternate routes on this reduced network. The operator has to use both options in order to do so i.e., negating the effect of soft constraints, and changing threshold values for the rules.

The routes generated are then prioritized and displayed to the operator:

Route 1: Link_6, Link_15, Link_23.

Route 2: Link_6, Link_8, Link_17, Link_25, Link_23.

These two routes vary considerably in their disutility values. Since there are no other routes available on the network, there are no options but to use these two routes in tandem. It may be possible to offer incentives to drivers to induce people to use the longer route.

5.5.3.3 Case 3

This incident is simulated under similar conditions except that the time of occurrence is the daytime off-peak. The incident has now occurred at 11:00 a.m. in the morning. Although link volumes are high as compared to the night time volumes, the volumes are still relatively less. The origin was N_3 and the destination this time was N_15.

The operator was able to generate alternate route during the off-peak and the problem of recurrent congestion on the links entering node 15 is not encountered here. The volume to be diverted is also lesser and is 1411 vehicles. The routes generated for the first period for this case are:

Route 1: Link_6, Link_15, Link_33 and Link_44.

Route 2: Link_6, Link_15, Link_26 and Link_35.

Route 3: Link_3, Link_11, Link_29, Link_40, Link_42 and Link_44.

These routes are independent of each other, except for link_44, which is common to routes 1 and 2. The volume to be diverted is also one that can be accommodated in two routes. Therefore, it is decided to use routes 2 and 3 for diversion of traffic.

The assignment is done on these routes and based on the new volumes generated, the alternate routes for the second period, are generated. These are as shown below:

Route 1: Link_6, Link_15, Link_33 and Link_44.

Route 2: Link_3, Link_11, Link_29, Link_40, Link_42 and Link_44.

It is shown here that a route not used during the previous period can be used for diverting traffic in the next period. This is an important result from the point of view of diversion. The availability of better routes at any point is detected by the model and presented to the user.

5.5.3.4 Assignment and Delay Calculations for Scenario 3

User-equilibrium assignment of the volume to be diverted, was carried out for the set of generated routes in each case. This provided information regarding the volumes on the different links once diversion has been carried out. The delay caused by the diversion process can thus be estimated and compared to the case where diversion was not carried out at all. The results are also compared to the case in which it was assumed that all the drivers in the network had prior knowledge that two lanes of link_13 were blocked. This case is equivalent to applying the user-equilibrium technique on the network that has two lanes of link_13 blocked. The results are as follows:

Table 5.6: Delay Values (in vehicle-hours) for Scenario 3 (Period 1)

Time of Day	Recurrent Congestion	100% Information	Network Generator
Peak Hour	292	257	291
Non-Peak	10	58	38

A similar analysis was also carried out using node 2 as the origin and node 8 as the destination. The delay for the peak-hour traffic was found to be close to 650 vehicle-hours during the first period. This is because of the small network used here. The model is unable to find good alternate routes originating from node 2. The model, therefore should be validated on a bigger network, which also has real-time/historic data about available traffic flow. The network generator is again able to contain the delays on the network. The delays are comparable to the case where the travelers are provided 100% information.

Table 5.7: Delay Values (in vehicle-hours) for Scenario 3 (Period 2)

Time of Day	Recurrent Congestion	100% Information	Network Generator
Peak Hour	292	700	391
Non-Peak	10	59	61

Providing 100% information to all the drivers, may not help reduce the delays. This is seen in the second period of diversion. This is because they have all decided the route to

follow to their destination, and many of them may follow the same route, causing congestion.

5.5.3.5 Case 4

This case deals with the incidents occurring concurrently. This happens quite rarely, but this could impose immense stress on the incident management team, if it is not prepared for it. In most cases not more than two incidents occur at the same time. The experts on the field have evolved a strategy, to combat this problem. They identify the most severe incident and concentrate on clearing that portion of the network. Then when the resources from this incident are freed, the other incident clearance process is started. This has been the idea used in the network generator, but the diversion process for the second incident is started.

Concurrent incidents are classified under two types:

- **Multiple Incident:** Although it occurs at the same time as the primary incident, it happens in a portion of the network that is distant from the primary incident. For this model, a multiple incident is treated as a different incident, which requires an alternate response and diversion plan.
- **Secondary Incident:** This occurs as a direct consequence of the initial incident. In most cases it occurs because of diversion volumes present on certain links of the network. The diversion to the incident link is now stopped, and the volumes on this link are moved out using the side streets.

When the spilled load incident occurs, a concurrent incident occurs on link_18. This is a vehicle collision, which involves 2 vehicles and no injuries for the drivers involved. The clearance time is estimated to be 0.5 hours, while the delay caused by the incident is calculated as 394 vehicle-hours. Since the spilled load incident causes greater delay than the collision incident, emphasis is placed on clearing the spilled load on link_13. There was no possible diversion route for the collision incident. For purposes of the spilled load incident, link_18 is designated to be blocked.

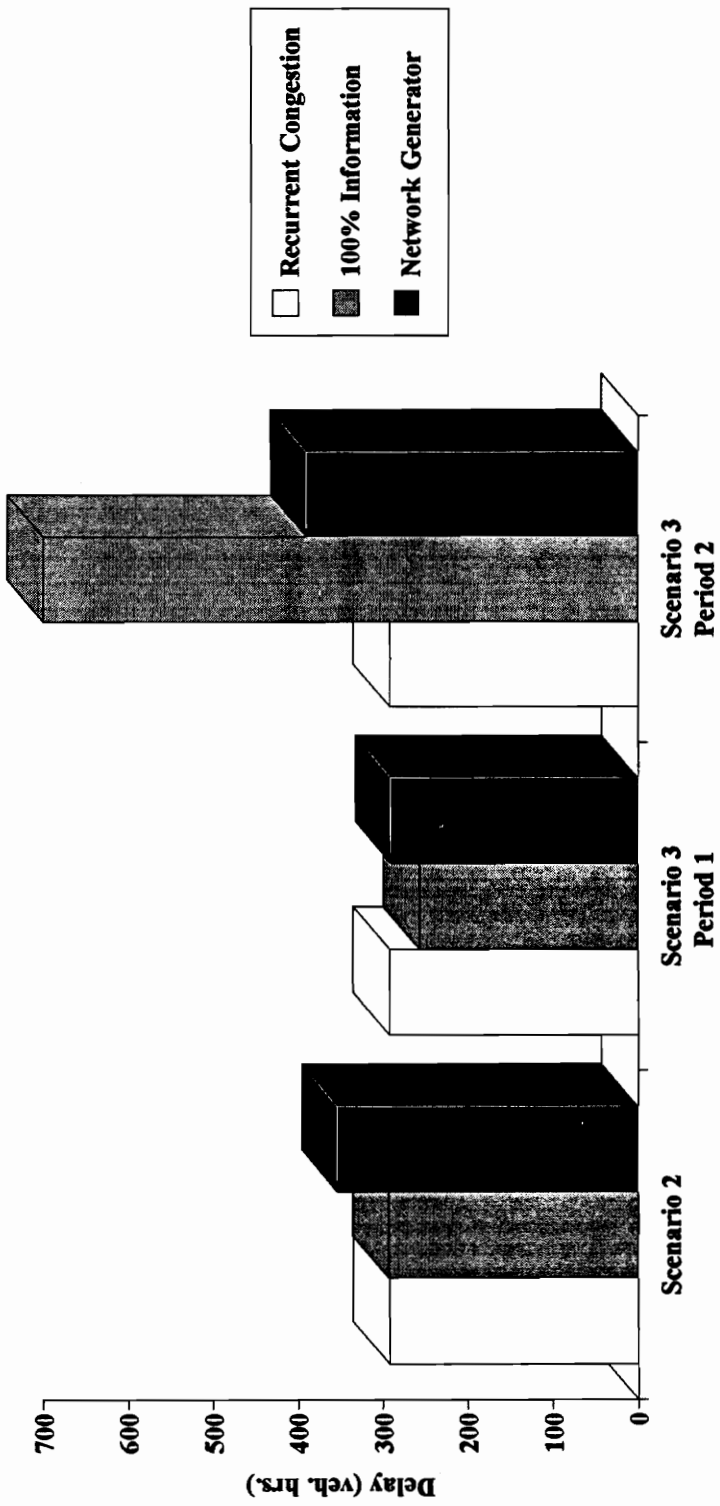


Fig. 5.2 Delay Characteristics (Peak Hour)

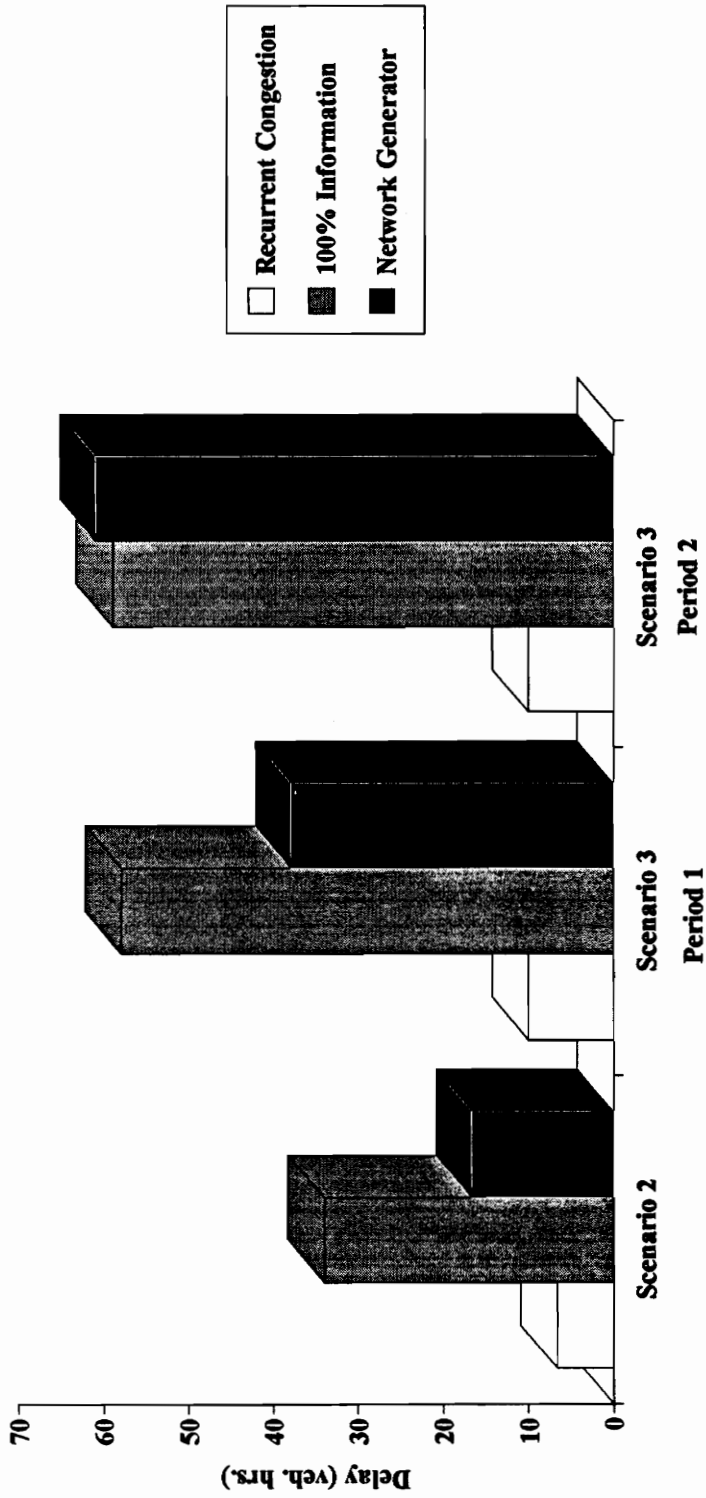


Fig 5.3 Delay Characteristics (Non-Peak Hour)

5.6 Conclusions

Diversion has been found to help, only when the delays caused by an incident are significantly high. This is especially true for incidents occurring in the peak periods. The model can handle diversion during both the peak and the off-peak periods. The results for the off-peak period show a significant difference. This is because there are more routes available to divert the traffic on. The delay characteristics for the off-peak and peak periods show that, on an urban network, there is a need to utilize the arterial capacity more efficiently.

It is found that the execution time for the user-equilibrium assignment is significantly lesser when the network generator is used. This has to be studied in greater detail for assignment techniques that are more computation intensive than the convex-combination algorithm.

The delay characteristics with respect to the peak and the non-peak can be studied from the graphs (**figure 5.2 and figure 5.3**). The delay saving is greater as the severity of the incident increases for the off-peak period. For the peak periods the delay savings increases with the severity above a certain threshold of delay. Below this value the network generator may not be so effective. The model selects the best possible routes available for any time period, based on the real-time characteristics.

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The state-of-the-art in diversion for incidents will not be able to handle the anticipated congestion problems, which is expected to grow. The analytical framework developed for the model within the context of ATMS has proved to be useful in reducing delays network-wide by selecting a set of feasible routes. This traffic management strategy has been implemented using a knowledge-based expert-system. This methodology has proved to be useful in handling the heuristic data associated with the model.

The model has demonstrated that the network-wide conditions can be used effectively in developing a strategy for diversion in real-time. The dichotomy between the operator's perspective and the driver's viewpoint, has been bridged to some extent. Tests on actual networks will show if user-compliance improves based on this model.

The results in the last chapter have shown us that:

- The model can handle multiple and concurrent incidents. This represents a significant improvement over diversion models developed elsewhere. The framework allows the diversion process to identify the diversion route based on real-time characteristics of the network.
- The model helps reduce the load on the operator, by providing timely warning messages and information about actions that are to be carried out. The network generator also provides an explanation facility, thereby serving as a good training tool, for new incident management personnel.
- As the severity of the incident increases the model significantly reduces the delays network-wide. This is consistent with the findings in literature.
- Routes generated during the off-peak usually have no common links. These completely independent routes help considerably in dissipating the traffic queues.
- The model works fairly well during the peak periods, and is able to find the best routes available.

- The analytical framework developed, identifies problems with the generated set of diversion routes, for the next time period. These problems are taken into account and a new set of routes may be generated if necessary.
- An operator is needed to supervise the working of the model. Currently, the process of generating the alternate routes has not been completely automated.
- Although the framework developed is independent of the network, some network-specific information is needed to calibrate the model. The thresholds for the different rules and functions are reset based on this information.

Although the model works well under most conditions, there are a few problems:

- The model should be validated on a larger network (preferably a real network). This involves significant effort in data collection. Also, the development version of KAPPA-PC used, cannot handle a huge network. Therefore, the model should be transferred to a better expert system shell.
- The “C” functions should be made entirely compatible with the code written in the expert system. One should be able to call these functions from within the expert system shell.
- The user equilibrium formulation, apart from being static, uses the BPR function, which is not very accurate when volume exceeds capacity. Therefore, this process may not be an ideal technique to use.
- Selection of origins and destinations is done by the operator. The knowledge of the expert in determining the origins and destinations should be examined carefully. The possibility of using multiple origins and destinations should also be studied.

6.2 Recommendations

The following are the recommendations for future research in this area:

- The process of selecting the origins (upstream of the incident) and destinations (downstream) for diversion needs to be studied. Automation of this feature would prove to be a significant improvement, for the model.

- The model should incorporate a better model to calculate the delays. The suggested model is a supply-demand curve. The scenarios with and without diversion, and using the static diversion maps can be modeled and the delays in each case calculated. This will help prove conclusively, the improvement as a result of using the network generator.
- The model should be transferred to a better expert system shell, which provides better features and easy integration with “C”. Nexpert Object would be a good candidate for this purpose. The use of a workstation would significantly aid the development process, especially because of the huge data requirements for the model.
- The assignment process used currently is static in nature. The use of a dynamic assignment model will represent the conditions more realistically. Also, this algorithm can be used to predict the flows on the network, after diversion has been carried out. The network generator helps in reducing computation times for the user-equilibrium assignment. The true computation time savings as well as the direct applicability of the reduced network (generated by the model) to the DTA should be studied carefully.
- The graphical user interface for displaying the network and the alternate routes can be improved. Nexpert-Object should also aid in the interface development process, through the use of the Application Programming Interface (API).

BIBLIOGRAPHY

1. IVHS AMERICA. "Strategic Plan for IVHS in the United States," May 20, 1992.
2. Hobeika A. G., "A Route Guidance System under Modified Network Conditions," Vehicle Navigation and Information Systems (VNIS) Conference Proceedings, Toronto, Ontario, Canada, 1989, pp. 170-175
3. Hobeika A. G., Subramaniam R. and Major. Ronald Miner, "A Framework to Model The Clearance of Freeway Incidents," Unpublished.
4. Federal Highway Administration, Demo Project no. 86, "Relieving Traffic Congestion through Incident Management- Wide Area Incident Management Expert System".
5. Subramaniam S., Ozbay K. and Hobeika A. G., "Implementation of a Real-Time Diversion Model for Advanced Traffic Management Systems," Presented at 1993 TRB Conference, Washington D.C.
6. Hobeika A. G., Sivanandan R., Ozbay K. and Subramaniam R, "Real-Time Diversion Strategies for Congested Urban Networks," Presented at the Second International Conference of IVHS America, Los Angeles, U.S.A, May 1992.
7. A. G. Hobeika and R. Subramaniam. Estimation of Delays Caused by Freeway Incidents to Initiate Real-Time Traffic Diversions. Proceedings of the 25th ISATA Conference, Florence, Italy, June 1992.
8. Hobeika A. G., Serali H. D., Sivanandan R., Ozbay K., Subramaniam R. and Zhang Y, "Real-Time Diversion Strategies for Congested Urban Networks-Phase 1 (1989), Phase 2 (1990) and Phase 3 (1991)," Project Report for the US DOT/VDOT UTCPC, Center for Transportation Research, Virginia Tech, Blacksburg, Virginia, 1991.
9. Hobeika A. G. and Ozbay K., "A Dynamic Traffic Assignment in the Context of Real-Time Driver Information Systems," Proceedings of the 24th ISATA Conference, Florence, Italy, 1991.
10. Hobeika A. G., and Zhang Y., "A Real-Time Traffic Diversion Model: A Conceptual Approach," Proceedings of the Second Conference on AATT, Minneapolis, Minnesota, 1991, pp. 393-397.
11. Hobeika A. G., Sivanandan R., Subramaniam R. and Ozbay K., "Real-Time Traffic Diversion Model: Conceptual Approach," ASCE Journal of Transportation Engineering, Vol. 119, No. 4 Jul./Aug. 1993, pp. 515-534.

12. Major Ronald Miner. Unpublished Interviews. Northern Virginia Freeway Incident Management Program.
13. Haghani A. E. and Daskin M. S., "Network Design Application of an Extraction Algorithm for Network Aggregation," *Transportation Research Record* 944, 1983.
14. Eash R. W., Chon K. S., Lee Y. J. and Boyce D. E., "Equilibrium Traffic Assignment on an Aggregated Highway Network for Sketch Planning," *Chicago Area Transportation Study*, 1983.
15. Jansen G. R. M. and Bovy P. H. L., "Spatial Aggregation Effects in Equilibrium and All-or-Nothing Assignments," prepared for TRB, January 1983.
16. Hearn D. W., "Practical and Theoretical Aspects of Aggregation Planning in Transportation Planning Models," *Transportation Planning Models*, pp. 258-287, Elsevier Science Publishers, 1984.
17. Taylor M. A. P., "Knowledge-Based Systems for Transport Network Analysis: A Fifth Generation perspective on Transport Network Problems," *Transportation Research*, Vol. 24A, 1990, No. 1, pp. 3-14.
18. Ritchie S. and Prosser N., "Development of a prototype real-time expert system for managing non-recurrent freeway congestion," *OECD Workshop on expert systems in Transportation*, Vol. 1, Finland, 1990, pp. 129-153.
19. Stack R. and Ritchie S., "Caltrans District 12 Real-Time Expert System for Freeway Incident Management," *Institute of Transportation Studies, University of California-Irvine*, 1993.
20. Ketselidou Z., "Potential use of Knowledge-Based Expert Systems for post-incident Traffic Control," Ph. D dissertation, University of Massachusetts at Amherst, 1989.
21. Gupta A., Maslanka V. and Spring G., "Development of a Prototype KBES in the Management of Congestion," Preprint No. 920302, Transportation Research Board, Washington D.C., 1992.
22. Garrison D. and Mannering F., "Assessing the traffic impacts of freeway incident and driver information," *ITE Journal* 60 (8) 1990, pp. 19-23.
23. Ben-Akiva M., dePalma A. and Kanaraglou P., "Dynamic Model of Peak-Period Traffic Congestion with Elastic Arrival Rates," *Transportation Science*, Vol. 20, No. 2, 1986.
24. Hendrickson C., Nagin D., Plank E., "Characteristics of travel time and dynamic user equilibrium for travel to work," *Proceedings of Eighth International Symposium on Transportation and Traffic Theory*, University of Toronto press, 1983, pp. 321-347.


25. Stamatiadis C. and Taylor L. W., "Travel Time Predictions for Dynamic Route Guidance with a Recursive Adaptive Algorithm," Preprint No. 940302, 73rd Annual Meeting, Transportation Research Board, Washington D.C 1994.
26. Stephanedes Y., Eil Kwon, Panos M., "On line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors," Transportation Research Record 1287, Transportation Research Board, Washington D.C., 1990, pp. 11-19.
27. Davis G., Kang J., "Estimating Destination-Specific Traffic Densities on Urban Freeways for Advanced Traffic Management," Transportation Research Board, Washington D.C., 1994.
28. Ullman G. L., Dudek C. L, Balke K. N., "Effect of Freeway Corridor Attributes upon Motorist Diversion Responses to Travel-Time Information," Preprint No. 940313, 73rd Annual Meeting, Transportation Research Board, Washington D.C 1994.
29. Heathington K., Worrall R., and Hoff G., "Attitudes and Behavior of Drivers Regarding Route Diversion," Highway Research Record, 3633, Highway Research Board, Washington D.C., 1971, pp. 18-23.
30. Mahmassani H., Caplice C., and Walton C., "Characteristics of Urban Commuter Behavior: Switching propensity and Use of Information," Preprint No. 890738, Transportation Research Board, Washington D.C., 1990.
31. Daniels E., Leven M., and McDermott J., "Improving Commercial Radio Traffic Reports in the Chicago area," Transportation Research Record 600, Transportation Research Board, Washington D.C., 1976, pp. 52-57.
32. Khattak A. J., Koppelman F. S. and Schofer J. L., "Stated Preferences for investigating Commuters' Diversion Propensity," 71st Annual Meeting, Transportation Research Board, Washington D.C 1992.
33. Cascetta E., Nuzzolo A., and Biggiero L., "Analysis and Modeling of Commuters' Departure Time and Route Choices in Urban Networks," Second International Capri Seminar on Urban Traffic Networks, July 1992.
34. A Regional Plan for Traffic Management on Northern Virginia Freeways. April, 1990, sponsored by VDOT, originally prepared by JHK
35. Roper D. H., Murphy R. J. and Zimowski R. F., "Alternate Route Planning: Successful Incident Traffic Management," TRB Special Report 153, Washington D.C. 1975 . pp. 141-144.

36. Roper D. H., "Freeway Incident Management," NCHRP Synthesis 156, Project 20-5 FY 1986 (Topic 18-08). Transportation Research Board, Washington D.C., December 1990.
37. TRB Committees on Freeway Operations, "Management of Freeway Emergencies and Special Events," Transportation Research Circular, No. 326, TRB, Washington D.C., December 1987.
38. Clive D. and Levitt R. E., "Knowledge Based Systems in Engineering," Mc-Graw Hill, New York, 1990.
39. James R. Slagle, "Artificial Intelligence: The Heuristic Programming Approach," Mc-Graw Hill Book Company, New York.
40. Yosef Sheffi, "Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods," Prentice-Hall Inc., New Jersey, 1985.
41. "Northern Virginia Freeway Management Team Operating Manual," originally prepared by JHK, 1989

VITA

Vijay Krishnaswamy was born on the eighth of February in Madras, India. He joined the Indian Institute of Technology, Madras in 1988 for a Bachelor's degree in Civil Engineering. He graduated in July, 1992.

In August, 1992 he joined the Department of Civil Engineering at Virginia Polytechnic Institute & State University, for a Master's with an emphasis in Transportation. He currently works with MFS Network Technologies, a leading firm in IVHS research, and would like to pursue a doctoral degree at a later stage.


07/29/94