A Framework for Dynamic Traffic Diversion During Non-Recurrent Congestion: Models and Algorithms

by

M.A. Kaan Ozbay

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Civil Engineering

APPROVED:

Dr. A.G. Hobeika, Chairman	Dr. H.D. Sherali, Co-Chairman
Dr. O. Balci	Dr. R. Sivanandan
Dr. I.L. Al-Qadi	Dr. P. Kachroo

April 30, 1996

Blacksburg, Virginia

Keywords: Diversion, Incident Management, Incident Duration Prediction, Expert Systems, Dynamic Traffic Routing, Feedback Control

A Framework for Dynamic Traffic Diversion during Non-Recurrent Congestion: Models and Algorithms

by

M. A. Kaan Ozbay

Civil Engineering

A. G. Hobeika, Chairman

Civil Engineering

H. D. Sherali, Co-Chairman

Industrial and Systems Engineering

(ABSTRACT)

Real-time control of traffic diversion during non-recurrent congestion continues to be a challenging topic. Especially, with the advent of Intelligent Transportation Systems (ITS), the need for models and algorithms that will control the diversion in real-time, responding to the current traffic conditions has become evident. Several researchers have tried to solve this on-line control problem by adopting different approaches such as, expert systems, feedback control, and mathematical programming.

In order to ensure the effectiveness of real-time traffic diversion, an implementation framework capable of predicting the impact of the incident on the traffic flow, generating feasible alternate routes in real-time, and controlling traffic in order to achieve a pre-set goal based on a system optimal or a user equilibrium concept is required. In this dissertation, a framework that would satisfy these requirements is adopted consisting of a

"diversion initiation module", a "diversion strategy planning module", and a "control and routing module" which determines the route guidance commands in real-time.

The incident duration data collected by the Northern Virginia incident management agencies is analyzed to determine major factors that affect the incident clearance duration. Next, prediction / decision trees are developed for different types of incidents. Based on the validation of these trees using the data that is not employed for the development of the trees, it is found that they perform well for the majority of the incidents. A simple deterministic queuing approach is used to predict the delays that will be caused by the incident for which the clearance duration is predicted using the prediction / decision trees.

The diversion strategy planning module, Network Generator, is developed as a knowledge based expert system that uses simple expert rules in conjunction with historical and real-time data to determine the incident impact zone, and to eliminate links that are not suitable for diversion. Finally, it generates alternate routes for diversion using this modified network. Network generator is tested using simulation on a small portion of the Fairfax network.

Finally, feedback control models for dynamic traffic routing models, both in distributed and lumped parameter settings, are developed. Methods for developing controllers for these models are also discussed. Two heuristic and analytic feedback controllers for the space discretized lumped parameter models are developed and their effectiveness for real-time traffic control is shown by simulating several scenarios on a simple network. An analytic feedback controller is also designed using a feedback linearization technique for the space discretized model. This controller also performed very well during simulations of various scenarios and proved to be an effective solution to this feedback control problem.

Acknowledgments

I sincerely express my gratitude to my committee chairman, Dr. Antoine Hobeika for inspiring, encouraging and guiding me through this dissertation. I am also extremely thankful to my committee co-chairman, Dr. Hanif Sherali, for his continuous support and encouragement during my studies.

I am extremely grateful to Dr. Pushkin Kachroo for his invaluable contributions to this work. Some parts of this work would not have been possible without his guidance, encouragement, and support. I would like to thank Dr. Osman Balci, Dr. R. Sivanandan, and Dr. Imad Al-Qadi for serving as committee members.

I would like to extend my appreciation to Mr. Sam Subramaniam for his contributions to the incident management project at the Center. I sincerely appreciate Mr. Arvind Narayanan, Mr. Yunlong Zhang, and Mr. Srikanth Jonnalagadda for working with me on different aspects of the incident management, especially on the development of the Wide-Area Incident Management Support System (WAIMSS) software.

Finally, I sincerely thank my wife, Ferzen, and my mother, Alev, for their continuous support, encouragement, and love.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Urban Congestion Problem

Urban traffic congestion has significantly increased in the recent decades. Traffic congestion by causing millions of hours of vehicle delays each year causes major losses in productivity and increases fuel use and environmental pollution. The monetary cost of congestion to the US economy has been estimated to be as high as $88 billion per year by 2005 (Institute of Transportation Engineers, 1992).

Urban freeway traffic congestion can be broadly categorized into recurrent and non-recurrent congestion. While the recurrent traffic congestion is due to high levels of traffic demand during the peak hours, the non-recurrent congestion is due to traffic incidents such as stalled vehicles, single and multiple vehicle accidents, spilled loads, construction operations, extreme weather conditions, or anything that causes a temporary reduction in the capacity of the roadway and an increase in delays. Non-recurrent congestion due to incidents are the major cause of freeway congestion. They account for over 60% of the traffic delays. This figure is likely to increase over the next ten years. (Cambridge Systematics, 1991).

Traffic incidents are generally random events that can greatly affect the traffic flow especially in the neighborhood of the incident link and cause long delays. Therefore, they need to be addressed in an effective and timely manner by the incident management teams. Thus, to coordinate the incident management process, many state DOTs formed incident management teams. These efforts have also been supported by the development of Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information

Systems (ATIS) or Advanced Traffic Management Information Systems (ATMIS) which include many surveillance, control, and information dissemination systems that aid in the detection, response, clearance, and handling of incident situations in the context of overall Intelligent Transportation Systems (ITS) research. Incident management as an important part of ATMIS helps in coordinating the various incident clearance agencies together so that they may perform as a team, thus expediting the clearance process and reducing delays caused by incidents.

1.2 Incident Management Process

Incident management refers to a set of strategies that are implemented to manage traffic flow and roadway infrastructure in the event of an incident. Incident management plans reduce the cost of the incident in terms of the delay and wasted fuel by hastening the recovery of the roadway capacity that was lost due to the incident, and by reducing the traffic demand on the facility through measures such as diversion, ramp metering, etc. Incident management typically involves the coordination of several agencies including Traffic Control Centers, DOTs, State and Local Police, Fire, Rescue, Hospitals, etc.

The incident management process consists of five sequential steps: incident detection, verification, response, clearance and recovery; each of which comprises a number of operations and coordinated decision making between the agencies involved. The provision of computer based support tools for the personnel involved, will help develop appropriate incident management strategies and increase the efficiency and expediency of the entire process. The components of incident management process can be summarized as follows:

Incident Detection and Verification: Major incidents are detected in less than 15 minutes. However , minor incidents may go unreported for more than 30 minutes or more. Although data collected by traffic sensors such as loop detectors can be used by automatic incident detection algorithms to detect most of the major and some minor incidents, one third to one half of reported incidents are detected by routine police patrols and the other half are reported to police from roadside call boxes, over citizen-band radio, and increasingly by mobile phones.

Incident Response: Although primarily state police handle most incidents on freeways, there is an increased awareness for coordinating the activities of other responding agencies such as fire and rescue, tow operators, ambulance services with the activities of police. Traffic management centers play an important role in coordinating the communication among these agencies. An increasing number of states and cities including Texas, Virginia, California and others, have developed or are in the process of developing incident response plans and strategies for tackling major incidents.

Clearance: Private tow truck operators clear the great majority of incidents. The most important aspect of the clearance operation is the correct diagnosis of the problem by the police and the request for the right tow equipment. While this is not a problem for minor incidents, it might constitute a problem for less frequent major accidents. The lack of expertise for handling large truck accidents can double or triple the time required to clear an incident.

Recovery: Recovery is the time from the complete clearance of the incident from the roadway and the restoration of physical capacity until the traffic flow reaches its pre-incident conditions. Recovery consists of three steps: restoring traffic flow at the site of the incident; preventing more traffic flowing into the area and getting trapped in the upstream queue; and preventing congestion from spilling across the metropolitan traffic network (Cambridge Systematics, 1991).

Congestion due to the incident can be minimized by the above actions through the use of effective traffic management strategies. These strategies include lane closings and openings, establishing and operating alternate routes and diversions, implementing traffic signal timing plans, and ensuring the safety of the incident personnel and victims. Unfortunately, traffic management is the least developed component of the overall incident management process and needs to be more carefully addressed for reducing the adverse effects of both minor and major freeway incidents..

1.3 Traffic Diversion

Diversion used in the context of freeway incident management operations refers to the real-time route guidance which is given to the motorists. Real-time route guidance differs from the diversion concept used in the planning concept in that it is not a long term policy decision concerning route choice, but rather, is a reaction to the immediate situation facing the motorists (Hall, 1974).

The main objective of a traffic diversion system is to allocate traffic in the network in order to obtain maximum utility . Route diversion is generally done in response to severe traffic incidents which cause substantial delays. Different diversion models have been developed in the past. The difficulties in the development and implementation of a route diversion strategy arise primarily from the inability to perform the following:

1. *Predict the impact of the incident on the traffic flow.* The lack of adequate data and the inexperience of the incident management personnel in the traffic management center (TMC), however, produce poor decisions especially during major incidents.

2. *Generate feasible alternate routes in real-time.* Currently, in the presence of a major incident, the incident management personnel at the TMC use pre-determined routes and strategies. However, sometimes these strategies and routes can be inadequate due to the real-time conditions. Therefore, they should be revised and appropriate adjustments in the diversion plan need to be made in real-time.
3. *Control traffic in order to achieve a pre-set goal such as a system optimal solution or a user equilibrium.* The volumes that need to be diverted to achieve one of the pre-set goals are generally predicted using static or quasi-dynamic traffic assignment techniques (Reiss et al., 1991). However, these static assignment methods are not suitable for time sensitive applications such as real-time traffic diversion. Moreover, the changes in traffic flow, travel times on alternate routes due to the control actions need to be continuously monitored and control actions should be revised in order to avoid instabilities and inconsistent route advisories that greatly reduce user compliance.

1.4 Problem Understanding

It is clear from the above discussions on incident management and traffic diversion, that the weakest element of existing incident management programs is the recovery stage, particularly the real-time development and implementation of the traffic diversion strategies. Although diversion strategies exist in many areas of the country, including the Northern Virginia area, these are pre-determined, static plans that are not capable of capturing real-time traffic conditions. In addition to the development and implementation problems discussed in the previous section, some of the known major shortcomings of the existing diversion strategies that are currently in use are as follows:

1. Since these plans are static and pre-determined, it is impossible or difficult to revise them in real-time.
2. They do not have, in general, capabilities for using the real-time data obtained by the new ITS technologies such as traffic sensors, probe vehicles, etc.
3. They are very limited in scope; in general they have one static alternate route for initiating a diversion after the occurrence of a severe incident on a limited number of freeway links (VDOT Northern Virginia Incident Management manual),
4. They do not have any dynamic models for controlling the traffic diversion. They have pre-determined messages such as "Congestion Ahead" or "Use the Alternate Route" that will be disseminated to the motorists via VMSs. However, they do not have any way of controlling these messages in a time-dependent manner to ensure the optimal diversion from the freeway to the alternate route(s).

Therefore, it is obvious that a dynamic traffic diversion / assignment / routing system is needed to develop real-time and effective diversion strategies and to control traffic during incidents.

1.5 The Scope of this Dissertation

The purpose of this dissertation is to develop a comprehensive framework that addresses all of the issues discussed in the previous section and to develop models that address each one of the problems cited in sections 1.3 and 1.4. At the heart of this framework there is the real-time traffic control model that uses a new feedback based approach for on-line traffic assignment / routing.

Traffic assignment has been one of the most important research topics in transportation engineering. For many years traffic assignment / routing models have been used only for planning oriented applications. These were static models since for long-term planning purposes there was not a clear need for time-dependent dynamic traffic assignment or routing. As a result of decades of research, different assignment models and solution algorithms have been developed (Sheffi, 1983). These models and algorithms are based on well established theories and they are very reliable. Although there have been several recent attempts to develop real-time algorithms that will perform on-line traffic control, a majority of the transportation research has focused on off-line planning problems.

However, with the advent of ITS, it is evident now that there is a need for real-time dynamic traffic assignment and routing (DTA / DTR) models that can be used in real-time traffic operations and control for developing and implementing effective diversion strategies. This problem presents grand modeling and algorithmic challenges to the researchers. Unfortunately, conventional static traffic assignment models and their dynamic models are not very useful for the purposes of real-time operations and control due to their modeling and computational limitations. The natural reaction of the transportation research community to this new challenge was to look at the well established static models and try to develop dynamic models that are extensions of these models by introducing time dimension and traffic dynamics into the existing models (Merchant and Nemhauser, 1972, Ho, 1980, Friesz et al., 1989). The efforts of these researchers substantially improved the understanding of the DTA problem. However, a unified model and well-accepted solution algorithm for this model could not be developed. Instead new DTA / DTR models that address some of the important problems of the existing models need to be designed in the context of real-time operations, an aspect that has been long neglected by the transportation researchers.

In this dissertation , we are proposing a unique approach and a framework for solving the DTA / DTR problem for on-line applications such as incident management and traffic control. This approach emphasizes the real-time on-line control problems and presents a DTA / DTR system that can provide realistic solutions to these problems. Our approach is specifically designed to solve real-time traffic management problems where the traffic diversion is used to alleviate the congestion due to the capacity reductions caused by the accidents.

1.5.1 Objectives

This dissertation mainly focuses on the diversion part of the recovery component of the overall incident management process. Other components of the incident management have been addressed by several researchers at Virginia Tech (Mastbrook., 1996, Subramaniam et al., 1995, Hobeika et al., 1994, Sherali et al., 1994, Sherali et al., 1996). This dissertation has two main objectives.

1. ***To formulate a practical framework for diversion for ATMIS***

For most of the incidents that are not severe enough to initiate diversion, the solution is to clear the incident as quickly as possible. On the other hand, diversion is shown to be an extremely effective way of dealing with severe incidents. Therefore, a dynamic and practical framework is needed to decide if the diversion is necessary based on the real-time conditions, and to develop real-time diversion strategies. These strategies include the location of diversion points, the determination of alternate routes for diversion, possible ways of controlling diversion using Variable Message Signs (VMS), Highway Advisory Radio, mainline and / or ramp metering, and the management of traffic signals to accommodate the traffic diverted to the alternate route(s).

It is clear that the development of a complete framework cannot be accomplished within the limitations of this dissertation. However, while certain components of this framework are emphasized and models and algorithms are developed for these components in the following chapters, for other components that are not studied in detail in this dissertation, the reader is referred to the work of other researchers working on these components. Therefore, the scope of the diversion framework described in this dissertation is to develop the following:

1. Incident Duration / Delay Prediction module.
2. Alternate Diversion Area / Route Generation Module (Heuristic Network Generator).
3. Real-Time Diversion Control Module.

2. *The Development of Models and Algorithms for each Module*

The models / algorithms developed for each module are a result of an effort for approaching the overall diversion process from a different perspective than the existing models in the literature. Different modeling and solution approaches have been selected in each module. The incident duration / delay prediction module uses real-world data and employs a simple prediction / decision tree approach to predict the incident durations. An enhanced version of the deterministic queuing model proposed by Morales (1986) is used to predict the delays caused by the incidents. The alternate route determination is achieved by developing a knowledge based expert system, named as "Heuristic Network Generator", that first determines the incident impact zone based on several factors such as the severity and type of the incident, and then eliminates links that are infeasible for diversion. Heuristic Network Generator then finds alternate routes using shortest path algorithms. Finally, the diversion is controlled in real-time using some new feedback control models and algorithms that are developed.

1.6 Dissertation Outline

This dissertation consists of 8 chapters. Chapter 2 provides an overview of the existing incident management systems, diversion models and dynamic traffic assignment and routing models. Chapter 3 presents a comprehensive framework for initiating and controlling traffic diversion during non-recurrent congestion and its components and their significance in this framework. Chapter 4 describes the methodology and data analysis used to develop the duration / delay prediction module, presents the prediction / decision trees for incident duration prediction and some validation results for the prediction / decision trees. Chapter 5 describes the Heuristic Network Generator, the knowledge based expert system developed for determining the incident impact zone and for prescribing alternate routes for diversion. Chapter 5 also briefly summarizes the software implementation of the Network Generator. Both Chapters 4 and 5 illustrate the working of the prediction / decision trees and the "Network Generator" by using a simple case study. Chapter 6 presents different ways of modeling the traffic system for developing different types of feedback controllers. Chapter 6 also presents two different heuristic ways of designing a feedback controller for a sample problem along with simulation results for three different scenarios. Chapter 7 describes the feedback linearization methodology for designing a feedback controller for the Ordinary Differential Equation (ODE) model proposed in Chapter 6 and presents the simulation results for several scenarios. Chapter 8 presents some concluding remarks and directions for further research.

2. Literature Review

2.1 Incident Management Systems

The development of incident management support systems has been attracting considerable attention over the past few years. These systems play an important role in aiding decision makers in a situation where they have to deal with cognitive overloads arising from the ill-structured nature of the problem, dynamic conditions and/or multiple operations.

Existing systems have employed the rule based reasoning capabilities of expert systems to present operators with high level analyses and recommendations concerning incident response. Ritchie (1990), discusses a knowledge-based decision support architecture for advanced traffic management. Incident management was an important component of the system. The proposed architecture employs a blackboard model to integrate knowledge sources at different agencies. The design envisions a network of real-time knowledge based expert systems (KBES) running on separate microprocessors at each agency. The design of this system provides decision support to traffic management personnel through four integrated modules - detection, verification, response planning, and recovery. Although this architecture addresses the temporal nature of the problem, it does not adequately address the spatial nature of the incident management process. Since the process in itself involves collecting and integrating large amounts of geographically referenced information such as, network information, agency locations and resources, the architecture is limited in its failure to design for the spatial component of the problem. In addition, the implementation details do not address the integration of the various agencies involved into a common platform to solve the problem.

In a subsequent paper, Ritchie et al (1991) describe a real-time decision support system for freeway incident management and control. The system, Freeway Real-Time Expert System Demonstration (FRED) currently operates on a simulated freeway network and addresses important decision support components such as, 1) management teams to be present at incident sites, 2) posting static alternate routes and 3) ramp metering. However, the emphasis of this system being on incident detection and verification, it does not provide adequate support for response plan generation. The resources needed for clearance of incidents are not recommended by the system. In addition, the system fails to consider dynamic diversion routes. This further limits the system because pre-defined alternate routes may not always be suitable for diversion. Its application for a comprehensive network is also overlooked. This is probably because of the data handling restrictions imposed by the expert system shell used (G2).

Gartner and Reiss (1987) discuss a traffic control system design for New York, that can be used for diversion, ramp metering and signal re-timing. However, this system is not comprehensive for incident management as it does not address other issues such as incident clearance, duration and delay prediction. Further, like the other two systems discussed above, they too neglect the spatial and coordination aspects of the problem.

The system developed by Siegfried et al. (1993) recognizes the spatial aspect of incident management by using an Geographic Information System (GIS) in the development of an automated incident management plan for the Houston area. The study evaluates the use of a GIS to relate incident locations with the transportation network and to make decisions and calculations for incident management. By using a GIS development platform, it is possible to develop interrelated maps, databases, and incident management applications. This platform also provides the ability to integrate other software like

PASSER-II and CAMEO, for data sharing and execution, through a common user-friendly interface. The prototype applications developed for the automated incident management plan are grouped into two. The first type pertains to incident management operations like alternate routing, incident response and resource-management. The second type pertains to planning and analysis for incident management, like network and incident information query. The prototype applications are developed using PC-Arc/Info on a PC/486 microcomputer. An important finding of this study is that a PC platform may not be suitable for real-time traffic management systems, using technologies such as automatic vehicle identification for traffic monitoring, where new data is received at short intervals of few seconds. The study strongly recommends the use of a workstation environment and the use of a programming language that parallels a higher level programming language.

By using a GIS as the development platform, the prototype application of Siegfried et al. (1993) aids decision making by providing facilities for querying and manipulating large volumes of geographically referenced data. However, the application lacks tools for high level analysis of this data using procedural and rule based reasoning. The overall architecture also fails to recognize the importance of coordinated decision making and information sharing between several agencies involved in the incident management process.

In summary, existing systems address some specific aspects of the incident management process using appropriate tools and methodologies. However, no system comprehensively addresses the processing of temporal and spatial information by a group of decision making agencies. Also, they do not have well-developed models and algorithms to support real-time diversion and traffic control.

2.2 Traffic Diversion

Real-time diversion and routing of traffic is one of the efficient ways of relieving non-recurrent traffic congestion. Several models have been developed for determining diversion routes and diverting the traffic onto these routes. Expert Systems, feedback control, mathematical programming models are among the approaches which have been used for developing real-time diversion and routing strategies.

The literature review reveals that there are several models of different complexity that are used to determine diversion routes for real-time applications. One of the first comprehensive studies on diversion is conducted by Hall (1974). This study explore the feasibility of real-time diversion in an intercity corridor between Baltimore and Washington D.C. Hall uses simulation to evaluate several diversion strategies for different scenarios of various states of the system. He proposes the use of either SCOT (Simulation of Corridor Traffic) or New Jersey Turnpike model developed by Sperry Rand Corporation for analyzing these different diversion scenarios. Among the system features he simulates are:

- Surveillance system including traffic detectors, aerial surveillance, closed circuit TV, call boxes and telephones, and highway patrol.
- Communication system.
- Variable message signs that will be used to disseminate congestion length, cause and alternate route information.
- Traffic control (management) center.
- Different incident scenarios such as the closure of the Baltimore / Washington Parkway at different points of Parkway for different levels of demand.

As a result of this study, Hall concludes that real-time diversions can provide a more comfortable driving environment and minimize the consequences of incidents. He also points out to some perceived drawbacks such as the difficulty of developing diversion plans in real-time and the cost of implementing these plans.

The diversion model developed by Berger et al. (1976) is also specific to the Baltimore beltway. He uses Sperry developed event-scanning traffic simulation based on the hydrodynamic traffic model with the capability of simulating multi-roadway freeway systems. They propose a simple control mechanism that tries to minimize delay difference, or delay rate difference, at alternate routes by switching the diversion messages on and off. This system suffers from the weakness that there exists only one alternative route that can be used for the entire beltway, in the event of an incident of sufficient severity on any link of the beltway. Further the controller proposed is heuristic in nature and there is no, guarantee that it will work under congested conditions .

Roper et al. (1984) describe a traffic management plan at Los Angeles aimed at bringing about voluntary traffic diversion upstream of a section of freeway that need to be closed for 6 hours for maintenance operations. A diversion plan that is specifically prepared for this closure has been implemented. Authors report that the operation is carried out smoothly, with no major problems. Although, this specific example shows that the pre-planning works for managing major special events that are known before they occur, it does not address the issue of generating diversion plans in real-time.

Zoe Ketselidou presents an expert system model for post-incident traffic control (1989). The model developed is based on pre-determined 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 best diversion strategy is chosen to be the one that balances the unused capacity according to three rules listed in order of decreasing priority:

- Wherever possible, the link demand will be kept below the capacity for all links in the system.
- 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.
- Diversion will be terminated when neither of the above rules is violated.

Other criteria that are employed for selection of alternate routes include proximity to main corridor routes, usefulness for access to ultimate destination, driving quality on route, impact on adjoining land-use, and jurisdictional problems.

The system is evaluated and tested by comparing the results obtained from the expert system and a simulation package called TRAFLO, which was developed by the FHWA. The Long Island, New York network which is basically a freeway corridor and which offers great potential for alternate routes assessment due to the availability of a number of freeways and associated arterioles in the network is used as the test network . The results of the evaluation and tests are reported to be fairly consistent with the results of the simulation runs

Although the explicit consideration of factors other than traffic volumes, in the determination of diversion routes is the salient feature of this model, the model is limited by several deficiencies such as the pre-selection process used to determine alternate routes.

The model developed by Gupta and Shuldiner (1992) considers information on traffic volumes in the arterioles as a very important influence in determining the diversion routes. Unlike the model by Berger et al. (1976), their model is dynamic in the sense that the diversion routes are updated in time. An appealing characteristic of this model is that trends in traffic conditions (like anticipation of the afternoon peak) are taken into account by applying historical patterns to real-time data to produce forecasted volumes. The functioning of the algorithm for dynamic network diversion is based entirely on a pre-defined set of diversion routes identified for each link of the freeway. The algorithm focuses on checking each route's v/c ratio. If the v/c ratio is less than a pre-defined critical value and if the time to reach the origin node of the diversion is less than the incident impact time the route is chosen as a diversion route, else it is not considered from diversion.) While features like recognition of the dynamics of route costs, severity of incidents and their impact on diversion route selection, and use of historical information for prediction, make this model a significant improvement over the others, the model does have its share of deficiencies. There are three major deficiencies associated with this modeling process.

1. Firstly the dynamics of factors, such as maintenance, bad weather, jurisdictional issues other than traffic volumes can render some of these pre-selected routes infeasible.
2. Secondly some other diversion routes may often be more suitable for diversion than the pre-selected routes based on the real-time traffic conditions.
3. And thirdly for application in real networks with hundreds of freeway links, explicit enumeration of alternate routes for each freeway link is very laborious, inefficient, and not realistic. Instead, efficient time-dependent disjoint or partially disjoint route determination algorithms can be used to generate alternate routes in real-time.

Further, the model will not be 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.

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. Taylor (1991) indicates some of the salient features that may be used to generate diversion routes using a knowledge based expert system. The problem data is divided into two parts: 1) the basic trip data about the origin, destination, time of day etc., and 2) the information the system may need about the characteristics of the traveler. The assumption 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 arterioles, 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.

In a recent study in Texas A&M, Ullman et al. (1994), it is 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 base 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. The results indicate that it will continue to be difficult for transportation agencies to predict how motorists will respond to specific travel time information unless individual preferences and diversion thresholds can be obtained directly from motorists.

Mahmassani et al. (1990), report that drivers who listened to radio traffic reports have a "greater propensity to switch routes". Drivers diversion propensity is also studied extensively by Khattak et al. (1992). The stated preference approach is used to study drivers diversion propensity. The empirical model based on a survey of downtown Chicago automobile indicates the relative importance of each variable in determining the diversion route. The results of the model show that drivers are willing to divert during incident conditions rather than for recurrent congestion. The second important result is that commuters are more willing to divert to an alternate route that is familiar, passes through safe neighborhoods, and with no traffic stops. Furthermore, drivers of higher income are found to be more willing to divert possibly due to their higher value of time.

Cascetta et al. (1992), present a study on the route choice characteristics of the users. The data from a road network in Torino is used to develop a statistical route choice model. The results of the analysis show that the significant variables in an individuals route choice model 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.

2.3 Dynamic Traffic Assignment / Routing

In this section, the literature on dynamic traffic assignment / routing is divided into two main sections: 1) dynamic traffic assignment models for planning and evaluation 2) dynamic traffic assignment / routing models for real-time control. his division of DTA models is important in order to better emphasize the motivation of this dissertation. The first category is comprised of models developed for planning / evaluation purposes that are generally extensions of classical static assignment models. They are relatively computationally expensive and not suitable for real-time applications. The second category of models, although small in numbers, are specifically developed for real-time applications such as real-time traffic control, route guidance, and others. They are designed to be computationally less demanding than the previous category of models and to take full advantage of the real-time sensor data.

2.3.1 DTA Models for Planning / Evaluation

The models that belong to the planning / evaluation category can be divided into two:

1. Heuristic / Simulation based models,
2. Analytic models.

Next sections discuss briefly models developed under each sub-category.

2.3.1.1 Heuristic / Simulation Based Models

SATURN (Simulation and Assignment of Traffic to Urban Networks) by Hall et al. (1980), and Van Vliet et al. (1981, 1990) is an early attempt to perform dynamic traffic assignment. It differs from the static traffic assignment methods by carrying out assignment and simulation phases iteratively. The period modeled is divided into time segments of 15-30 minutes. These time segments are treated independently but, in the simulation, vehicles queuing at the end of one time segment can be passed to the next as initial queues of fixed route traffic. The iteration loop starts with an assignment using arbitrary flow / delay relationships and is terminated when a satisfactory agreement between the assigned flows in the successive iterations on a sufficiently large proportions of links and turns. Flow / delay relationships are obtained from the simulation phase and are input to the assignment phase. Minimum cost paths through the network are determined and loaded using these relationships. Then, new link flows and turning movements based on this loading process are derived and fed into the simulation model that uses them to determine a new set of flow / delay relationships. The final assignment gives the estimated routes and flows on the links. Simulation and assignment modules handles only the traffic within one time period, generally independent of other time periods. Hence, the SATURN model becomes quasi-dynamic.

CONTRAM (CONtinous TRaffic Assignment Model) by Leonard et al. (1982) is one of the first heuristic dynamic simulation / assignment models. CONTRAM is a time-dependent traffic assignment model, based on traffic flow interactions rather than a simulation of individual vehicle movements in real-time. The flow of traffic on any link is the sum of the flows in the individual packets traversing the link in the interval concerned. CONTRAM splits each fixed demand O-D matrices into packets (each packet is formed of ten vehicles) and assigns these packets to the network using an iterative process until a stable assignment is obtained.

In the first iteration, a packet is assigned to its shortest path from its origin to its destination. Then, a list of links are stored to determine the time slice in which it traverses each link along its route. The network conditions are updated by adding the number of vehicles in the packet to the flow estimate on each link in its path, in the appropriate time slice. This process is repeated until all packets have been assigned. Subsequent iterations of the assignment procedure for individual packets are similar to the first iteration. But, before the contributions of a packet to the flows on the links in its route, its contributions during the previous period are subtracted to avoid double counting. CONTRAM lacks a well-defined formalism for its assignment and attempts to reach equilibrium by successive iterations. It also suffers from the lack of advanced time-dependent shortest path generation routines and the way it models time dependence through very long time intervals. Thus, due to all these drawbacks, it is only considered to be a quasi-dynamic model.

Another simulation / assignment model, DYNASMART (Mahmassani et al. 1985, 1991, 1995), is developed to study the effects of real-time information in urban traffic networks. This model evolved from the discrete-event magneto-hydrodynamic particle code developed for plasma physics (Leboeuf et al. 1979) and the macroparticle traffic simulation model that is based on it (Chang et al. 1985). The vehicles are moved in groups, called macroparticles. Concentrations, speeds and travel times on each link are updated at each simulation step. The model keeps track of the positions of each vehicle but, other microscopic details are not considered for the sake of simplicity. Different routing strategies are used to assign vehicles to the network. Among these are:

1) **Home based pre-trip information** where the user can select its path only at the origin.

2) En-route information where users have only access to information along the way. Under this strategy, they can only enter the network through their primary highway path and can switch through available paths.

3) **Both home-based and en-route information** combined where they have access to both home-based and en-route information. They can select their initial path and switch to better routes along their ways.

In addition to these strategies, different user behavior models , market penetration values for guided vehicles are tested to understand the evolution of the network conditions under different rules. The model does not have any capability of predicting future travel times as a function of present conditions. Therefore, the guidance information does reflect the future formations. On the other hand, the strategies that allow the users to switch along their path enable the model to recover the mistakes of the previous routing information based on the current conditions at that time.

Mahmassani et al. (1995) present algorithms developed for system and user optimal dynamic assignments over a congested network using the simulation / assignment framework, DYNASMART. They also study and evaluate the effects of network performance under real-time information systems for electronic route guidance. Peeta et al. (1995), present a rolling horizon framework for capturing the short-term changes in O-D trip tables. They present two formulations and a common solution technique for the multi-user dynamic traffic assignment problem. The first formulation does not allow the re-routing of motorists whereas the second one does. One of the main disadvantages of using DYNASMART for solving the above problems is the fact that its simulation component is not evaluated using real-world data and different network topologies. Another important drawback is the computational requirements which might be very demanding for real-time applications. However, DYNASMART proves to be a very valuable tool especially for off-line evaluation of traffic patterns under real-time information and different route guidance approaches.

Koutsopoulos et al. (1992) present a simulation based assignment model to evaluate and compare the effectiveness of different routing strategies. The model is formed of a simulation and a routing module. The emphasis is on the routing strategies. The shortest path determination for the routing information is based on 3 different travel time types:

a) Shortest path with current travel times.

b) Shortest path with simple projection used by ALI-SCOUT system (Koutsopoulos et al. 1992).

c) Shortest path with projection using information discounting (Koutsopoulos et al. 1992).

Travel times on links are determined using Greenshield's speed density equation. The information given to the vehicles entering the network is updated at every 1 minute. They are allowed to change their original routes at predetermined nodes in order to recover from projection errors at the previous intersection. In addition to the routing strategies mentioned above, to avoid the overloading of the shortest route, an "ideal policy" is used. According to this policy, at every information update not all vehicles are routed to the shortest path; an optimal split ratio is determined and the equipped vehicles are assigned either to the link with the least travel time, or to several alternative links so that all guided vehicles experience the same travel times. This avoids the congestion of the shortest route. The combination of these different strategies and policies are tested in a small network to provide a better understanding of ATIS systems. But, a final model description is not provided. This model is still in the development stage.

An iterative routing / assignment method for anticipatory real time route guidance is developed by the IVHS group in University of Michigan (Kauffman et al. 1991). The iterative routing-assignment procedure begins with forecasted travel times and calculate route guidance information for assigning guided vehicles. This assignment provides new travel time values that are used as forecasted values for the next iteration. In the simulation phase, a microscopic traffic simulation model, called INTEGRATION is used. INTEGRATION is created by M. Van Aerde (1993) for investigation of congestion due to interactions between freeway corridors and surface street areas. It maintains information for each vehicle in the network including "location, origin, destination, and ability to receive route guidance". The route guidance is based on the determination of the optimal next link to be taken by a vehicle as a function of its current node location and destination and the current time epoch. This process of determining the next best link from the previous iteration's travel time values (forecasts) for each vehicle reaching a node is called anticipatory route guidance. The iteration between the assignment and the simulation terminates when the travel times of the current iteration are sufficiently close to the travel times of the previous iteration. Outputs of this model include time-dependent link volumes and travel times, delays, pollution outputs and other measures of travel quality.

2.3.1.2 Analytical Models

Analytical models can be broadly divided into three main sub-category based on the mathematical forms which have been used to construct DTA models:

1. mathematical programming
2. optimal control
3. variational inequalities

Each sub-category can also be divided into two more distinct categories based on the on the formulation :

1) System optimal
2) User optimal

In the next section, DTA models categorized based on the classification scheme presented above will be briefly described and their merits and demerits will be outlined.

2.3.1.2.1 Mathematical Programming

2.3.1.2.1.1 System Optimal Formulations

Merchant and Nemhauser (Merchant and Nemhauser, 1978) present a model of the dynamic network flow. They adopt the system optimal view and present a system optimal traffic assignment model. It is a discrete time, single-commodity, non-linear, non-convex minimization model for a single destination network. The congestion is represented explicitly in the constraints. In their model which can be viewed as the partial generalization of the well-known static, system optimizing, assignment model, travel costs, route choices and flow rates are independent. In other words, the flow that exits a link "a" depends only on the flow on this link "a" but, in reality it should depend on the flows and the capacities of the other links of the network. The most important problem associated with this model is its nonconvexity which causes analytical and computational problems. A stepwise linear version of this model is proposed and solved by Ho (1980).

Carey (1987) reformulates Merchant and Nemhauser's model as a convex nonlinear program. This new formulation has the following advantages over Merchant and Nemhauser's model:

1) It is a well-behaved convex programming model that can be solved using any well-known algorithm already developed for convex programs (a piecewise linearized version of the model is a standard linear program).

2) Karush-Kuhn-Tucker conditions are shown to be both necessary and sufficient to characterize an optimal solution, whereas in the Merchant and Nemhauser' model they are not sufficient and have not been shown to be necessary.

Carey (1990) presents several methods for extending the model to multiple destinations and multiple commodities. The extensions also include flexible departure times and elastic demands. But, most of the problems that originates from the original Merchant and Nemhauser formulation continue to exist. Carey (1992) shows that the nonconvexity of the dynamic assignment problem when it is extended to the multiple destinations and multiple commodities is essentially due to the additional constraints introduced to satisfy the first-in-first-out (FIFO) requirement. He suggests several ways of dealing with the nonconvexity problem by approximating the FIFO conditions in continuous rather than the well accepted and more natural discrete formulations. Especially, assumptions related to the flow exit functions continues to create problems.

Chang et al. (1988) present a time-space network formulation for the system optimal time-varying flows in urban commuting networks. They use a simplified representation of the traffic flow that is still sensitive to queue build-ups, dissipation and spillbacks. They also suggest the use of the available large-scale network optimization algorithms for solving the problem.

Ghali and Smith (1995) propose a deterministic queuing assignment model which seeks to minimize total traveler delays in a network by routing motorists according to local time-dependent marginal travel times. This model can be applied to multi-commodity networks with many origins and many destinations. The model has two basic parts: 1) network loading procedure which consists of the simulation of the interaction between vehicles, the calculation of the entry flow for each link for a given set of route inflows, and the evaluation of network performance, and 2) an assignment procedure which is concerned with the calculation of vehicles' routes. Numerical results for two example are also given in the paper. The approach presented in this paper does not guarantee system optimality and has some limitations to certain assumptions in queuing. Also, the computational effort required to determine the solution of the model is reported to be very high for real-sized networks and ATMIS applications.

2.3.1.2.1.2 User Equilibrium

Janson (1990) first formulates and evaluates a heuristic dynamic traffic assignment. A key assumption to this model is that the route choice decisions are made at the time of trip departure based on projected link impedances that account for changes in travel demand over future time intervals. Then, a dynamic user equilibrium (DUE) model is developed. This problem is defined as follows: "Given a set of zone-to-zone trip tables containing the number of vehicle trips departing from each zone and head towards each zone in the successive time intervals of 10-15 minutes each, determine the volume of vehicles on each link in each time interval of a network connecting these zones that satisfy following two conditions:

1. All paths between a given pair of zones used by trips departing in a given time interval must have equal travel impedances.

2. All paths between a give pair of zones not used by trips departing in a given time interval cannot have lower travel impedances. "

This dynamic user equilibrium (DUE) problem is first formulated as nonlinear mixed-integer program in terms path flows. Then, DUE is reformulated in terms of link flows. The advantage of link flow decomposition is that it does not implicitly assume complete enumeration of all possible paths between zone pairs and it decomposes into inner and outer sub-problems being solved successively by the convergent dynamic algorithm presented in the paper. Convergent dynamic algorithm uses methods of linear combinations to find successive solutions to DUE while holding temporal link uses from each origin fixed. Both formulations, link flow and equivalent path flow, prevent temporally discontinuous flows that arises because of not knowing how the time interval will change in the vicinity of each node with respects to flows originating from different origins in different time intervals.

In subsequent papers, Janson (1991, 1992) formulates the dynamic user equilibrium traffic assignment problem as a bi-level problem with an upper problem which solves the multi-interval time-varying demand traffic assignment and a lower problem which determines the reachability of the of each node in each time interval He then solves each problem interdependently until a pre-determined convergence criteria is reached. He also introduces several new ideas in order to handle the endogenous path-link variables by fixing them with a shortest path problem and then assigning the time-varying demand onto the constrained expanded time-space network. His model has several drawbacks such as the use of BPR functions to calculate link costs, the inaccuracy in defining node time intervals when trips overlap between two intervals and inability to represent correctly the traffic dynamics.

A space-time network flow model that represents traffic flows over time for a capacitated road transport system is presented by Zawack and Thompson (1987). The travel is modeled as a piecewise linear convex function of the initial flow. Each route's capacity is determined by the physical space available at that time on that route.

Two solutions are proposed. The first solution is for a multiple-source single-destination space-time network. The objective of the space-time network flow model is to find a set of dynamic flows over time which minimizes the total travel time. To ensure the solution corresponds to the original requirements of the problem, the dynamic network flow problem is solved as a multi-commodity network flow problem where each set of vehicles leaving from the starting location and bound to a given destination is a separate commodity.

The second solution is the "shortest path solution". This solution is based on the assumption that each user wants to follow the fastest route to its destination. The algorithm starts by determining the shortest path for the first arriving cars, then dispatches them along these routes. The new time-space traffic flow network is obtained by only reducing the capacities of the used links. The shortest paths for the new cars are determined using this new network. This process of generating a new time-space traffic flow network, finding shortest routes for the vehicles at the origin, and the dispatching them to these route continues till all the vehicles are assigned to a minimum route. This algorithm which is only applicable to a single-source single-destination network can be iteratively applied to the resulting network to find out one consistent shortest path solution.

The approach proposed by Drissi-Kaoutini et al. (1989) can be looked as an improvement of the of the dynamic maximal flow models developed by Ford and Fulkerson (1962) and Zawack and Thompson (1987). The dynamic traffic assignment model is mainly based on the assumption that the time spent by a vehicle on a link may be decomposed into a fixed travel time plus a waiting time. The fixed travel time corresponds to the free flow travel time over the link. Then, the vehicle is put in an exit queue until it becomes possible to enter the next link. This decision depends on the link capacities and costs. This model leads to specific network structure which is called "the temporal expansion of the base network". The temporal expanded network corresponds to the flow conservation constraints of the model. "Using the expanded network, it is possible to express the dynamic traffic assignment problem as a static assignment problem over this temporal expansion of the base network". Then, the problem is solved as a simple static user equilibrium problem over static temporal expanded network.

A new dynamic model for the congested network is described in Smith (1982). The model determines time-varying link flows and route costs, in terms of given time varying route inflows. A cost vector "C^X" of route costs that depends in a continuous way on the vector "X" of route inflows is introduced. Then, the dynamic user equilibrium is defined as follows:

"For each OD pair and each time "t", there exists a dynamic user equilibrium, if and only if more costly routes are not used."

The statement above is the major constraint of the minimization problem described by the authors. An objective function "V" that minimizes the total cost of travel for each user while satisfying the above constraint is established. A descent algorithm similar to the one for steady-state user equilibrium is suggested. Meanwhile, several problems related to this model remains to be studied furthermore:

1. The differentiability properties of the proposed objective function.
2. The proof of existence of a unique global minimum for the problem.
2. The computational feasibility of the overall model.

In a more recent paper, Smith (1993) specifies in detail a new dynamic model of a single complete peak period or a single complete day. He gives a proof of the existence of dynamic user equilibrium if his model is used to determine route costs. The paper also gives several optimization formulations of the dynamic equilibrium problem and suggests imprecise descent algorithms for calculating dynamic equilibria without giving any formal proof of convergence. Finally, the paper provides a family of alternative methods of determining dynamic flows which satisfy FIFO conditions by employing different ways of specifying the priority of a vehicle or a group of vehicles. One of the major shortcomings of this model is the computational effort required to solve it.

Recently, Jayakrishnan et al. (1995) present a dynamic user equilibrium traffic assignment model with traffic flow relationships based on a bi-level optimization approach. They have used link cost functions derived from the modified Greenshield's speed-density relationships. The model is solved using a heuristic algorithm that takes full advantage of its bi-level formulation structure. The formulation differs from similar formulations in that the traffic flow relationships are explicitly incorporated into the model itself and the use of short time intervals allows better representation of traffic dynamics. Although this model is a good attempt to bridge the gap between traffic assignment algorithms and traffic flow theory, it appears to be computationally very expensive for real-sized networks and real-time ATMIS applications due to the extra constrains and integrality requirements for some variables such as the "equilibrium nodal arrival estimates".

In addition to the deterministic dynamic user equilibrium formulations, there have been several attempts for developing dynamic stochastic user equilibrium formulations. Birge and Ho (1993) modify the Merchant and Nemhauser (1978) model by treating the O-D flows as random variables. They formulated the new model as a multi-stage, stochastic, non-linear, and non-convex mathematical program.

2.3.1.2.2 Optimal Control

2.3.1.2.2.1 System Optimal

The model by Friesz et al. (1989) adopts a system optimal perspective for the traffic assignment. The same paper also presents a user optimization version of the system optimal formulation. The model formulates the problem of dynamic traffic assignment as a "continuous time optimal control problem" for the single destination case. The optimality condition of the dynamic model requires only equalization of instantaneously perceived unit costs. It only considers the route choice decision making because travel time at each instant is assumed to be known and inelastic. The particular extension to the model by Wie et al. (1991) is to include elastic travel demand which implies departure time choice. Wie et al. (1994) present the development and test of a Lagrangian method for solving dynamic system optimal traffic assignment models formulated as optimal control problems. Their numerical method has been proven to efficient since no path enumeration is required during the computation and the optimal control framework provides a natural way for problem decomposition by time period.

The dynamic traffic assignment model developed by University of Michigan is also in the system optimal paradigm (LaFortune et al. 1991). The dynamic assignment problem is considered in a network with multiple origins and destinations. This network is modeled over the time horizon as a discrete time dynamical system. The objective of the model is to assign the vehicles to links over time in order to minimize the total travel time experienced by all the vehicles using the network. By making the assumption that the demand is known beforehand and the time horizon is fixed (no incidents), an optimization problem that can be solved completely is obtained. This optimization problem is an optimal control problem for a nonlinear integer valued discrete-time dynamical system. It is amenable to an algorithmic solution based on dynamic programming. This model and the solution technique is applied to simple network but, its limitations for a large network are not known yet Another question is related to the real-time feasibility of this approach where demands are changing continuously.

2.3.1.2.2.2 User Equilibrium

An equivalent continuous time optimal control formulation of the dynamic user optimal traffic assignment problem on congested multi-destination networks is given by Wie et al. (1990). They also give the necessary optimality conditions and establish that the necessary conditions are also sufficient. Finally, the model is shown to be a time-dependent generalization of Beckman's mathematical programming problem for a static user equilibrium traffic assignment in the form of an optimal control problem by the equilibration of instantaneous user route costs.

The dynamic traffic assignment problem by Ran, Boyce and LeBlanc (1993) is solved from the user optimal point of view. The dynamic user optimal problem is defined as a dynamic generalization of the conventional static user optimal problem. The objective of the dynamic user optimal traffic assignment problem is described as to find the dynamic trajectories of link states and inflow and exit flow variables given the time-dependent OD

flow requirements, the network and the link travel time functions. The problem is formulated based on the underlying choice criterion that each traveler uses the route that minimizes his/her instantaneous travel time when departing from the origin or any intermediate node to his/her destination. Finally, all the users that depart at the same time from the same origin to the same destination should experience the same travel times.
An equivalent optimization model of the dynamic user-optimal traffic assignment problem that will satisfy the above statements is formulated using optimal control theory. The objective function consists of four terms. The first two terms are related to the travel times of vehicles in the network, the third term is a penalty cost function for vehicles remaining on links at time "t", the fourth term is a salvage cost function for vehicles remaining on the links at the end of the analysis period. This penalty and salvage cost terms are standard in optimal control theory. The constraints of the model include the state equations for each link, the equations for flow conservation at each node, the relationship between travel times, link flows, nonnegativity and boundary conditions.

The dynamic user optimal problem is converted into a nonlinear problem by dividing the assignment time interval into small time increments. Hence, the optimal control problem is reformulated as a discrete nonlinear problem that can be solved by the efficient Frank-Wolfe algorithm also used for solving the static user equilibrium model on urban networks. An algorithm, similar to the static user equilibrium algorithm, is presented for the dynamic user equilibrium problem. The original network is converted to "an expanded time space network". One significant aspect of the algorithm is that by using this space-time expanded network, "the Frank-Wolfe LP sub-problem requires only the solution of minimal cost route problems for each OD pair". Thus, this expansion technique allows standard algorithms for static traffic assignment to solve dynamic traffic assignment problems.

This algorithm is tested for a small network; it needs to be tested for larger networks and more realistic cost functions needs to be used. Other important improvements can be accomplished by modifying the problem formulation, especially for the formulation of constraints that uses estimated link travel times.

Ran et al. (1993) formulate a bi-level programming of the dynamic user optimal departure time and route choice problem by using the optimal control theory. This model extends the dynamic user optimal model described above to the case where both departure time and route over a general network must be chosen. Although the optimality conditions that are consistent with dynamic user optimum departure time and route choice are given in the paper, a solution algorithm is not provided.

Ran et al. (1992) formulate two dynamic user optimal route models as equivalent optimal control programs. The first model determines driver's route choice based on the instantaneous travel time between the decision node and the destination. In the second model, for each origin-destination pair, a driver's route choice is based on the perceived route travel time experienced between the origin and destination. The development of efficient solution algorithms and the calibration of model parameters are both left out in this paper as future research.

Another dynamic user equilibrium optimal control model for many to one travel demand is proposed by Lan and Huang (1995). An efficient algorithm that generates the approximate solution to the network optimal control problem has been developed. A numerical example for a real road network is also given.

2.3.1.2.3 Variational Inequalities

2.3.1.2.3.1 User Equilibrium

Friesz et al. (1993) model the dynamic traffic assignment model in which each user wants to find the best path and optimal departure time in order to arrive at the destination on time, considering variable road conditions over the planning horizon. They show that there is a variational inequality formulation of dynamic user equilibrium with simultaneous route choice and departure time decisions.

Bernstein et al. (1993) present an alternate formulation of the simultaneous route and departure time choice problem which does need the inverse exit time functions. Then, they present a heuristic path swapping algorithm to solve this new model for the deterministic queuing network case. Important issues such as the performance of the proposed solution algorithm and its convergence are identified as future research topics.

Wie et al. (1995) formulate the discrete time dynamic user equilibrium problem on a general network with many origins and destinations as variational inequality problem. They use arc exit functions to model vehicle flow through state dynamics of the kind used in Merchant and Nemhauser (1987) together with nest cost operators introduced in Wie (1989) and Friesz et al. (1993) to determine unit path costs that reflect the route and departure time choice decisions of network users. A heuristic algorithm for obtaining an approximate solution and some numerical results for Sioux Falls network are also presented.

Drissi-Kaoutini (1993) formulates a dynamic traffic assignment model which incorporates explicit link queues as a multi-commodity network equilibrium problem using variational inequalities.

Ran et al. (1996) formulate a multi-class dynamic user equilibrium model as continuous time variational inequality (VI) problem. They identify three distinct classes of users namely, fixed used travelers, stochastic user optimal travelers, and dynamic user optimal travelers. To solve the discretized version of the continuous time VI, they propose an algorithm that combines relaxation, Frank Wolfe and Method of Successive Averages (MSA) techniques. The determination of the correctness of the results, the testing of the proposed solution algorithm using larger networks and the calibration of model parameters are identified as topics for future research.

2.3.2 DTA / DTR Models for Real-Time Operations

There is very little research on the development of DTA / DTR models for real-time operations. Papageorgiou and his colleagues have been the pioneers in this area. In addition to their work, recently Peeta et al. (1995) have proposed the use of rolling horizon approach in conjunction with the simulation / assignment model, DYNASMART. Unfortunately, the addition of the rolling horizon approach does not eliminate problems inherent to DYNASMART discussed in detail in the critique section of this chapter. Therefore, the main discussion in this section will be on the models developed by Papageorgiou and his colleagues during the last decade.

The application of closed loop control systems to the control of freeway systems is described in Papageorgiou (1983). In addition to the general traffic model for freeways and intersections, the overall multilevel approach to the solution of dynamic nonlinear optimal control problems and its applications to control problems is also presented. In Papageorgiou (1991), dynamic traffic phenomena is modeled macroscopically for a multi-destination traffic network. The basic assumption of the model is that traffic demand at the origins of the network is deterministic and independent of the traffic conditions in the network, meaning the demand is not elastic but time-varying. The model presented in the paper consists of three parts: 1) a macroscopic traffic flow sub-model that describes the dynamic evolution of the traffic in the network, 2) a traffic composition part that describes the propagation of traffic compositions in the network, and 3) a dynamic assignment part that routes traffic streams to obtain user optimum conditions.

The most important part of the overall model is the dynamic assignment part that uses a feedback methodology to establish user optimum dynamic traffic assignment conditions. This methodology adopted in this paper uses a multivariable feedback regulator with integral parts and a simple bang-bang controller to achieve a dynamic user optimum. The simple feedback regulator attempts to equalize travel times on alternative routes by re-routing them at decision points. This approach is attractive, but is difficult to design for nonlinear systems. One approach is to linearize the system, but then the results will be valid only in the linear region. Another method suggested in Papageorgiou (1990) is the application of Linear Quadratic (LQ) optimal control on the linearized model of the system, which has the same difficulty. The main drawback of this approach is that the feedback regulator may be insensitive to sudden changes, especially in case of severe congestion.

In a subsequent paper, Messmer and Papageorgiou (1995) have formulated the control task as dynamic, non-linear, discrete optimal control problem and solve it by gradient based search. In this approach feedback control is realized by solving nonlinear optimization using gradient search over a sufficiently long future time horizon at each control interval (1995, 1994). This method obviously involves more computation and also might have difficulty with the proof of performance characteristics. They also evaluate the controller's computational efficiency and control benefits in terms of global network performance switching performance of VMS and user optimality using simulation. The system is also shown to be robust with respect to modeling inaccuracies.

2.3.3 Summary and Critique of the DTA / DTR Models

A short and critical review and assessment of the DTA / DTR literature presented in the previous section can be done through the discussion of these models with respect to some key features:

1. Dynamic modeling of traffic and user behavior
2. Cost functions
3. Solution techniques
4. Qualitative properties

Dynamic Modeling of Traffic and User Behavior

All DTA models discussed in the previous section, both analytical and simulation based, have adopted some dynamic traffic models and constraints. Analytical DTA models have explicit types of dynamics such as the link entrance flow variables and link exit flow functions or flow variables similar to the dynamics presented in Merchant and Nemhauser (1978), or integral equations based on link exit time functions similar to the ones presented in Friesz et al. (1993). Although the original dynamics proposed in Merchant and Nemhauser (1978) provide a convenient way to represent traffic flow by assuming that the rates of change of volumes on links equal to the net difference of link inflows and outflows, for longer links this approximation fails to capture real dynamics of traffic flow. Merchant and Nemhauser try to account for queuing at the link exit locations by introducing link exit functions which depend on link volumes. These functions, in turn, are shown by Carey (1992) to cause violations of FIFO queue discipline requirements. Later, Ran et al. (1993, 1995) have attempted to modify the Merchant and Nemhauser approach for modeling traffic dynamics by redefining flow and exit rates as decision variables and adding a flow propagation constraint which is shown by Daganzo (1994) to cause anomalies with respect to the fundamentals of traffic flow theory. Friesz et al. (1993) proposed a system of integral equations involving arc exit time, not exit flow functions, to represent the link dynamics. However, these integral equations are difficult to solve even for off-line applications.

All the DTA models also include some user behavior assumptions to model the route choice, departure time choice, re-routing decisions, and equipment choice. One way to deal with this variety of behavioral assumptions is to divide users into different classes of users and decide the percentages of users belonging to each user group. The route choice is generally modeled through the use of user equilibrium or system optimal assumptions. The departure time choice is frequently modeled through the use of pre-determined time dependent O-D trip tables (Mahmassani et al. 1995). Other models model the route choice through the use of departure time parameters that are adjusted on a trial-and-error basis by the modeler using some kind of heuristic such as the diagonalization approach proposed by Ran et al. (1992). On the other hand, en-route re-routing is modeled approximately by Friesz et al. (1989) and Ran et al. (1992) by assuming route diversions at each intersection. DYNASMART (Mahmassani et al. 1995) models en-route re-routing through the determination of kth shortest paths at each intersections based on the prevailing network conditions. These kth shortest path calculations are reported as the most time consuming part of DYNASMART and they also present several theoretical problems due to the fact that the existing kth shortest path algorithms share many links and they are really not too different from each other with the exception of few links. It is clear that the modeling of en-route re-routing presents many challenges and needs to be studied more in the future.

The extension of existing single class DTA models is generally done by determining several user classes that behave differently based on the equipment they own. For example, Mahmassani et al. (1995) and Peeta et al. (1995) suggest the existence of four user classes namely, equipped drivers who follow system optimal routes, equipped drivers who follow user equilibrium routes, equipped drivers who behave according to the boundedly rational rule for making switching decisions, and equipped drivers who follow pre-determined routes. Very similar grouping of users is also used by Ran et al. (1996) as an extension of the dynamic traffic assignment model proposed by Boyce et al. (1992), Ran et al. (1993). The introduction of multiple user classes complicates both analytical and simulation based DTA models to a degree that they become very difficult to solve for any real-sized networks.

On the other hand, feedback based models proposed by Papageorgiou et al. (1990a, 1991) adopts a macroscopic model of traffic and develops feedback control models. Since the goal of these feedback models are quite different from most of the predictive DTA models discussed above and that they do take full advantage of real-time sensor information, the macroscopic modeling of traffic are reported to perfectly serve the purposes of these feedback controllers (Papageorgiou et al. 1990b, 1990c).

Cost Functions

Development of realistic link cost functions / link impedance functions for both freeways and arteriales has been an active research area for many years. These functions are not only required to be accurate enough to account for the present traffic, they are also required to capture the effects of future traffic which will be encountered during future time periods. There are mainly three types of approaches used to determine the link costs namely, close form impedance functions (Janson, 1993), point queue models (Friesz et al. 1994, Bernstein et al. 1994, Drissi-Kaoutini 1993), and combined simulation assignment models in which link costs are determined by the simulation (SATURN, CONTRAM, DYNASMART).

Although BPR impedance functions are the most widely used closed form functions, they are developed by FHWA for steady state conditions and it is generally accepted that they are not suitable for DTA models. Although Ran et al. (1994) proposed some realistic new cost functions for freeways and arterioles, they always used some simplified functions that make their models mathematically tractable but that are not based on any empirical basis (Ran et al. 1992, 1993, 1996).

Solution Techniques

It is obvious that the solution techniques used to solve the DTA models largely depend on the formulation of the model itself. Models such as CONTRAM and SATURN try to reach convergence by mainly iterating many times. DYNASMART on the other hand uses the multiple successive averages (MSA) method which is first proposed for static user equilibrium problems by Daganzo and Sheffi (1977). Mahmassani et al. (1995) show successful convergence results by using MSA. Simulation based models, due to the nature of the simulation naturally do take longer time to solve. Moreover, the results are of heuristic nature and some nice properties such as the convergence and finiteness of the solution algorithms cannot be proven formally.

The most used approach to model DTA is the optimization approach known as the mathematical programming with discrete variables. Also, other modeling approaches such as variational inequalities, discrete optimal control can all be shown to have a mathematical programming representation. Moreover, Ran et al. (1993) showed the equivalence of the continuous optimal formulations to the finite dimensional mathematical programs. Therefore, we can describe all these models as mathematical programming / optimal control / variational inequality models. For this category of problems, several authors proposed several different solution methods. Boyce et al. (1991) and Ran et al. (1992), (1995) proposed an algorithm based on Frank-Wolfe (F-W) (1956) and diagonalization techniques for the nonlinear program which is obtained by the time discretization of the continuous time optimal control problem. Janson (1991) used F-W algorithm to solve his dynamic traffic assignment model. Jayakrishnan (1995) also used a heuristic algorithm that is based on the F-W. Bernstein et al. (1994) used a heuristic path swapping algorithm to solve the variational inequality formulation of their DTA formulation. Wie et al. (1994) developed and tested an augmented Lagrangian method for solving for solving dynamic traffic assignment models formulated as optimal control problems. The main challenge still remains to solve these models for real networks. Almost all of the solution techniques described above are applied to small size test networks. The validity of the solution are not examined carefully. None of the above studies compared their results with real-world data, basically due to the unavailability of such data.

In summary, the need for the development of efficient solution techniques for existing DTA / DTR models emerges as a very important and timely issue. Although simulation based models allow better representation of traffic dynamics and obviate the need for problematic exit functions, they introduce several new problems such as the increased computation times, need for extensive validation of simulation component, inability to prove the existence and uniqueness of the solutions. Another important issue is the reliable comparison of model results with real-world data and assess the realism of the modeling assumptions and model results.

Qualitative Properties

The above DTA / DTR models can also be categorized as:

1. **Predictive / Descriptive models:** Most of the DTA models belong to this category. They are mainly used for planning and evaluation purposes and are not suitable for real-time application due to the drawbacks that discussed in detail in the previous sections.
2. **Prescriptive / Normative models:** Some of the models that belong to the first category are also modified for dealing with issues such as real-time guidance, etc. However, these models are originally developed for off-line purposes are not suitable for real-time operations. However some researchers have attempted to develop normative models from scratch. They use methods that are more suitable for the purposes of real-time guidance and information provision (Papageorgiou et al. 1983, 1990a, 1990b, 1990c; Messmer and Papageorgiou 1995; Cremer and Fleischman 1987). These models through the use of feedback can very effectively deal with issues such driver compliance in the presence of routing information and the real-time traffic conditions on the network links.

Clearly, prescriptive / normative models that use feedback are more suitable for real-time diversion which is the focus of this dissertation. Details of feedback control and different solution techniques for designing feedback controllers will be presented in the sixth, and seventh chapters of this dissertation.

3. System Framework for Traffic Diversion

3.1 The Scenario

Consider the following scenario for the small portion of the Fairfax county in Northern Virginia network shown in Figure 3.1. This network which covers 9.5 miles of the I-66 freeway section and neighboring arterials consists of 57 nodes and 111 links. A major accident that involves multiple cars and trucks has occurred on Interstate 66 during the peak hour (Northern Virginia Incident Logs, 1994). One lane of the two lanes and the shoulder of I-66 were closed due to this incident. Although the detection and verification of the incident do not take more than 10 minutes and the response begins immediately, restoration of the interstate traffic to normal flow can at least take one hour. The queue has begun to build up from the blocked point. Motorists who approach the incident location start feeling the effects of the incident in the form of slow traffic. The variable message signs display "Major Accident Ahead" and "Use Alternate Route" Messages. The traffic management center has to act fast and start diverting these motorists that are going to pass through the incident location. The traffic management center at Arlington is responsible for coordinating the response as well as for he selection and implementation of the best traffic management measures.

3.2 The Solution Approach

The solution approach calls for the development of an integrated, multi-layered framework which will assist the traffic management operators in developing and implementing diversion strategies in real-time. The traffic manager at the traffic management center needs to answer some very though questions very quickly. First (s)he

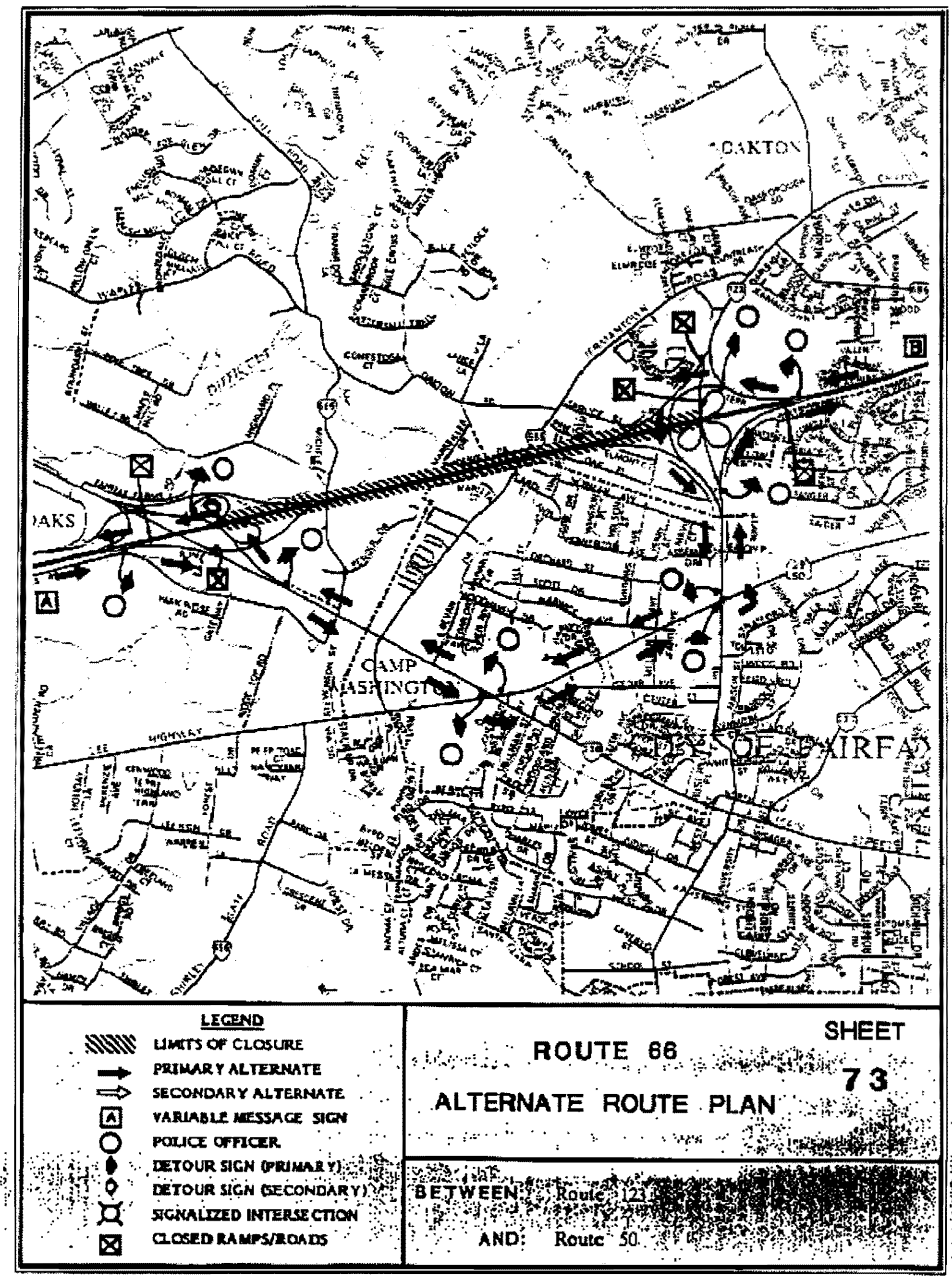


Figure 3.1: Sample Fairfax Network for the Scenario (Source: Northern Virginia Incident Management Manual)

needs to predict quite accurately the duration of the incident and decide if there is a need for diversion. Then, (s)he needs to decide the diversion point(s) and alternate routes to be used for diversion. Given the time and equipment limitations, it is generally difficult to implement diversion for more than one diversion points. That's why it is extremely important to choose the best diversion point for the specific incident and determine best routes for the chosen diversion point. Finally, the variable message signs need to be actuated in such away that the traffic flow on the alternate routes is controlled dynamically and over reaction of the motorists to the incident and routing advisory is avoided.

In addition to the existing diversion and traffic assignment / routing models discussed in the literature review Chapter, there are also very recent projects that attempt to develop operational diversion strategies for real-life situation. Among these are the new dynamic traffic assignment project named as "an operational dynamic traffic assignment system" and the VDOT project titled as the "development of traffic control strategies for the Suffolk management system". The former attempts to build a comprehensive dynamic traffic assignment system that addresses many situation including traffic diversion during incidents, traffic routing under normal traffic conditions, and planning purposes. The latter attempts to develop traffic control strategies for a two alternate route system using the results of off-line traffic simulation studies. However, to the knowledge of this author, none of these new projects proposed a practical yet complete framework to deal with freeway incidents in real-time.

3.3 System Framework

The main objective of this research is to develop a practical and real-time dynamic traffic assignment / routing system that will be used by the traffic control center during incident conditions. In order to achieve this goal, the system must:

- Be realistically and robustly designed to tackle real world problems and to model the real diversion process during the incidents,
- Be fast enough to develop effective traffic diversion strategies in real-time,
- Make use of both the historic and real-time sensor data,
- Be flexible so that it can be used in different cities and towns for different types of incidents.

In order to develop a realistic and effective system framework for traffic diversion during incidents, we have to first determine different steps followed during the diversion process. These are:

1. after the incident occurred, is detected and confirmed, predict the duration of the incident based on the incident characteristics;
2. based on the incident duration prediction, predict the delay that will be caused by this incident;
3. if the delay is more than a threshold value then initiate the diversion based on the diversion plans developed either off-line or on-line, and
4. implement and control the diversion using different tools such as variable message signs (VMS), highway advisory radio (HAR), and other tools.

Each step of this diversion process described above requires realistic and efficient models and algorithms for real-time implementation. This dissertation will discuss and describe in detail these models and algorithms that are developed as part of this research work.

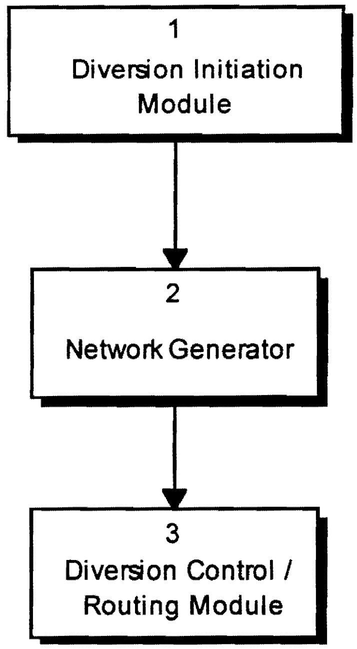


Figure 3.2: Three Modules of the Proposed Framework for Performing the Diversion Process during Non-Recurring Congestion

3.4 System Components

The diversion system architecture is divided into three basic modules. Each module has several sub-modules. The three main modules are "diversion initiation module", "diversion strategy planning module (Heuristic Network Generator)", and "diversion control / routing module". Diversion initiation module has two sub-modules, namely "incident duration prediction sub-module", "incident delay estimation sub-module". Diversion strategy planning module is also called "Heuristic Network Generator" (Ozbay et al., 1994; Hobeika et al. 1993). This module has “link elimination”, “route generation”, and “route prioritization” sub-modules. Diversion control / routing module has several sub-modules in the form feedback control models that use the real-time sensor data to control the diversion process to achieve a pre-set diversion objective. The overall structure and modules of the proposed system is shown in Figure 3.2.

3.4.1 Diversion Initiation Module

Traffic diversion is an effective tool to alleviate non-recurrent congestion. However, the decision to use it for real world application is not trivial. In this research, we propose a diversion initiation logic depicted in Figure 3.3. According to this logic, the duration of the actual incident is first predicted by the “incident prediction module”. The models used to perform the incident duration predictions are developed using real incident data collected by the incident management personnel from Northern Virginia (Subramaniam et al. 1994). The model development efforts and the validity of these models are discussed in detail in Chapter 4. The incident predictions are then used to predict the delays on the freeway. A general framework for delay prediction is proposed by Mastbrook (1996). The full implementation of this framework is not complete yet. In this dissertation, an extended version of a more simplistic approach that uses the deterministic queuing model for delay estimation (Morales, 1986) is adopted. The delay is predicted using the predicted incident duration as well as the lane blockage information and the traffic demand during time it takes to clear the incident.

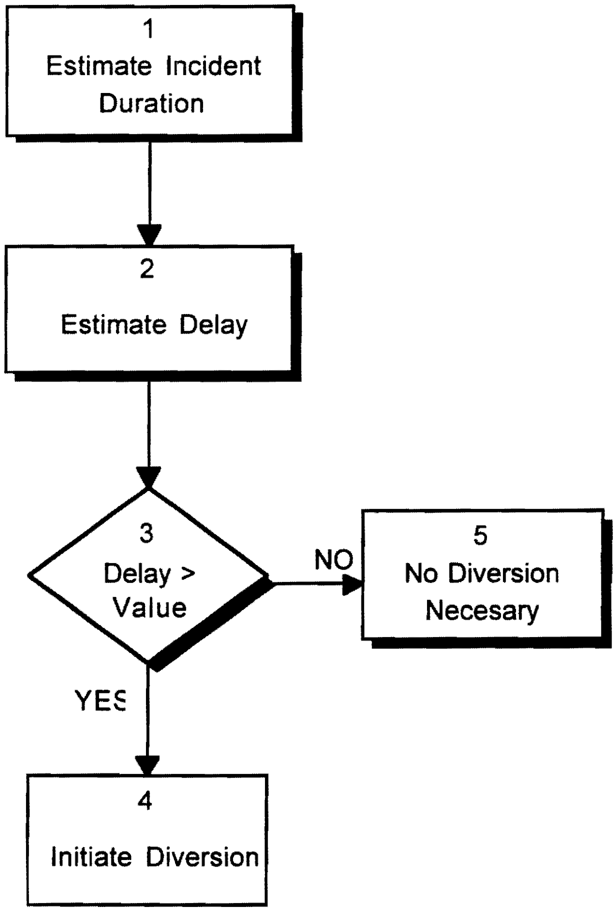


Figure 3.3: Diversion Initiation Module

The severity of delay is used as the mechanism to trigger the diversion process. The threshold values are set up using the expert knowledge obtained from the Northern Virginia incident management personnel. However, these threshold values will be different for different networks depending on the level of daily recurring congestion, the incident management procedures of the area and other site specific considerations.

3.4.2 Diversion Strategy Planning Module: Heuristic Network Generator

It is clear that under incident conditions, it is not necessary to consider the entire network for generating effective diversion strategies. An effective approach to address this problem is to determine the "incident impact zone" which will be smaller than the entire network for the majority of incidents. It is also well known that some of the links are not suitable for diversion due to many reasons, such as the already high v/c ratios, on-going construction, lack of equipment needed to route motorists or jurisdictional considerations

Network Generator is a knowledge based expert system developed to address these issues by pre-processing the network and determining a set of feasible routes that can be used for diversion (Figure 3.4). The knowledge base is developed using the knowledge acquired from the experts in Northern Virginia and the existing literature. The knowledge base consists of simple rules that determine the "incident impact zone" based on the incident delay and incident characteristics and rules that eliminate infeasible infeasible links for diversion based on the real-time and historic data and the network specific knowledge. A hybrid architecture that consists of a Geographical Information System (GIS), Arc-Info, an Expert System shell, Nexpert-Object, and C programs is developed for the implementation of Network Generator. This architecture supports the real-time working of the system by allowing the real-time manipulation of large network data through the use of GIS and supporting the decision making process through the use of the expert system. This architecture also allows a client / server implementation of the Network Generator which proves to be very useful in the multi-agency environment of the incident management process.

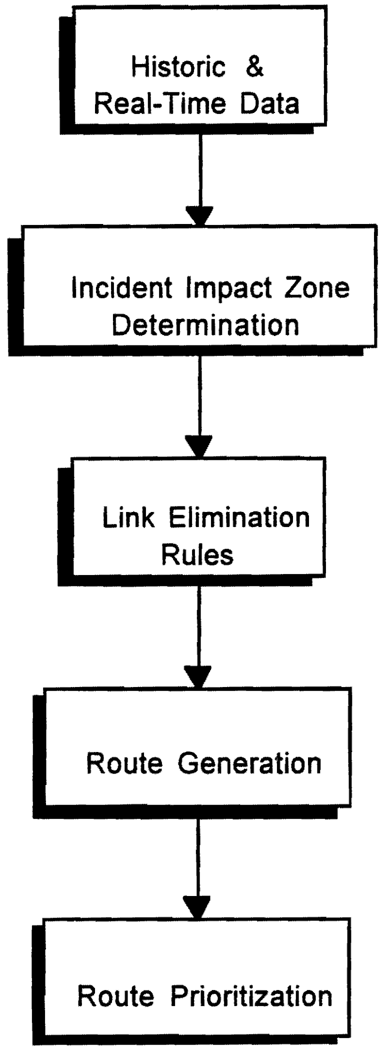


Figure 3.4: Heuristic Network Generator

3.4.3 Diversion Control / Routing Module

The heart of the proposed system is the effective and real-time control of the traffic diversion. The major goal of this module is to control the diverted traffic volumes in order to prevent over loading of some alternate routes and to reach a pre-determined goal such as system optimal or user equilibrium. This goal is achieved by using feedback control. The feedback control models developed in this research are still simplistic and need to be enhanced/calibrated for real-world implementations. On the other hand, they present a new and efficient alternative to the existing network-wide optimization models. These control models will use the sensor data to decide about the control commands given to the motorists. These control commands will be disseminated through variable message signs or highway advisory radio. The control commands will be revised at every sampling time based on the sensor data obtained from the traffic sensors on the roadway

The feedback control models developed in this research are designed for controlling the point diversion of traffic to multiple alternate routes determined by the "Heuristic Network Generator". The system dynamics and modeling issues of these feedback control models are described in detail in Chapter 6 of this research. Two heuristic controllers and the simulation results are also described in Chapter 6. Chapter 7 describes the feedback linearization of a simple highway network model and presents simulation results.

4. Diversion Initiation Module

This module is comprised of two major sub-modules: incident duration prediction and delay prediction sub-modules. Diversion initiation logic is based on the severity of incident delay which is determined by using the output of these two sub-modules.

4.1 Incident Duration Prediction

4.2 Background

In this section, several major studies on incident data analysis and duration are briefly reviewed for summarizing the previous work in the area of incident duration. Methodologies used in each study and their findings and conclusions are also briefly described.

In the study by Jones et al. (1991), in the Seattle area, the primary source of accident duration data is the State Police Dispatch records over a two-year period. The important concept is one of conditional probability in which the authors seek to determine the probability that an incident lasts 't' minutes given that it has already lasted '(t-1)' minutes. The incident duration is approximated by a log-logistic function as opposed to the log-normal distributions chosen elsewhere, even though the two are closely related.

The major drawback of this model is the use of several unrealistic variables in the model. For example, data like the age of drivers is not easily obtained during real-time incident conditions. It is highly improbable and unpractical to assume the availability of these variables in predicting the clearance times when the incident clearance process is ongoing. The basis for the log-logistic function used above is specific to the Seattle region. The incident duration function is modeled as log-normal in some other cases. Golob et al. (1987), in a specific study that deals with the truck accidents, developed this further from theoretical considerations of the incident occurrence process. The analyses are organized according to three categories: type of collision, incident severity, incident duration and lane closures.

Nine thousand five hundred and eight truck accidents obtained from Traffic Accident Surveillance and Analysis System (TASAS) in California are analyzed for the model development. In this data set, there are 4436 personal injuries and 120 recorded fatalities. Relationships between the type of collision and selected truck accident characteristics are explored using the method of log-linear modeling. The first analysis is the position (highway or ramp) of occurrence of the various collision types. Overturns, broadsides and hit-objects are more common on ramps, while rear-end and sideswipes are predominantly located on highways. The duration of an incident is hypothesized as being comprised of many sequential stages such as 1. Detection, 2. Initial Response, 3. Injury Attention, 4. Emergency Vehicle Response, 5. Accident Investigation, 6. Debris, 7. Cleanup, 8. Recovery.

It is assumed that the amount of time required for completion of each of these activities is influenced by the time required for preceding activities. The duration of the nth activity is of the following form:

$$Y_n - Y_{n-1} = Z_n \, Y_{n-1}$$

where,

Y_n = time of completion of nth response activity, measured from the start of the incident.

Z_n = random factor that relates duration of the nth activity to the cumulative time for the preceding activities.

Then,

$$Y_n = Y_{n-1} \, (1+Z_n)$$
$$= Y_0 \, \Pi(1+Z_i).$$

This represents the log-normal distribution (which can be proved by applying log on both sides). In this study, for experimental purposes a comparison is made between the log-normal and the log-uniform distributions. The log-normal distributions is shown to have higher relevance than the log-uniform distributions.

Another research effort for modeling freeway incident clearance time, is part of a project that has been carried out at Northwestern University. The research by Wang et al. (1991), is an effort to develop an initial capability to provide successively improved incident clearance time predictions as time progresses. In this study, 121 incident records provided by the IDOT Communications Center are used. Because the data set is not large enough, the 1988 Chicago Area Expressway Accidents Annual Report (ACC), published by Chicago Police Department, is used to examine the validity of the study data before applying these data to make any interpretation. A series of χ^2 tests show that the study data set has a good representation and similar distribution to the large ACC data set. Incident reports are the original data used in this study. The first objective of the study is to find out what variables are important for the prediction of incident clearance time. Before making a detailed analysis for each variable, an overall incident clearance time regression model involving 22 variables is developed. Based on statistical assessment, nine variables are found to show strong relationship with the dependent variable, incident clearance time. These statistically significant variables are:

1) Operational Factors:

   Heavy Wrecker (WRECKER), Assistance From Other Response Agencies (OTHER), Sand/Salt, Pavement Operations (SAND).

2) Incident Type Factors:

   Number of Heavy Vehicle Involved (NTRUCK), Heavy Loading (HEAVY), Liquid or Uncovered Broken Loadings in Heavy Vehicles (NONCON), Severe Injuries in Vehicles (SEVINJ), Freeway Facility Damage Caused by Incident (RDSIDE).

3) Environmental Factors:

   Extreme Weather Condition (WX).

Other variables are found to be statistically insignificant. Among them are Number of Vehicles, Roll-Over Vehicle, Ratio of Lanes Blocked, Number of severe Injuries, Day of Week, Distance from CBD, Time of Day, Response Time (RESP), Second Patrol Arriving Time, Incident Report (HAR)

All the variables that are found to be statistically significant are included in the revised prediction model. The variables RESP and HAR, though statistically insignificant, are assumed to be important in improving incident operation study, thus are kept in the main model. The following is the main model (with HAR) developed by the study:

$$\begin{aligned} CLEAR &= 14.03+35.57*HEAVY+16.47*WX+18.84*SAND-2.31*HAR \\ &\quad +0.69*RESP+27.97*OTHER+35.81*RDSIDE+18.44*NTRUCK \end{aligned}$$

+32.76*NONCON+22.90*SEVINJ+8.34*WRECKER

Further, the concept of time sequential incident clearance time prediction models is adopted in the study for the reason that as time proceeds, further incident information becomes available for use in models, and more precise prediction results can be provided. A time sequence of the entire incident operation procedure and of those significant factors is defined. The variables are then grouped into four sequential incident clearance time prediction models. Predictions made by each model are the most accurately predict incident clearance times based on the most important information that would be available before a specified time point. According to the availability of the information at different time stages, 4 sequential incident prediction models are developed.

The number of incidents considered for the analysis (109 in total) is not significant enough for the model results to be considerate very reliable. The validation of these results have also been done with only 12 incidents, which is very few for a reliable validation process.

Another recent study on incident clearance prediction is also conducted in Northwestern University as part of the ADVANCE Project task (Sethi, et al., 1994). Information of 801 roadway incidents from Northwest Central Dispatch (NWCD) was used in the study. All the incidents are classified into the following incident types:

- Traffic stop/arrest (TSA)
- Motorist assist (stalled vehicles, flat tires, etc.) (MA)
- Accident with personal injuries (ACPI)
- Accidents involving property damage (ACPD)

- Severe incident (SI) including incidents with entrapment, with load spills and all fire related incidents.

Preliminary data analysis is performed to find the distribution of incident clearance times for all incidents. It is found that most incident clearance times, more than half the sample, are very short, under 10 minutes, but a substantial fraction of incidents (19.8%) lasts more than 40 minutes with some lasting more than an hour (10%). The preliminary analysis also examines the difference in average incident duration for different incident types, different roadway types and/or different levels of congestion on the roadway, represented by time of the day. The results of the data analysis show that incident duration is strongly associated with incident type but differences in clearance times by time of day and/or roadway type are relatively small for each incident type.

Detailed analysis is used to determine the differences in incident types within each group, roadway types (arterial vs. collector/local), traffic conditions represented by the time of the day, and the intensity of the incident measured by the number of vehicles dispatched. Analysis of Variance (ANOVA) is used to measure and test the statistical significance of differences in clearance times for each of these explanatory variables. Based on the statistical analysis results, an incident duration estimation decision tree shown below is developed.

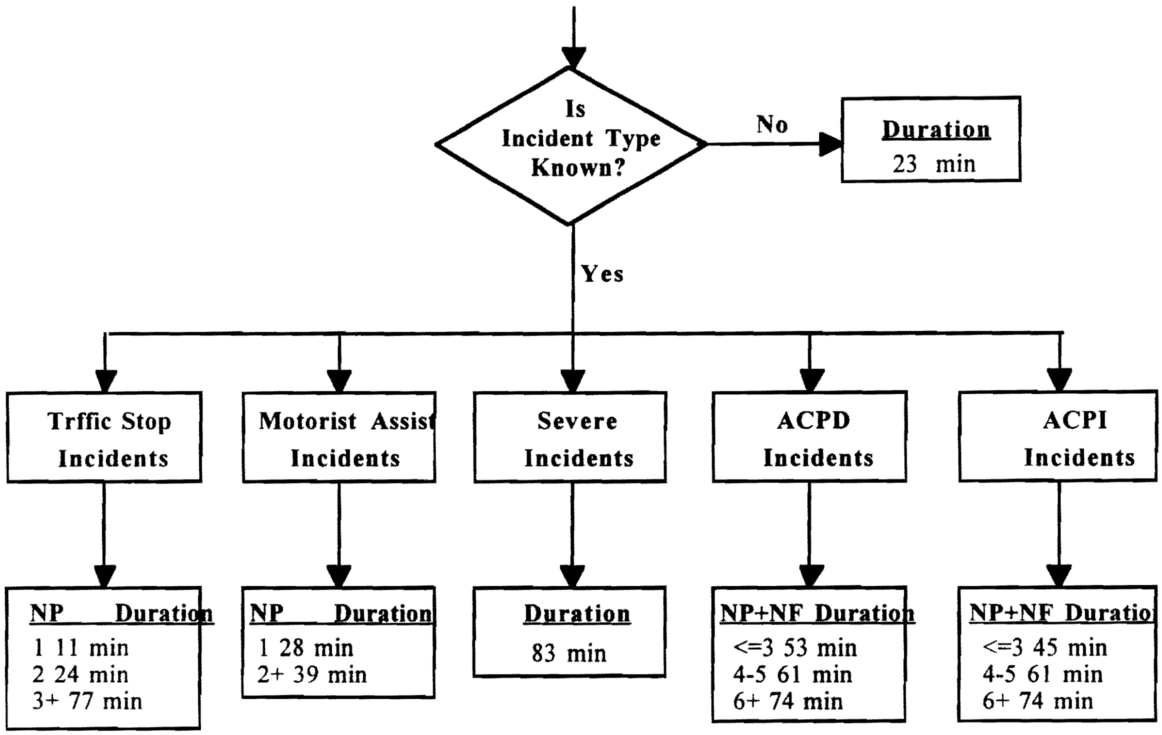


Figure 4.1: ADVANCE Project's Decision Tree for Incident Duration Prediction

The approach used in the ADVANCE Project is a simple, practical and effective. However, the classification of incidents is oversimplified. For example, classification of severe incidents is not clear. Severe incidents may include hazmat incidents, vehicle fire, truck incidents, fatality incident, and their corresponding clearance times may vary significantly. For each incident type, significant factors affecting clearance times are not fully explored. To better understand clearance characteristics, further study is needed.

4.2.1 Methodological Structure

In this research, survey forms are used for investigating factors which influence incident clearance duration and to develop a model to predict incident clearance duration for different incident types. The methodology is summarized in Figure 4.2. There are several advantages of using surveys in this study. It is the only way to obtain the type of data needed in this study. Incident management teams in Northern Virginia do not collect detailed data for each incident. As a matter of fact, the only type of data that is collected in Northern Virginia is the location, type, approximate starting and ending time of an incident, and the involvement of a tractor trailer. However, in this research, with the help of the domain expert, we have identified many parameters that might affect the incident clearance times. Thus, the survey research provides adequate richness of data and ability to study factors that affect clearance times at a detailed level.

A shortcoming of using survey forms is the errors caused by the participants namely, incident management personnel. Other important problem is the existence of inconsistencies among very similar incident cases which might be due to the operational differences among different agencies and individual teams in dealing with similar incident cases.

4.2.2 Structure and Design of the Survey Forms and Data Collection

Collection and analysis of incident data were undertaken as part of the initial effort of this incident management project to understand incident clearance characteristics. The details of these efforts can be found in Subramaniam et al. (1994). The aim of this section is to briefly summarize the early work for designing the survey forms and collecting the incident data. However, the emphasis in this chapter is on the analysis of new data, and the development of incident clearance prediction / decision trees using the results of this analysis.

Description of the Design of Survey Forms and First Phase of Data Collection

Survey forms were prepared as a result of extensive meetings with the Northern Virginia Incident management experts and the review of available incident management literature. The details of this effort is described in Subramaniam et al. (1992) and will not be discussed here.

Data was collected from incident clearance survey forms filled out by members of the participating agencies. The following agencies are involved in the project: Virginia State Police, VDOT Safety Service Patrol, Fairfax County Police Department, and Fairfax County Fire and Rescue Department.

About 10,000 survey forms were distributed to the above mentioned agencies to be completed whenever members of these agencies participate in clearing an incident.

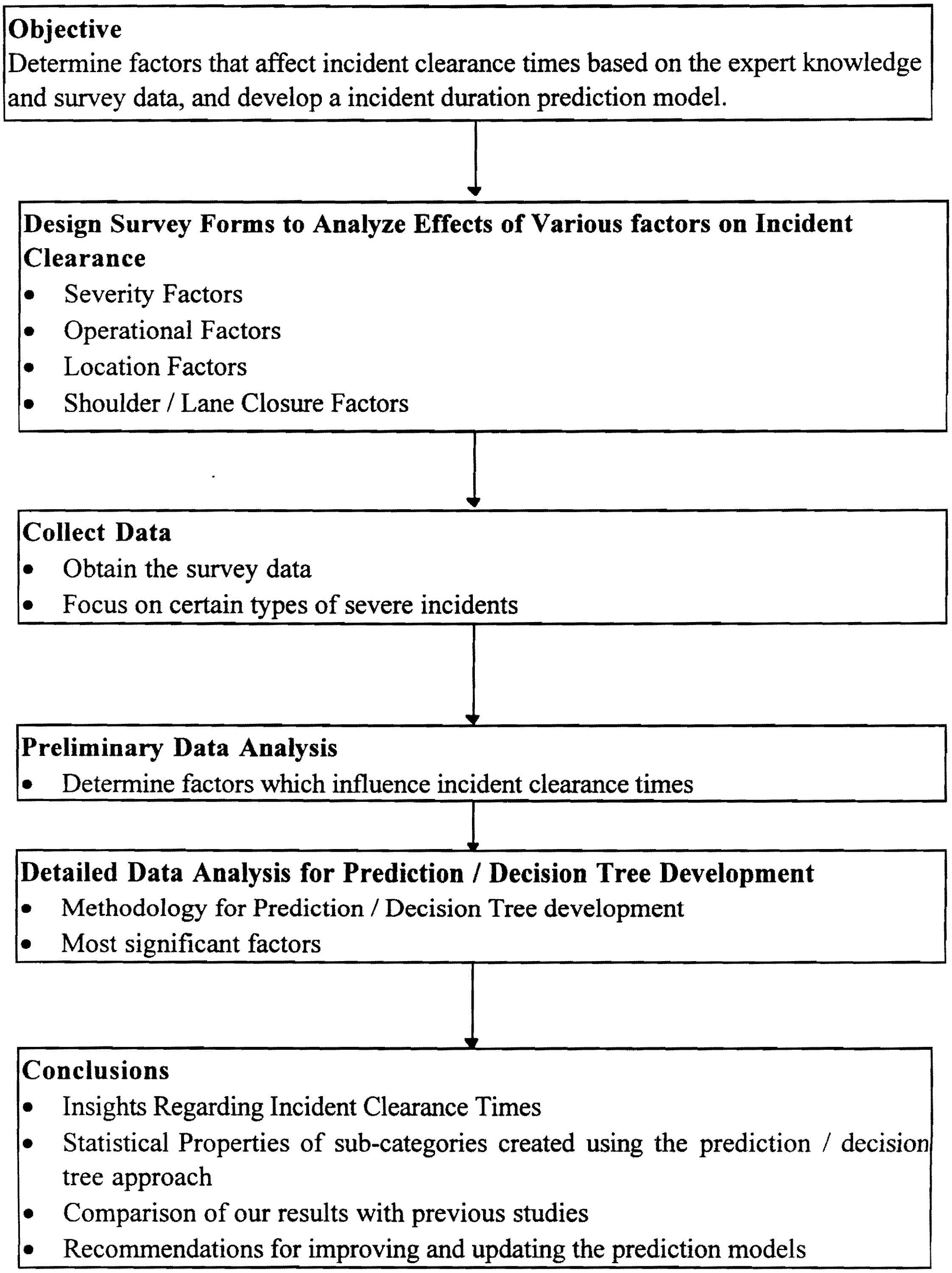


Figure 4.2. Methodology for Incident Duration Prediction

Members were also instructed on how to complete the survey forms. Over 1000 personnel belonging to different agencies participated in completing the survey forms during roadway incidents. The data collection began in April, 1994. Persons filling out the survey forms were asked to submit completed forms to their respective supervisors, who mailed the forms back to Virginia Tech, where the information contained in the forms was encoded into computer data files to be analyzed.

Samples of Data Collected

The information collected from the survey forms includes the name and agency of the person filling out the survey form, date and time of incident occurrence, incident details such as the type and nature, number and type of vehicles involved, number of injuries and fatalities, prevailing weather conditions, location characteristics, lane closure information, resources used for clearing the incident, and the clearance time.

Incident Clearance Time

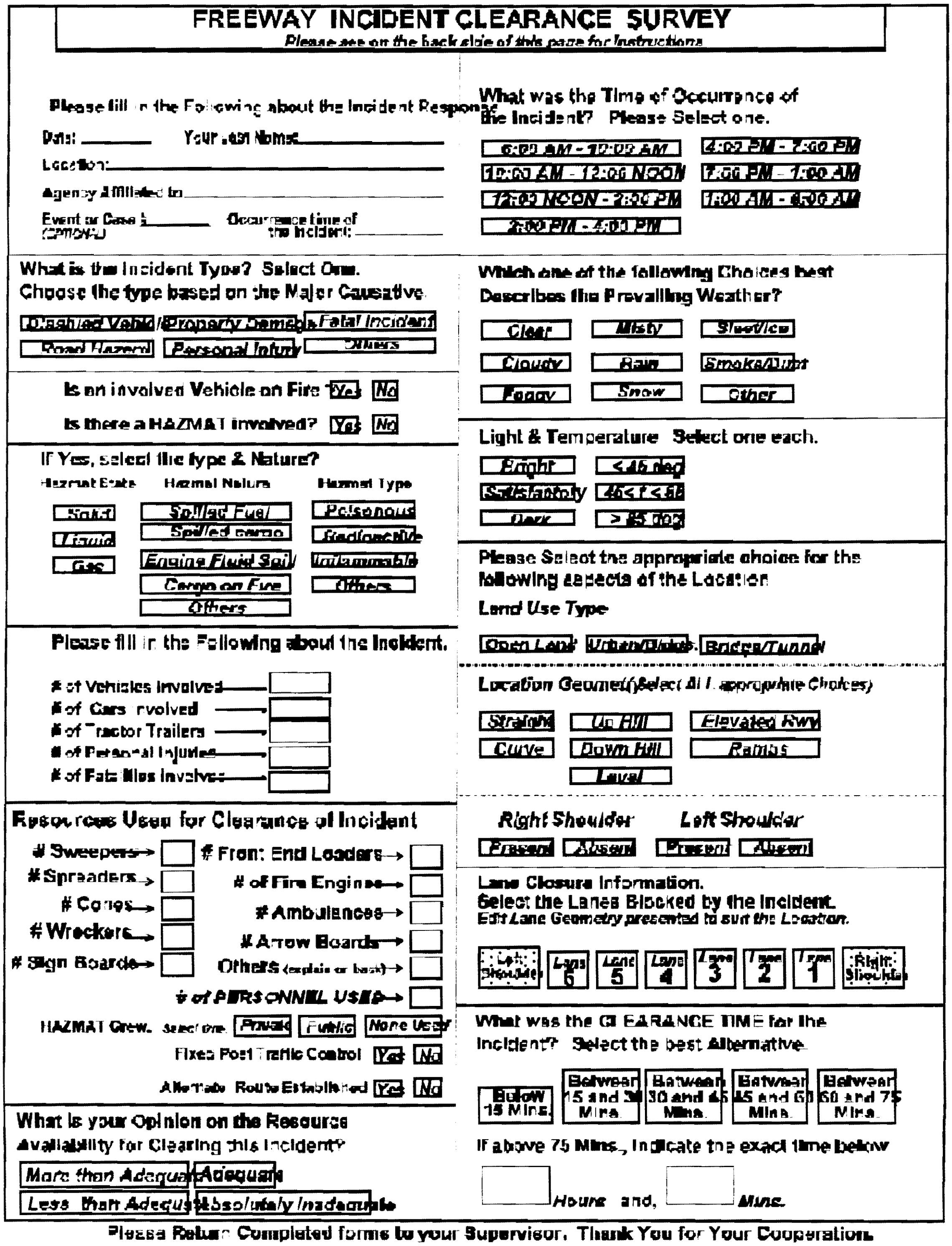
Incident clearance time is defined as the time from incident identification to the time at which the last emergency vehicle leaves the scene. Incident clearance time is used as a proxy for incident duration which should include the time from beginning of an incident until the normal roadway capacity is restored. The time from the moment that an incident occurs to the time that incident is verified and response starts is normally short in Northern Virginia Area and does not vary significantly for different incident cases, thus, this time interval is ignored in our analysis. The time interval between the moment that an incident is cleared (response units leave the scene) to the time when normal capacity is restored is affected by the traffic flow condition of that time and number of delayed vehicles. A sample survey form is shown in Figure 4.3.

Data Issues

Various issues involving the transfer of information from the returned survey forms into computer data files, the quality of the data, and the problems faced in extracting quality data from the survey forms were needed to be solved.

As the first step in analyzing the incident data, information contained in the completed survey forms was transferred to computer data files. One of the biggest problems faced was ensuring the quality of the data. Many survey forms were incomplete. For example, in the case of incidents involving disabled vehicles, in many survey forms the type of vehicle was not marked. Lane closure information, which is one of the most important factors influencing incident clearance times, was not properly filled out in a large number of survey forms. In some survey forms, the prevailing weather, light, and temperature conditions were missing, while in some others, more than one box was checked for these conditions.

Records with such incomplete and conflicting information about incident clearance, which, if included, would have seriously affected the outcome of the data analysis, were removed from the data sets. Only about a half of the returned survey forms completed by VDOT Safety Service Patrol were of some quality and were therefore usable. About 70 percent of the survey forms filled out by Fairfax Co. Police Department were usable. 88 percent of the forms filled out by Fairfax Co. Fire and Rescue Department and 97 percent of the forms filled out by Virginia State Police were usable.



FREEWAY INCIDENT CLEARANCE SURVEY

Please see on the back side of this page for Instructions

Please fill in the Following about the Incident Response:

Date: ______ Your Last Name: ______

Location: ______

Agency Affiliated to: ______

Event or Case # (OPTIONAL) ______ Occurrence time of the Incident: ______

What was the Time of Occurrence of the Incident? Please Select one.

6:00 AM - 10:00 AM | 10:00 AM - 12:00 NOON | 12:00 NOON - 2:00 PM | 2:00 PM - 4:00 PM | 4:00 PM - 7:00 PM | 7:00 PM - 1:00 AM | 1:00 AM - 6:00 AM

What is the Incident Type? Select One. Choose the type based on the Major Causative

Disabled Vehicle | Property Damage | Fatal Incident | Road Hazard | Personal Injury | Others

Is an involved Vehicle on Fire? Yes | No

Is there a HAZMAT involved? Yes | No

If Yes, select the type & Nature?

Hazmat State	Hazmat Nature	Hazmat Type
Solid	Spilled Fuel	Poisonous
Liquid	Spilled cargo	Radioactive
Gas	Engine Fluid Spill	Inflammable
	Cargo on Fire	Others
	Others	

Which one of the following Choices best Describes the Prevailing Weather?

Clear | Misty | Sleet/Ice | Cloudy | Rain | Smoke/Dust | Foggy | Snow | Other

Light & Temperature Select one each.

Bright | < 45 deg | Satisfactory | 45 < T < 85 | Dark | > 85 deg

Please Select the appropriate choice for the following aspects of the Location

Land Use Type

Open Land | Urban/Dense | Bridge/Tunnel

Location Geometry (Select All appropriate Choices)

Straight | Up Hill | Elevated Rwy | Curve | Down Hill | Ramps | Level

Please fill in the Following about the Incident.

\# of Vehicles Involved →

\# of Cars Involved →

\# of Tractor Trailers →

\# of Personal Injuries →

\# of Fatalities Involved →

Resources Used for Clearance of Incident

\# Sweepers → | # Front End Loaders →

\# Spreaders → | # of Fire Engines →

\# Cones → | # Ambulances →

\# Wreckers → | # Arrow Boards →

\# Sign Boards → | Others (explain on back) →

\# of PERSONNEL USED →

HAZMAT Crew. Select one. Private | Public | None Used

Fixed Post Traffic Control Yes | No

Alternate Route Established Yes | No

Right Shoulder Present | Absent *Left Shoulder* Present | Absent

Lane Closure Information.

Select the Lanes Blocked by the Incident.

Edit Lane Geometry presented to suit the Location.

Left Shoulder | Lane 6 | Lane 5 | Lane 4 | Lane 3 | Lane 2 | Lane 1 | Right Shoulder

What is your Opinion on the Resource Availability for Clearing this Incident?

More than Adequate | Adequate | Less than Adequate | Absolutely Inadequate

What was the CLEARANCE TIME for the Incident? Select the best Alternative.

Below 15 Mins. | Between 15 and 30 Mins. | Between 30 and 45 Mins. | Between 45 and 60 Mins. | Between 60 and 75 Mins.

If above 75 Mins., Indicate the exact time below

______ *Hours* and, ______ *Mins.*

Please Return Completed forms to your Supervisor. Thank You for Your Cooperation.

Figure 4.3: Sample Survey Form (Source: Subramaniam et al. 1994)

Experimental Design and Statistical Analysis of the Old Data (First Phase Data Collection)

Statistical Analysis Software (SAS) was used for analyzing the data (Subramaniam et al. 1994). Incidents were grouped into six different categories as Disabled Vehicles, Road Hazards, Property Damage, Personal Injuries, Fatalities, and Other Incidents.

As part of preliminary analysis of the data, the average clearance times were calculated to compare the clearance times by incident type and agency. Table 4.1 shows the results of this analysis. The average clearance times by all data sets (agencies) for each incident type is as follows:

Disabled Vehicles:	17.5 minutes;	Property Damage:	33.6 minutes;
Road Hazard:	19.7 minutes;	Personal Injury:	41.9 minutes;
Fatalities:	65.6 minutes;	Other Incidents:	15.2 minutes.

Table 4.1: Average Clearance Times In Minutes By Incident Type and Agency

(Figures in parentheses are the number of incident cases)

	Fairfax Co. Police Dept.	**Fairfax Co. Fire&Rescue**	**Safety Service Patrol**	**VA State Police**	**Surveys from Sept. 1992**
Disabled Vehicle	21.5 (46)	35.4 (7)	13.5 (373)	17.9 (739)	19.7 (400)
Road Hazard	36.0 (10)	- (0)	12.5 (24)	21.8 (43)	17.9 (46)
Property Damage	30.7 (203)	29.3 (51)	24.0 (59)	39.6 (165)	36.6 (135)
Personal Injury	40.7 (85)	33.7 (75)	52.5 (23)	53.6 (54)	39.6 (63)
Fatalities	- (0)	67.5 (1)	67.5 (1)	105.0 (2)	45.0 (4)
Other Incidents	16.5 (10)	22.5 (2)	9.5 (2)	14.4 (128)	16.3 (68)

It can be seen that clearance times vary considerably with incident type, and are comparable between agencies. In the case of Fairfax County Fire and Rescue Department, the clearance times are a little higher because fire and rescue units are called upon only in case of special circumstances, thus increasing the clearance times.

Fairfax County Police and Fire & Rescue Departments have mostly dealt with incidents involving property damage or personal injuries, while a great majority of incidents that involved Safety Service Patrol and Virginia State Police were disabled vehicles. Less than one percent of all incidents involved fatalities. Once the fact that incident type considerably affects clearance time was established, for each incident type regression analyses were undertaken to model clearance times on various factors that were thought to affect clearance time from consultations with experts. Clearance time for each incident type was modeled based upon the following factors.

- lane closure
- number of cars involved
- number of trucks involved
- number of personal injuries
- number of fatalities
- hazmat involvement
- fire involvement
- time frame of occurrence
- prevailing weather
- prevailing temperature
- land use type

This statistical analysis is conducted to study the effect of the various "hypothesized" variables on the clearance of incidents. As expected, several of these variables have significant effects on clearance times. And, these follow different priorities or patterns depending upon the incident types. In general, this study clearly establish the fact that the clearance times of incidents are affected by the incident types. This was quite evident from the "clearly" different mean and median values of clearance times obtained. Besides, the low variances observed reinforces the "representativeness" of the mean in the data sets. As part of this study, linear regression models are developed to model the incident duration. These models that are presented in detail in Subramaniam (1995) have unfortunately very low R-square values and the possible reasons for this are also explained in the same report. Based on the results of these surveys and expert advise, it is decided that the second phase of the surveys commencing from July 15, 1994 is to be different from the initial survey effort.

New Survey Form

Starting on April 1, joint survey on incident management operations is conducted by Virginia Tech, VDOT, Virginia State Police, Fairfax County Police Department, and Fairfax County Fire and Rescue. The second phase of data collection started on July, 1994 and continued until the beginning of 1995. Survey forms have been filled out and data analysis aimed at developing models to predict incident clearance times is being developed at Virginia Tech. Based on the preliminary statistical data analysis and examination of problems found in the forms during data entry and feedback obtained from the agencies involved in the survey, a new survey form is designed to try and obtain better results. The following are some of the major changes implemented in the new survey:

- Removal of ambiguous and redundant questions,

- Additions of questions that are more relevant based on old survey results,
- Elimination of "open choices" and changing question options to "force" responses into the study variable categories or requirements.
- Clarification of instructions to minimize incomplete responses and improve responsiveness, and,
- To discontinue survey of disabled cars on shoulders of roadways since ample data is already available and to reduce workload on respondents.

Various changes are made to the old survey form to generate the new one. The reasons for the changes and a copy of the new survey form are presented in Subramaniam et al., (1994). The most important changes made on the new form are:

1. the splitting of disabled vehicle into disabled car and disabled truck,
2. the omission of disabled car on shoulder incident type since normal disablement on shoulders accounts for over 60% of all incidents and data is already available in plenty
3. the addition of disabled truck and disabled car in travel lane as two different incident types instead of disabled vehicle item on the old form.
4. the addition of exact clearance time question instead of time intervals motivated by the statistical analysis of the old data.

4.2.3 Analysis of New Incident Data

Over 900 formed have been filled by the northern Virginia incident management personnel and returned to the Center for Transportation Research during the second phase of data collection. The data items with serious errors or with vital attributes (e.g., incident type) missing are eliminated during the data entry process. As a result of this processing, a new data set consisting of 650 incidents is obtained. This new data set is used for the data analysis and the development of decision / prediction trees described in this section.

First, linear regression technique is used to model the incident duration. However, due to the wide variation of clearance time data, a low R-square value (approximately 0.35) is obtained by using different combinations of variables in prediction model. The data is also tested to determine if it follows either log-normal or log-logistic distributions described respectively in Golob et al. (1987) and in Jones et al. (1991). The result of statistical significance tests show that the data does not follow either one of these distributions. Next, new approaches for this prediction including neural networks and prediction / decision trees are undertaken. Among these approaches, decision / prediction tree approach have provided a good and simple way of modeling the duration prediction problem with the available data.

In the following sections, first the methodology adopted for developing the prediction / decision trees is described. Then, the variables considered are described, followed by the preliminary and detailed data analysis results. Next, results and conclusions of the data analysis are presented and discussed. Finally, prediction / decision trees developed to estimate incident clearance times and their validation are presented.

4.2.3.1 Methodology for Developing Prediction / Decision Trees

The methodology adopted for developing prediction / decision trees is very similar to the one used by the ADVANCE Project. However, there are some major improvements:

- The data set does not contain normal car disablements on shoulder and minor traffic violation stop and arrest incidents. These incidents account for the majority of the total incidents and do not have much impact on traffic. Consistent results have been achieved in the first stage of data analysis (Subramaniam et al., 1994).
- An improved and more detailed incident type classification method is used in this study.
- Larger number of significant variables have been explored within each incident type group.

The methodology for developing prediction / decision trees is a simple one. The process first considers the complete data set X. It then examines all possible splits of X into two or more subsets based on the values of the independent variables and the selection of best split. The decision for the best split at each step is made by using the Analysis of Variance (ANOVA). A tree is then constructed through a series of splits. This process of constructing a prediction / decision tree is very similar to the classification / regression tree approach (CART) described in detail in Breiman et al., (1984) and Stewart R.J., (1996). The major difference is that CART uses a binary split while our classification methodology allows non-binary splits. Also, CART uses the mean squared error as the cost function associated with a tree having K terminal nodes. Each binary split in CART is chosen as the split that produces the greatest reduction in mean squared error cost. This is also similar to using ANOVA which can be described as a technique to analyze total variation among different populations.

4.2.3.2 Variable Description

The following are candidate variables considered to have impact on incident clearance times.

Incident Type factors

INCTYPE Incident Type

INCTYPE = 1:	Road Hazard Incident
INCTYPE = 2:	Personal Injury Incident
INCTYPE = 3:	Personal Injury Incident
INCTYPE = 4:	Vehicle Fire Incident
INCTYPE = 5:	Fatal Incident
INCTYPE = 6:	Cargo Spill Incident
INCTYPE = 7:	Hazmat Incident
INCTYPE = 8:	Weather Related Incident
INCTYPE = 9:	Construction/Maintenance
INCTYPE = 10:	Disabled Car in Travel Lane Incident

Incident Detail/Severity Factors:

NCARS:	Number of Cars Involved in the Incident
NTRUCKS:	Number of Trucks Involved in the Incident
NINJUR:	Number of Injuries Involved in the Incident
NFATAL:	Number of Fatalities Involved in the Incident

Operational factors

POLICVEH:	Number of Police Vehicles on the Scene
FIREENG:	Number of Fire Engines on the Scene

AMBUL: Number of Ambulances on the Scene

WRECKER: Number of Wreckers Used

ARRBRDS: Number of Arrowboards Used

MEDVAC: MEDVAC Helicopter Requested? (y/n)

FIXPOST: Fixed Post Traffic Control? (y/n)

ALTROUT: Alternate Route Established? (y/n)

Location Factors

RWTYPE: Roadway Type

RWTYPE = 1:Freeway

RWTYPE =2: Nonfreeway

LANDUSE: Land Use Type

LANDUSE = 1: Open Land

LANDUSE =2: Urban/Buildings

LANDUSE = 3: Bridge

LANDUSE = 4: Tunnel

Shoulder/Lane Closure Factors:

SHOULDER: Shoulders Information

SHOULDER =1: Shoulder Presence

SHOULDER =0: Shoulder Absence

LANCLOSR: Ratio of Lanes Closed to Total Number of Lanes

Environmental Factors

WEATHER: Weather Condition

WEATHER = 1: Clear/Cloudy

WEATHER = 2: Rain

WEATHER = 3: Foggy/Misty

WEATHER = 4: Snow

WEATHER = 5: Sleet/Ice

LIGHT: Light Condition

LIGHT = 1: Bright

LIGHT = 2: Satisfactory

LIGHT = 3: Dark

TEMP: Temperature

TEMP =1: <45 Degree

TEMP =2: 45~85 Degree

TEMP =3: >85 Degree

Variable CLT is the value of incident clearance time in minutes.

4.2.3.3 Preliminary Analysis

The frequency distribution of incident clearance times and summary information for all incidents is shown in Figure 4.4. This figure shows that most incident clearance times are within one hour, but still a substantial fraction (15.0%) of incidents lasts more than 1 hour, some severe incidents (about 0.8%) even last more than 3 hours. The mean is 44 minutes and the standard deviation is 33.8 minutes. The distribution is skewed as opposed to approximately normal in the case of some other studies such as Jones et al. (1991); this may be because our data set does not contain large number of longer incidents. Also, the data set at this point is not homogenous enough and contains many types of incidents. Most of the previous incident duration studies where a normal or approximately normal distributions are found concentrated on one type of incidents such as, truck incidents (Giuliano G., 1989).

The first step of the preliminary analysis is determine the characteristics of incidents that have different clearance times and to examine the differences in average incident clearance times for different incident types and different roadway types. The frequency distributions of incident clearance times for freeway and non-freeway incidents are presented in Figures 4.5 and 4.6. The frequency distributions of incident clearance times for different types of incidents are presented in Figures 4.7-4.12. In the whole data set, there are only three cases of Hazmat incidents with clearance times of 135, 180, 417 minutes respectively. Only two cases of weather related incidents are recorded, with clearance times of 15 and 150 minutes respectively. Therefore, we don't have enough data points for analyzing Hazmat and weather related incidents. Average clearance times by incident type are summarized in Table 4.2.

CLT	Frequency
10	50
20	92
30	142
40	51
50	110
60	107
70	22
80	29
90	17
100	2
110	4
120	12
130	1
140	2
150	1
160	0
170	1
180	2
More	5

Observation	650
Min	7
Max	417
Mode	30
Std	33.84678
Mean	45.03077

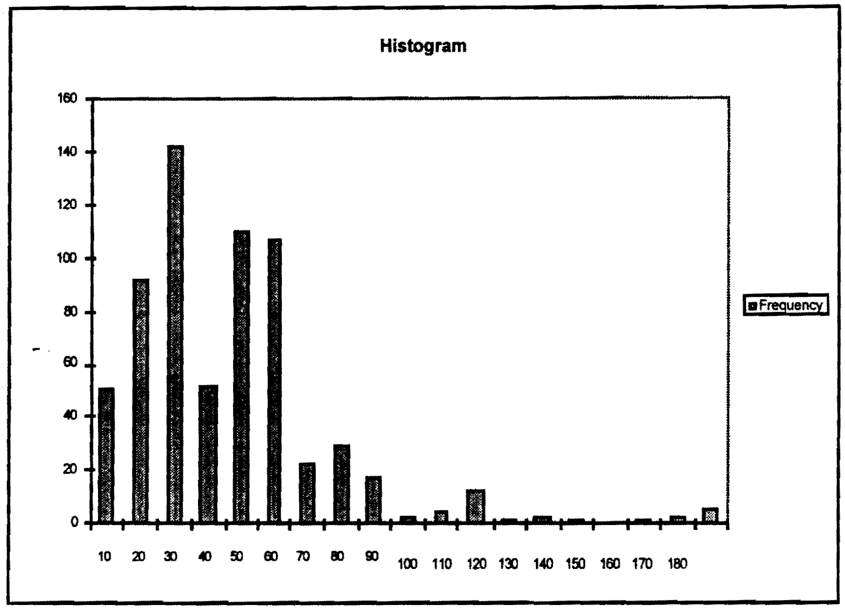


Figure 4.4: Incident Clearance Time Distribution

CLT	Frequency
10	40
20	68
30	124
40	42
50	101
60	97
70	15
80	24
90	17
100	0
110	1
120	12
130	0
140	1
150	1
160	0
170	1
180	2
More	4

Count	550
Min	7
Max	285
Mean	45.25
Mode	30
STD	30.84

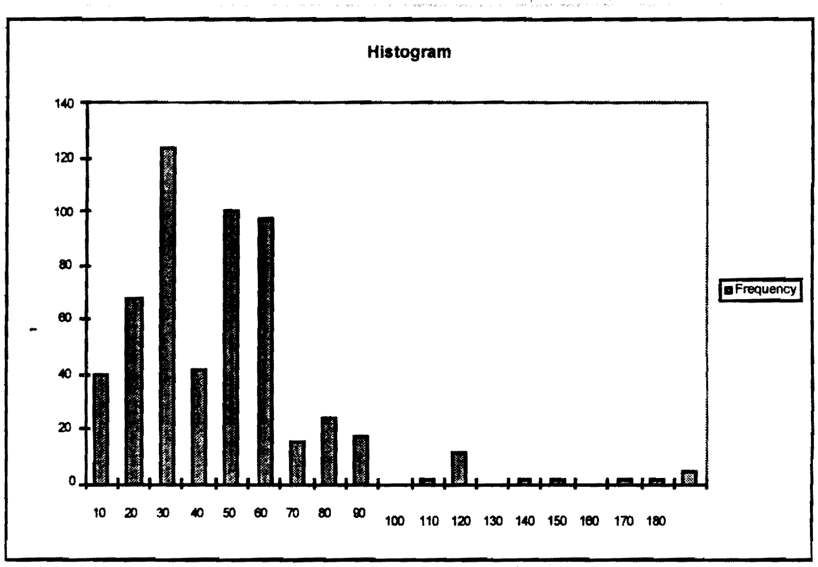


Figure 4.5: Freeway Incident Clearance Time Distribution

CLT	Frequency
10	10
20	24
30	18
40	9
50	9
60	10
70	7
80	5
90	0
100	2
110	3
120	0
130	1
140	1
150	0
160	0
170	0
180	0
More	1

Count	100
Min	7
Max	417
Mean	43.8
Mode	30
STD	47.25794

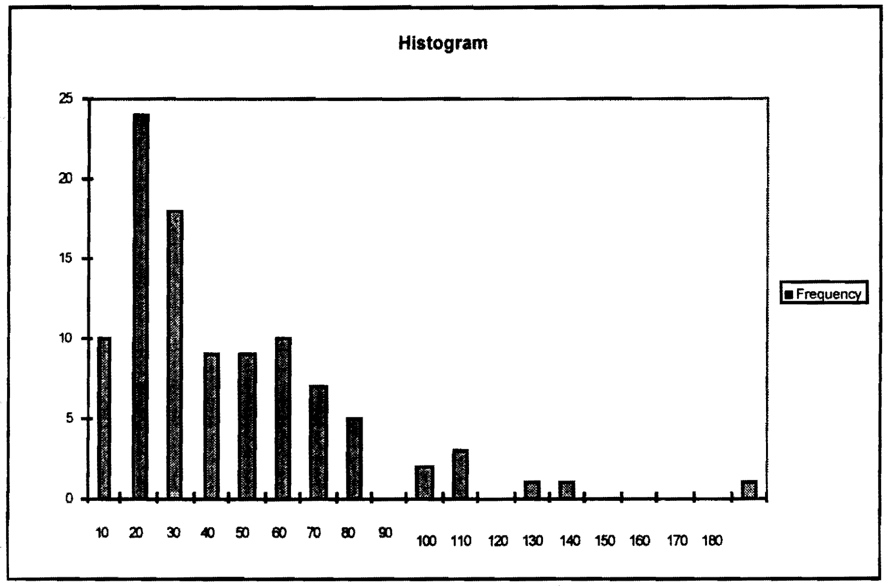


Figure 4.6: Non-Freeway Incident Clearance Time Distribution

CLT	Frequency
10	2
20	4
30	0
40	1
50	0
60	1
70	1
More	0

Count	9
Min	7
Max	61
Mean	26.89
Mode	15
STD	20.61

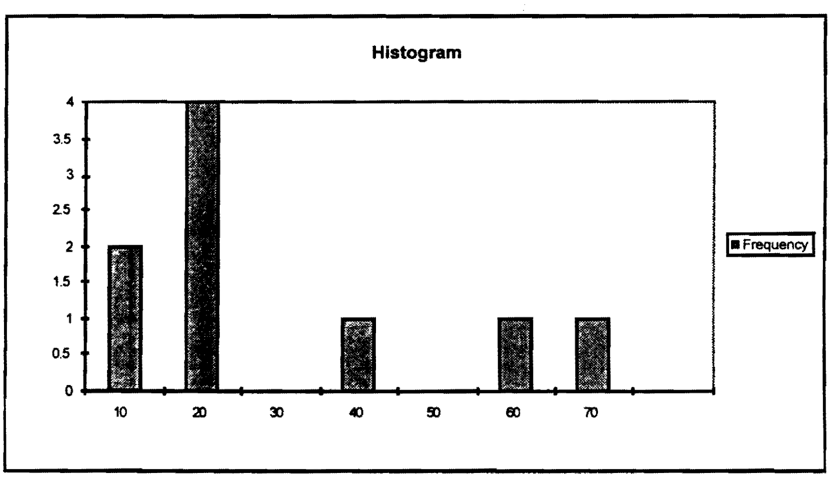


Figure 4.7: Road Hazard Incident Clearance Time Distribution

CLT	Frequency
10	31
20	52
30	101
40	26
50	69
60	73
70	16
80	14
90	7
100	0
110	4
120	6
130	0
140	0
150	0
160	0
170	0
180	0
More	2

Count	401
Min	7
Max	270
Mean	42.61
Mode	30
STD	26.67

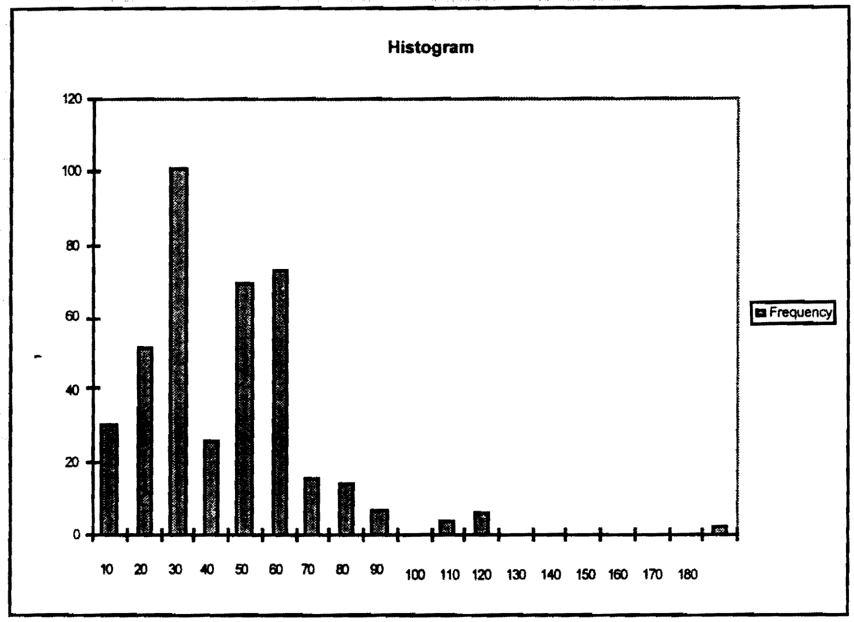


Figure 4.8: Property Damage Incident Clearance Time Distribution

CLT	Frequency
10	9
20	13
30	31
40	19
50	34
60	29
70	4
80	14
90	10
100	2
110	0
120	4
130	0
140	1
150	0
160	0
170	1
180	1
More	1

Count	173
Min	10
Max	195
Mean	50.81
Mode	45
STD	30.24

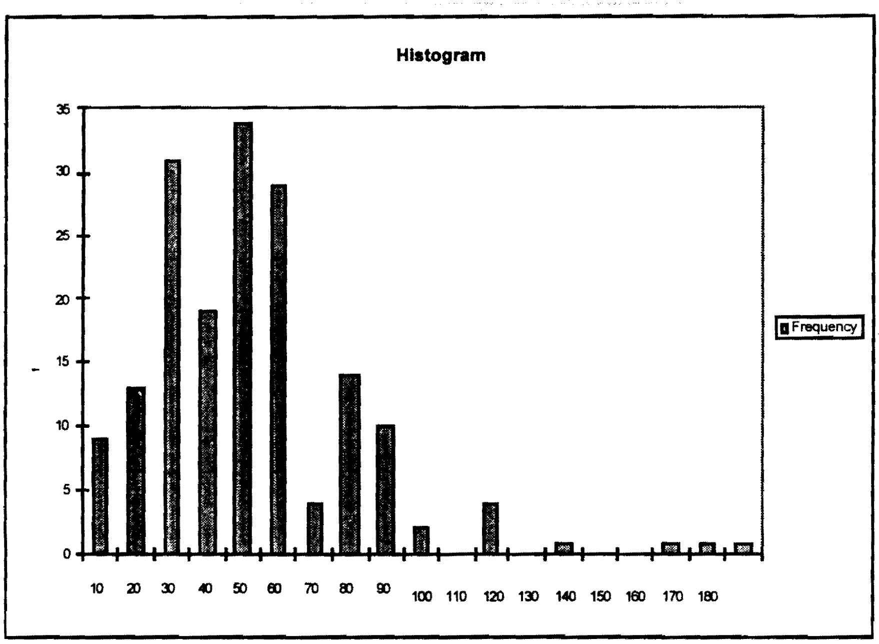


Figure 4.9: Personal Injury Incident Clearance Time Distribution

CLT	Frequency
10	3
20	5
30	1
40	0
50	0
60	0
70	1
80	0
90	0
100	0
110	0
120	2
130	1
140	0
150	0
160	0
170	0
180	0
More	1

Count	14
Min	8
Max	285
Mean	60.29
Mode	8
STD	79.68

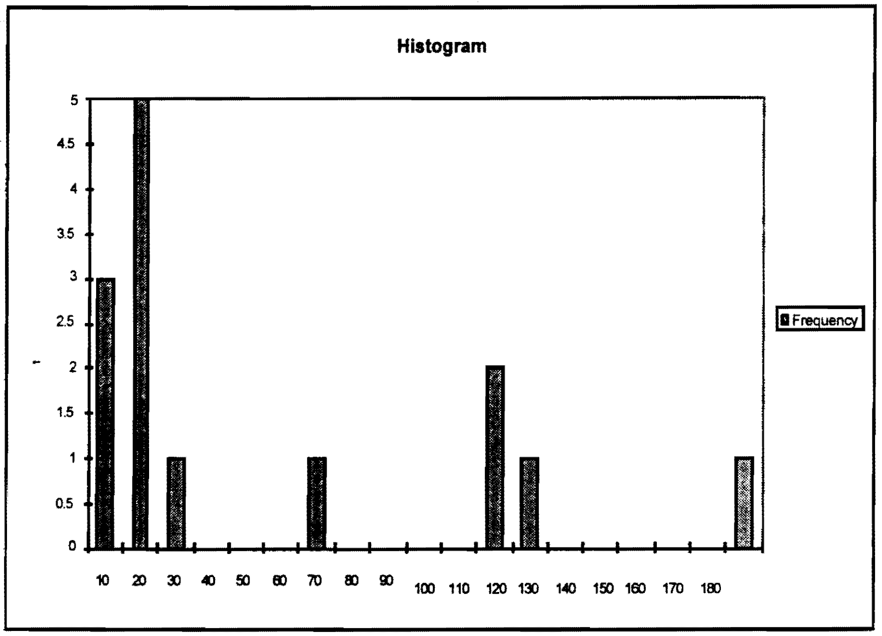


Figure 4.10: Disabled Truck Incident Clearance Time Distribution

CLT	Frequency
10	0
20	2
30	0
40	1
50	2
60	1
70	0
80	1
More	0

Count	7
Min	15
Max	75
Mean	42.57
Mode	45
STD	21.05

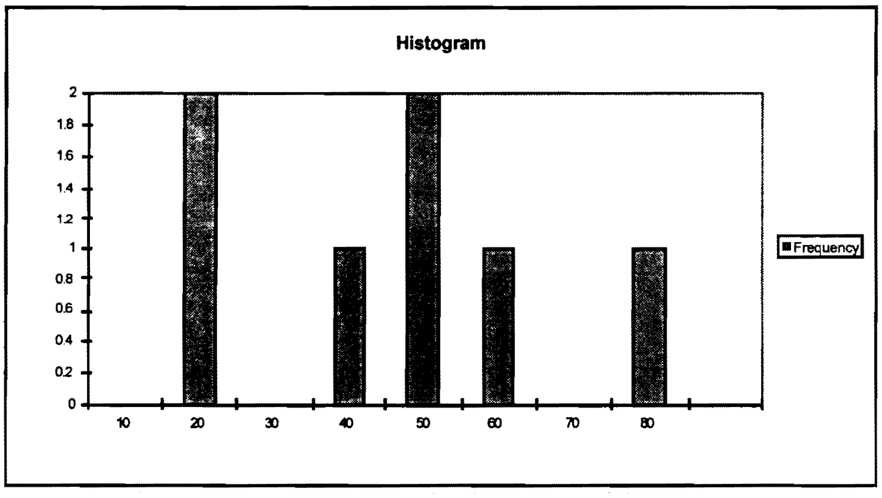


Figure 4.11: Vehicle Fire Incident Clearance Time Distribution.

CLT	Frequency
10	5
20	15
30	9
40	4
50	5
60	3
More	0

Count	41
Min	7
Max	60
Mean	27.17
Mode	15
STD	15.10

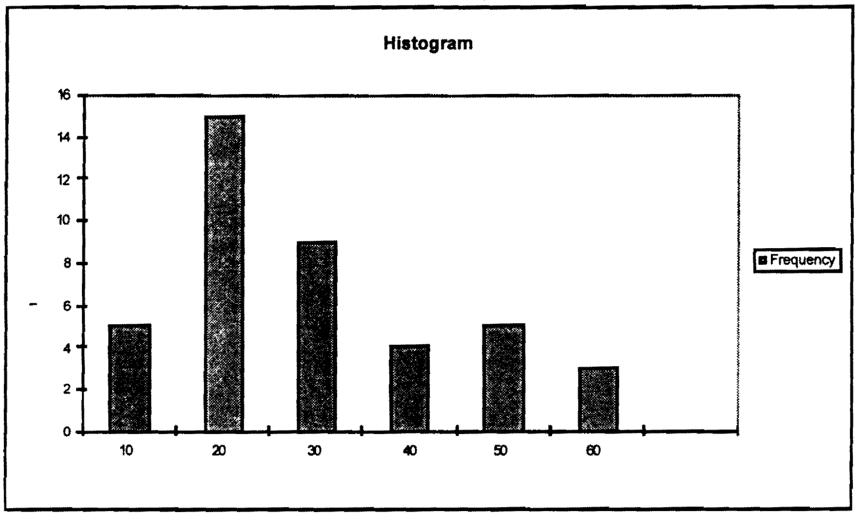


Figure 4.12: Disabled Car in Travel Lane Incident Clearance Time Distribution

Table 4.2: Average Clearance Times By Incident Type (Minutes)

Incident Type	Road Hazard	Property Damage	Personal Injury	Disables Truck	Vehicle Fire	Hazmat	Weather Related	Disabled Car	All
Mean (min.)	26.89	42.61	50.81	60.29	42.57	244.00	82.50	27.17	45.03
# of Cases	9	401	173	14	7	3	2	41	650

The results of preliminary analysis show that incident clearance times are strongly associated with incident types but differences in clearance times by roadway type are relatively small. Both the average values or distribution patterns differ significantly for each incident type for most cases. For example, the average clearance time for Hazmat incidents is over 6 hours while car disablements only last less than half hour on average.

The ANOVA results (Appendix-A.0) also strongly suggest the need to consider the incident type as the major factor for additional analysis.

Further, the results of the preliminary analysis of the incident clearance times indicate that there is a wide distribution of incident clearance times even for incidents of the same type. This supports the need for additional analysis using several factors such as, incident type, severity of incidents, operational, roadway type, and environmental factors.

4.2.4 Detailed Analysis

The aim of the detailed analysis is to determine the effects of factors such as, incident type, incident severity, operational, location, lane/shoulder closure, and environmental factors on the incident clearance times. Analysis of Variance (ANOVA) is used to measure and test the statistical significance of differences in incident clearance times for each of these explanatory variables. Prediction / Decision Trees are then built based on the results ANOVA analysis. For groups that don't have large enough sample size to run statistical significance tests, average values and an interval of 25% and 75% percentile as lower and upper limits are used for the purpose of prediction / decision tree development.

4.2.4.1 Road Hazard

The first type of incident type that is going to be analyzed is the "road hazard". Since there are only 9 cases of road hazard incident, carrying out statistical significance test cannot be carried out. Average values, and 25% and 75% percentiles are used as estimates of the clearance times (Table 4.3).

Table 4.3: Clearance Time Estimation for Road Hazard Incidents

Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
7	61	27	15	35

4.2.4.2 Property Damage

The second type of incidents that are studied are "property damage" incidents. Many incident detail / severity factors described in the previous section that might affect the clearance duration of "property damage" incidents are considered. The impact of these factors on clearance time is investigated using ANOVA tests. Among these factors, truck involvement and number of vehicles involved are found to affect the incident clearance times the most. The results of ANOVA tests (See Appendix-A.1) show that truck involvement has more impact on clearance times than the number of vehicles involved. Thus, the data is first grouped into two sub categories: "Incidents with truck involvement" and "incidents without truck involvement". Table 4.4 shows the impact of truck involvement on clearance time estimation

Table 4.4: Impact of Truck Involvement on Clearance Time Estimation

	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
Truck Involved	8	210	51	30	60
No Truck Inv.	7	270	42	30	70

The next step is to determine other significant factors for each sub-category using ANOVA. As shown in Appendix-A.2, it is found that number of vehicles involved in the incident is a significant factor affecting clearance time of property damage incidents without truck involvement, especially, when number of vehicles involved is four or more, the clearance time increases significantly. However, in the sub-group of property damage "incidents with truck involvement", neither number of trucks involved nor number of vehicles involved turn out to be significant (See Appendix-A.3 and A.4). Table 4.5 shows the impact of number of vehicles involved on clearance time estimation for property damage without truck involvement.

Table 4.5: Impact Of Number Of Vehicles Involved On Clearance Time Estimation

No. of Vehicles Involved	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
1-3	8	210	41	30	60
4, or 4+	7	270	68	40	80

<u>Operational Factors</u>

Among the operational factors, number of police vehicles on scene (POLICEVEH) is found to be significant for both the sub-categories of number of cars involved, namely "1-3 cars involved" and "4 or more cars involved" (Appendix-A.5 and A.6). Table 4.6 shows the impact of number of police vehicles involved on clearance time estimation for property damage incidents without truck involvement.

The next step is to study the effect of other agencies' involvement on "property damage" incident clearance times. The involvement of other agencies is reflected by the involvement of wreckers, ambulances and fire engines. For property damage incidents,

Table 4.6: The Impact of Number of Police Vehicles Involved on CLT Prediction

No. of Vehicles Involved	POLICEVEH	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
1-3	1	7	120	38	25	50
	2	10	90	46	35	60
	3, or 3+	10	120	65	45	90
4, or 4+	1	15	80	39	20	60
	2, or 2+	30	270	97	60	110

only the impact of wrecker involvement is considered since ambulances are requested normally for personal injury incidents and fire engines are only requested for vehicle fire incidents. The results (Appendix-A.7) show that wrecker and ambulance involvement significantly increases incident clearance times for the cases of 1-3 vehicle involved and 1 police car involved. However, for the cases of more than 1 police vehicle involved or more than 3 vehicles involved, the impact of wrecker involvement is either statistically insignificant (Appendix-A.8) or unknown due to insufficient sample size. Table 4.7 shows the impact of wrecker involvement on clearance time estimation for property damage without truck involvement and 1-3 vehicles and 1 police cars involved. Tables 4.8 and Appendix-A.9 show that the impact of wrecker involvement on clearance time estimation for property damage incidents with truck involvement is quite significant.

Table 4.7: The Impact of Wrecker Involvement on CLT Prediction for Property Damage Incidents without Truck Involvement and 1-3 Vehicles and 1 Police Car Involved

Wreckers Involved ?	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
No	7	105	35	25	45
Yes	7	120	44	35	60

Table 4.8: The Impact of Wrecker Involvement on CLT Prdiction for Property Damage with Truck Involvement

Wreckers Involved ?	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
No	8	80	42	35	60
Yes	20	210	68	40	90

The other operational factors found to be important while studying the whole data set are "Arrowboard used" or "Alternate Routes" established. However, due to the limited data points at this sub-category, the statistical significance tests cannot be performed.

<u>Location Factors</u>

The two location factors that have an impact on the clearance time are found to be roadway type and land use type. Incidents that occur on freeways last shorter than the arterial incidents, however, the difference is very small, thus ignored. Incidents that occur on rural areas last a little longer than urban incidents but the difference is again very small, thus ignored too. Only location factor significantly affecting clearance time is found to be LANDUSE3, i.e., incident on bridge or tunnel. Incidents on bridges or in tunnels take about 15 minutes more in average to clear. However, in this data set there are few incidents with this property. Therefore, this result will be included in the final knowledge base as an heuristic rule that differentiates regular incidents from incidents occurring in tunnels or on bridges.

<u>Environmental Factors</u>

The only environmental factor affecting clearance times is inclement weather condition. The incidents occurring under inclement weather condition take on average about 7 minutes longer to clear.

4.2.4.3 Personal Injury Incidents

Similar methodology is followed to develop prediction / decision trees for estimating incident clearance durations for the personal injury incidents.

Incident Severity/Detail Factors

ANOVA tests show that both the truck involvement and number of vehicles involved (Appendix-A.10 and A.11) are statistically insignificant. However, the impact of number of injuries is found to be significant for "personal injury incidents" (Appendix-A.12). Thus, the data is grouped into two sub categories: incidents with 1 to 2 injuries and incidents 3 or more injuries. Table 4.9 shows the impact of number of injuries on clearance times.

Operational Factors

Number of police vehicles on scene (POLICEVEH) is found to be significant for both the sub categories of 1-2 injuries (Appendix-A.13) and 3 or more injuries (Appendix-A.14). Table 4.10 shows the impact of number of police vehicles involved on clearance time prediction for personal injury incidents.

Table 4.9: Impact of Number of Injuries on CLT Prediction for Personal Injury Incidents

Number of Injuries	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
1-2	10	180	49	31	60
3, or 3+	20	195	71	60	80

Table 4.10: The Impact of Number of Police Vehicles Involved on CLT Prediction for Personal Injury Incidents.

No. of Injuries	POLICEVEH	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
1-2	1	10	120	42	31	60
	2	10	135	52	35	60
	3, or 3+	20	90	60	35	69
3, or 3+	1, 2	20	100	56	60	80
	3, or 3+	60	195	107	75	160

The next step is to study the other agencies' involvement on incident clearance times. The involvement of other agencies is reflected by the involvement of wreckers, ambulances and fire engines. For personal injury incidents, only the impact of wrecker and ambulance involvement is considered since fire engines are normally requested for vehicle fire incidents. The results in Appendix-A.15 show that total number of wreckers and ambulances is a significant factor affecting incident clearance times for the cases of 1-2 injury and 1 police car involved. However, for the cases of more than 1 police vehicle involved or more than 2 injury, the impact of total number of wreckers and ambulances is either statistically insignificant (Appendix-A.16) or unknown due to insufficient sample size. Table 4.11 shows the impact of total number of wreckers and ambulances on clearance times for personal injury incidents with 1-2 injury and 1 police car involved.

Table 4.11: Impact Of Total Number Of Wreckers And Ambulances on CLT For Personal Injury Incidents With 1-2 Injury And 1 Police Car Involved

WRECKER + AMBULANCE	Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
0-1	10	95	35	20	45
2, or 2+	10	120	51	35	60

Similar to the analysis of "property damage" incidents, some other operational, location, and environmental factors are found to be statistically significant for the whole personal injury data set. However, due to the limited data points at each sub-category, the statistical significance test can not be performed at this level. Some adjustments to the mean values and interval limits were made according to the statistical analysis results on the whole personal injury data set. These values are shown on the final decision trees.

4.2.4.4 Disabled Truck Incidents

Since there were only 14 cases of disabled truck incidents, carrying out statistical significance test is unnecessary. Therefore, average values, 25% and 75% percentiles are used as estimates of the clearance times. From Table 4.12, it can be seen that the variance of clearance times is quite large. However, due to lack of data, no further statistical significance analysis can be performed.

Table 4.12: Clearance Time Estimation for Road Hazard Incidents

Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
8	285	60	13	120

On the other hand, form Table 4.13, it can be seen that wrecker involvement increases clearance time significantly.

Table 4.13: Impact of Wrecker Involvement on CLT of Disabled Truck Incidents

Wrecker Involved?	Average Clearance Time (min.)
No	32
Yes	76

4.2.4.5 Vehicle Fire Incidents

Since there were only 7 cases of vehicle fire incidents, carrying out statistical significance test is not possible. Average value, 25% and 75% percentiles are used as estimates of the clearance times (Table 4.14).

Table 4.14: Clearance Time Estimation for Vehicle Fire Incidents

Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
15	75	43	38	60

4.2.4.6 Hazmat Incidents

Since there were only 3 cases of hazmat incident, carrying out statistical significance test is not possible in this case. Average of three incidents is used as the default estimate value (Table 4.15).

Table 4.15: Clearance Time Estimation for Hazmat Incidents

Min	Max	Average
135	417	244

The extremely wide range of clearance time distribution shows that more study is needed for this incident type.

4.2.4.7 Weather Related Incidents

There were only 2 cases of weather related incidents, with clearance times being 15 minutes and 150 minutes respectively. Thus, it is not possible to comment on this type of incidents.

4.2.4.8 Disabled Car in Travel Lane

There are 41 cases of Disabled Car in Travel Lane incidents, no statistically significant factor on clearance times can be identified. Average value, 25% and 75% percentiles are used as estimates of the clearance times (Table 4.16).

Table 4.16: Clearance Time Estimation for Disabled Car in Travel Lane Incidents

Extreme Low	Extreme High	Average Value	Lower Limit of the Estimate Interval	Upper Limit of the Estimate Interval
7	60	27	20	40

4.2.5 Summary of Detailed Analysis

Property damage and personal injury incidents are the majority of the incident data collected and analyzed in this study. Due to the large sample size of the data of these two incident types, detailed significance analysis are performed for each category. For road hazard, vehicle fire, disabled truck, disabled car in travel lane incidents, the sample sizes are small, however, small variations and consistent results are found for these incident types. Therefore, clearance time prediction can be made with acceptable accuracy based on the conclusions drawn from these fairly small data samples and the expert knowledge. For hazmat and weather related incidents, the number of incidents for each category is very small for drawing any meaningful conclusions. More data and study are needed to determine the factors that affect the clearance times of these incident types.

4.2.6 Development of Incident Clearance Time Prediction / Decision Trees

Based on the detailed analysis in the previous section, prediction / decision trees for incident clearance time prediction have been developed. The first split of the overall data set is made according to incident type, then all the other significant affecting factors are considered and the data is further split into smaller sub-sets. Figures 4.15, 4.16, 4.17 show the final decision trees. As mentioned before, there are some factors that have been found to be important but not included in the decision trees due to insufficient number of cases. Adjustments should be made on those factors. Table 4.17 shows those factors.

Table 4.17: Adjustments of Other Affecting Factors on Clearance Times

Affecting Factor	Arrow Board Used and/or Alternate Route Established	Inclement Weather Condition	Incidents on Bridge or Tunnel
Adjustment on CLT (min.)	+12	+10	+8

It should be pointed out that with more incident data becoming available for analysis and with the help of expert knowledge, the decision trees can be refined and further expanded.

Disablement Incidents

For disablement incidents, the data came from survey forms of two stages. Both old data and new data were used. Old data were used mainly for disablement on shoulder (no lane closure), while the new data sets are used for more severe disablement cases namely, disablements in travel lane with lane closure and truck (tractor-trailer) disablements. Figure 4.18 shows the prediction / decision tree developed for disablement incident duration prediction.

It can be seen clearly that truck disablement lasts much longer than normal car disablement. Especially, if wreckers are needed, the truck disablement can no longer be considered as a minor incident. Another interesting point that should be mentioned is the significance of lane blockage (closure) on the clearance duration of car disablement incidents. Unlike the cases of property damage, personal injury and other severe incidents in which lane closure is found to be insignificant with regard to clearance time, lane blockage plays a major role in car disablement. Car disablement occurred on shoulder deserves little attention and takes little time to clear, while car disablement in the travel lane with lane blockage takes considerably longer time to clear.

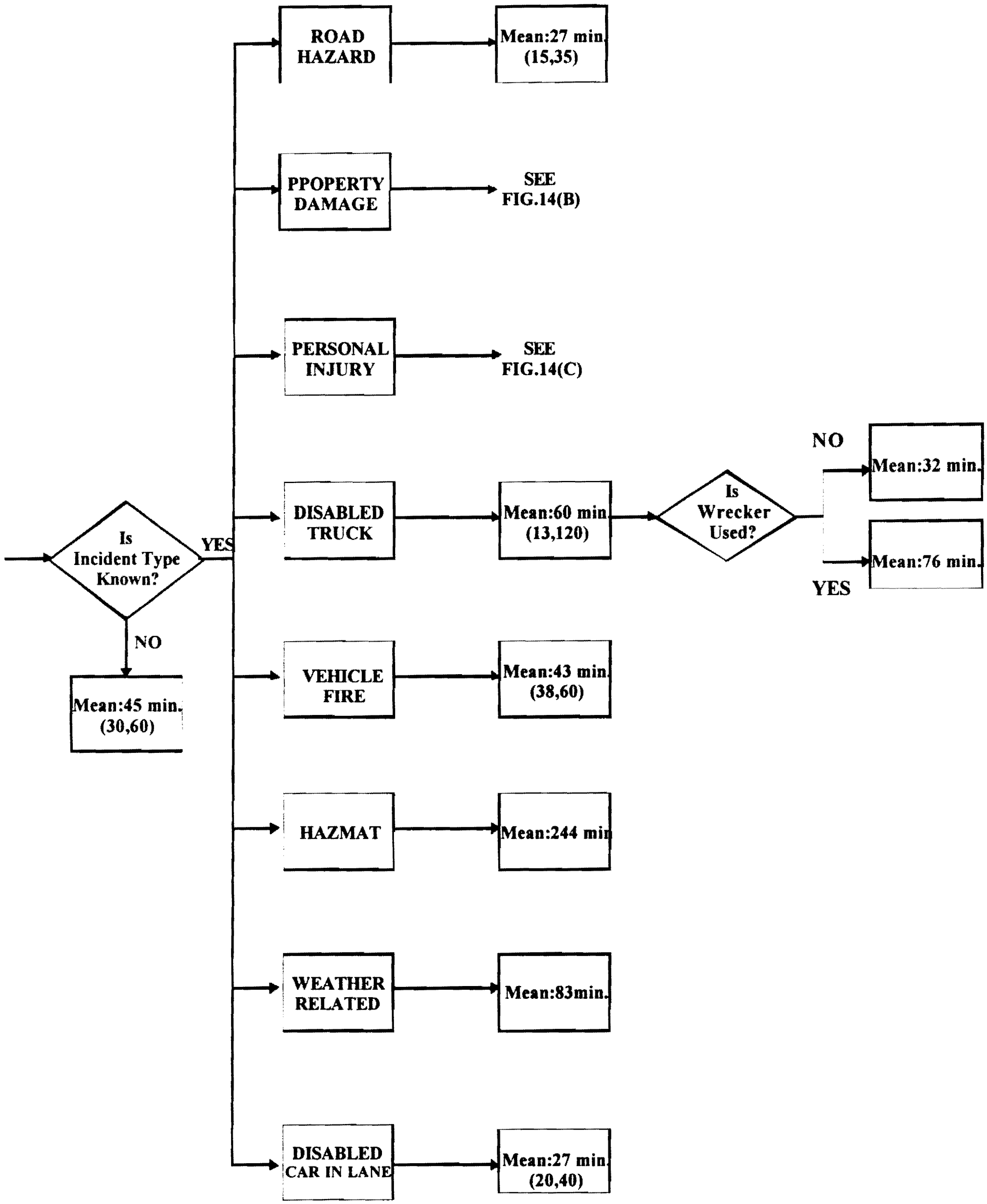


Figure 4.13: Decision Tree for Incident Clearance Time Prediction

(Numbers in parenthesis are interval limits in minutes)

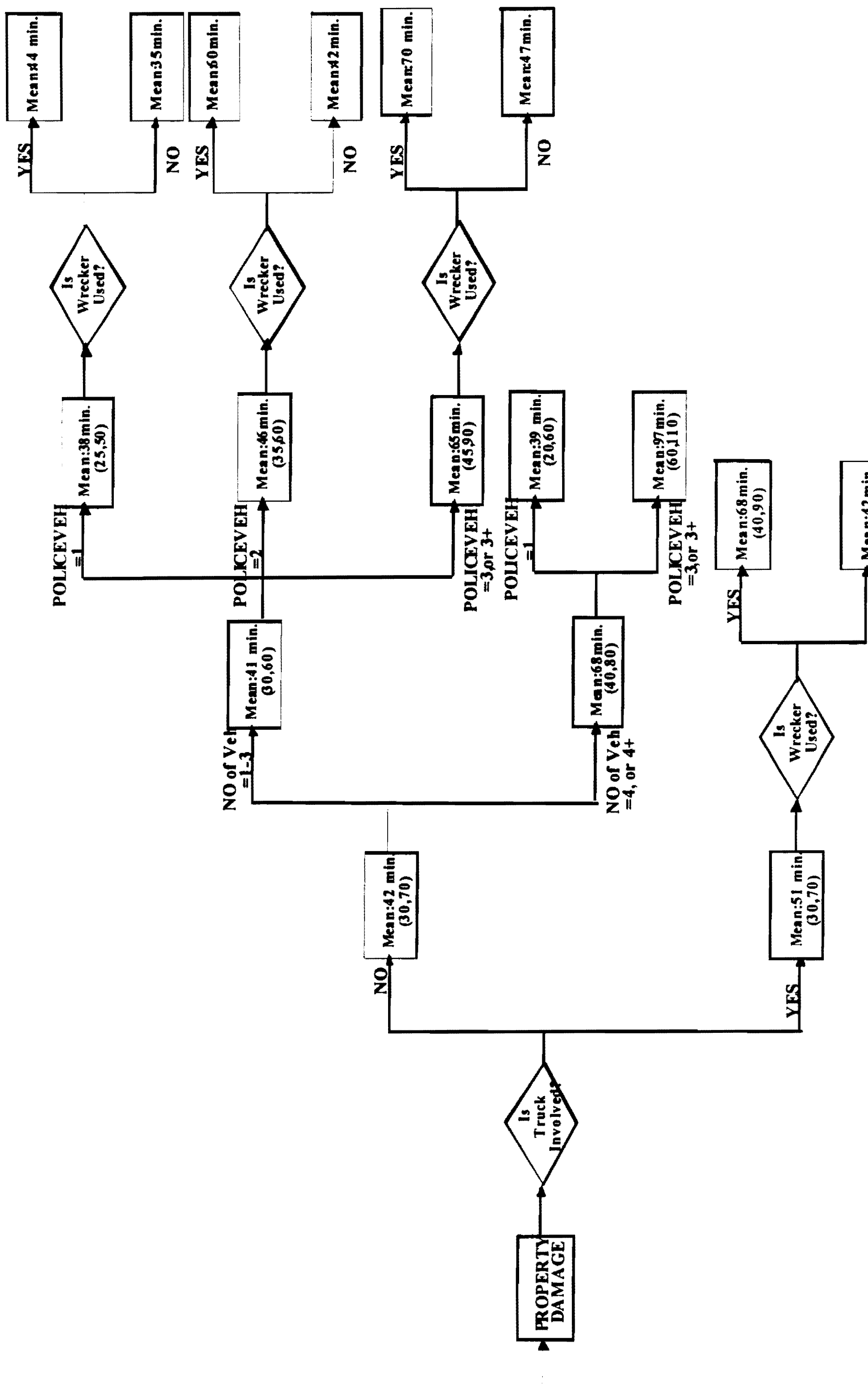


Figure 4.14: Prediction Tree for Incident CLT Prediction of Property Damage Incidents

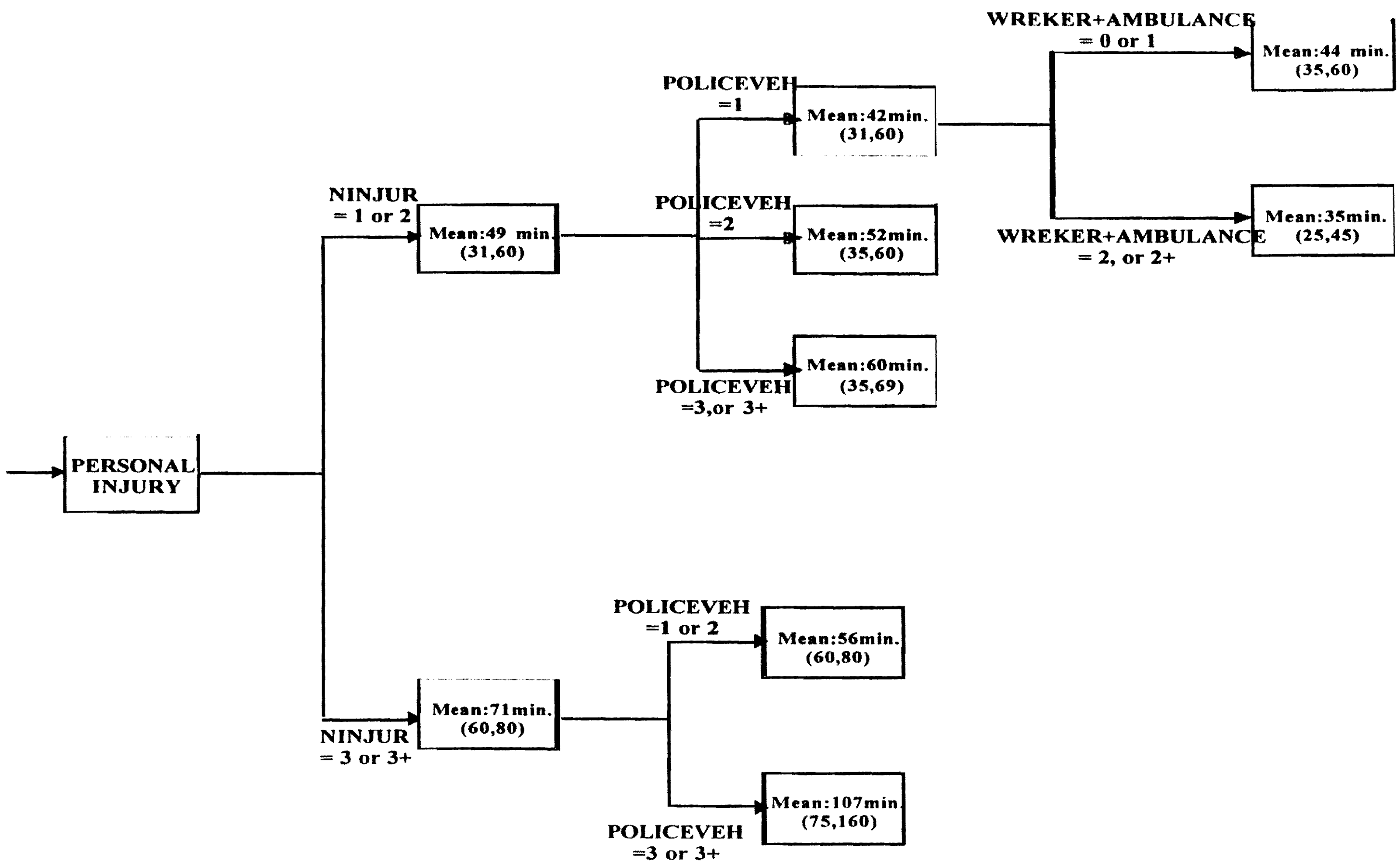


Figure 4.15: Decision Tree for Incident Clearance Time Prediction of Personal Injury Incidents

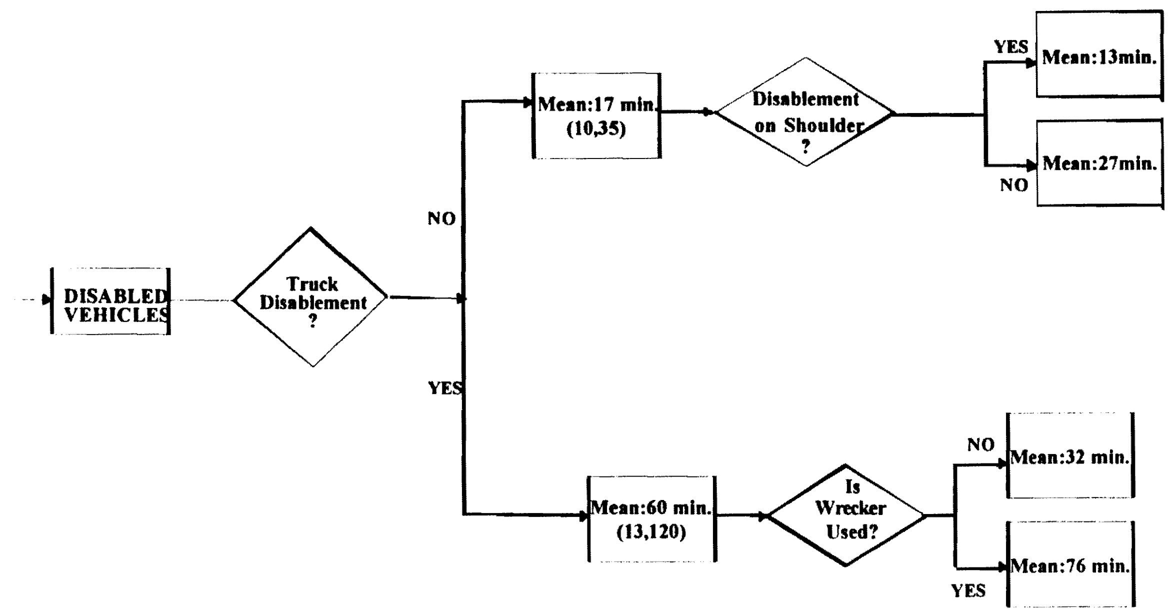


Figure 4.16: Decision Tree for Incident Clearance Time Prediction of Disablement Incidents

4.2.7 Validation of Prediction / Decision Trees

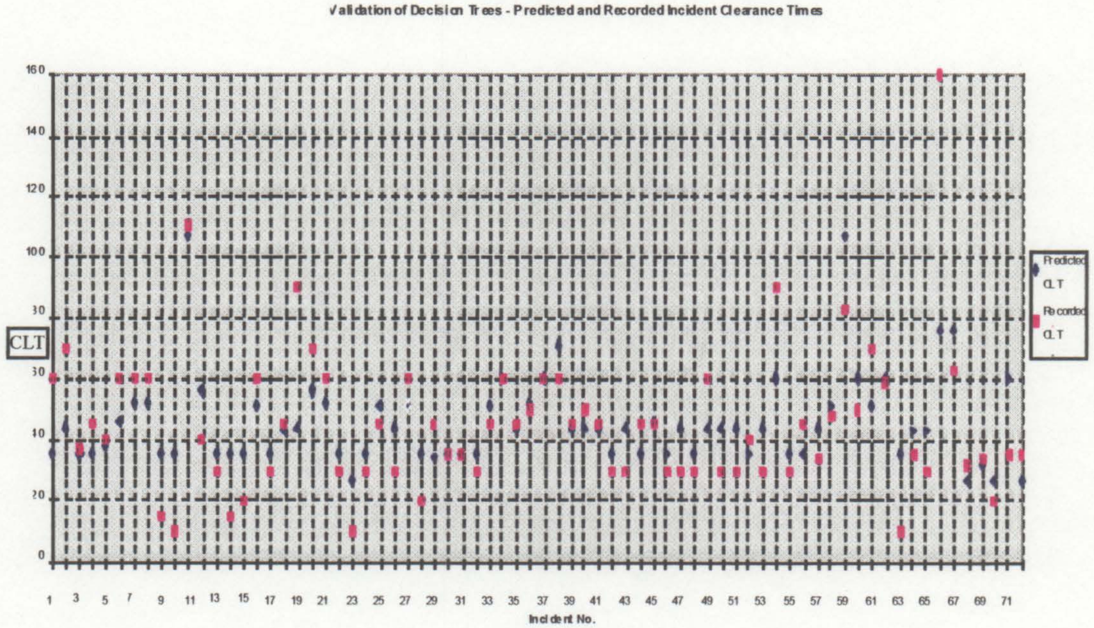
Data from the new survey forms is used to validate the decision trees. For obtaining more convincing results, only the data which is not used for developing the decision trees is used for the validation purposes. The following tables and figures show the accuracy and validity of the decision trees developed. It is shown that the developed prediction / decision trees have satisfactory precision in prediction duration of most incident cases. 44 incident cases out of 73 have been predicted with less 10 minutes of prediction error. For the incident management process, this is a very acceptable result since these predictions are only going to be used as advisories and yard sticks for making diversion decisions. However, some outliers which large difference between recorded and predicted incident duration exist. This problem is briefly discussed before in this Chapter and it is largely due to the individual differences of incident management teams in clearing similar incidents.

Table 4.18: Summary of Validation Results of the Prediction / Decision Trees

Differences	Number of Test cases
Predicted CLT-Recorded CLT<=10 min	44 Cases
15 min.< Predicted CLT-Recorded CLT<=20 min.	6 Cases
10min.< Predicted CLT-Recorded CLT<=10 min.	14 Cases
Predicted CLT-Recorded CLT>30 min.	2 Cases
20< Predicted CLT-Recorded CLT<=10 min.	7 Cases
Total No. of Cases	73 Cases

Table 4.19: Validation Results

Incident No.	Predicted CLT	Recorded CLT	Difference
1	35	60	-25
2	44	70	-26
3	35	37	-2
4	35	45	-10
5	38	40	-2
6	46	60	-14
7	52	60	-8
8	52	60	-8
9	35	15	20
10	35	10	25
11	107	110	-3
12	56	40	16
13	35	30	5
14	35	15	20
15	35	20	15
16	51	60	-9
17	35	30	5
18	43	45	-2
19	44	90	-46
20	56	70	-14
21	52	60	-8
22	35	30	5
23	27	10	17
24	35	30	5
25	51	45	6
26	44	30	14
27	51	60	-9
28	35	20	15
29	34	45	-11
30	35	35	0
31	35	35	0
32	35	30	5
33	51	45	6
34	60	60	0
35	44	45	-1
36	52	50	2
37	60	60	0
38	71	60	11
39	44	45	-1
40	44	50	-6
41	44	45	-1
42	35	30	5
43	44	30	14
44	35	45	-10
45	46	45	1
46	35	30	5
47	44	30	14
48	35	30	5
49	44	60	-16
50	44	30	14
51	44	30	14
52	35	40	-5
53	44	30	14
54	60	90	-30
55	35	30	5
56	35	45	-10
57	44	34	10
58	51	48	3
59	107	83	24
60	60	50	10
61	51	70	-19
62	60	59	1
63	35	10	25
64	43	35	8
65	43	30	13
66	76	160	-84
67	76	63	13
68	27	32	-5
69	32	34	-2
70	27	20	7
71	60	35	25
72	27	35	-8



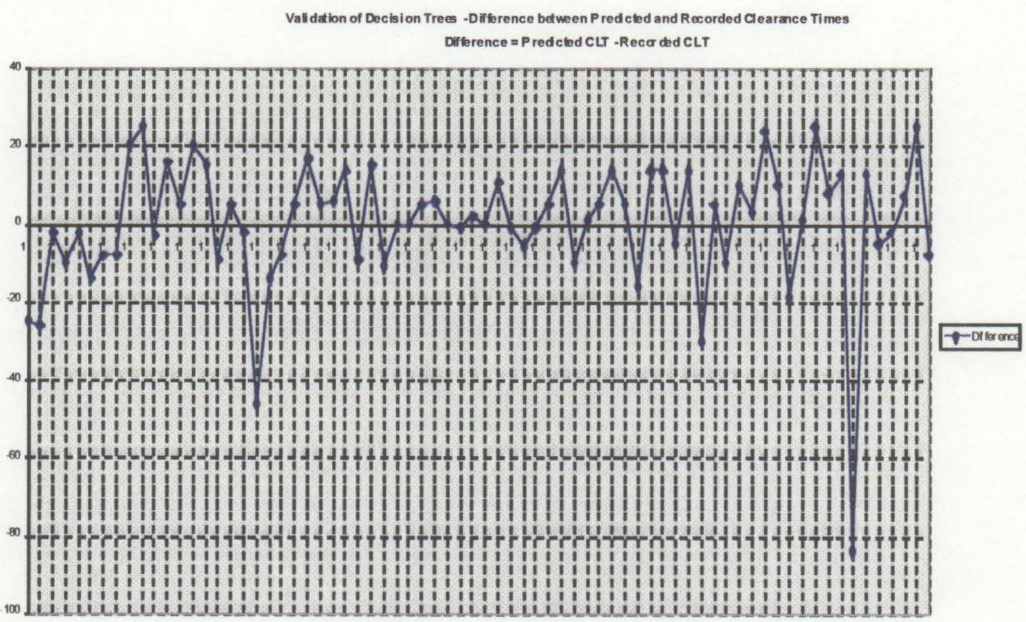


Figure 4.17: Incident Clearance Times: Recorded vs. Predicted by the Prediction / Decision Trees

4.2.8 Distribution Properties of Incident Duration Data

The study of incident duration distribution properties provides meaningful assistance in decision making during incident management process. With the validated incident duration distribution, both the unconditional probability of an incident taking the probability that an incident will last more than some time period can be assessed. This information can then be used by the incident management teams for both on-line and off-line studies. One way of using the probability that an incident will last more than some pre-determined value can be used to decide if a diversion is needed or not.

Different researchers studied the distribution properties of incident durations for different types of incidents (Golob et al., 1987, Giuliano, 1988, Jones et al., 1991). To better understand the incident clearance characteristics specific to Northern Virginia area, incident duration distribution properties are also studied in our project. Incident data from joint survey conducted by Virginia Tech, VDOT, Virginia State Police, Fairfax Fire and Rescue Department, and Fairfax county police department in Northern Virginia area is used for this study.

It is found that for a homogeneous sub-set with enough data samples, the incident clearance time generally conforms normal distribution. Normal or approximate normal distribution assumptions have been validated by chi-square tests for sub-categories of property damage incidents with similar major attribute values (e.g., No. of vehicles involved, No. of police vehicles on the scene, etc.), and other subsets of major incident types.

It is found that for the whole data set or data sets classified according to major incident types, the incident clearance time distribution curves shift to the left due to the fact that the majority of the incidents are less severe ones with short duration. These curves show some trend of log-normal distribution similar to the previous studies by other researchers, but the log-normal distribution assumptions are rejected by statistical tests. However, when the data sets are further divided into smaller sub sets based on values in such a way that each sub-set contains the incident data of the same nature and similar severity, a normal distribution trend becomes apparent and the normal distribution assumption is also confirmed by statistical tests.

It should be pointed out that chi-square tests rather than other higher power tests are used to conduct normality tests in our study. Normal distribution is by nature a continuous distribution and statistical tests such as the Anderson-Darling have generally high power for testing the normality. However, our duration data is obtained from survey forms and are not recorded very accurately. In most cases, "nice" numbers like 30 minutes are used even though the actual duration times could be 27, 32, 33.5 etc. Therefore, it is appropriate to group the incident duration data into some intervals of 15 minutes and use chi-square test which becomes a ready to use approach to conduct normality test. Table 4.20 summarizes the results of normality tests for different groups of incidents.

Table 4.20: Summary of Statistical Test Results for Incident Duration Distribution Properties

Incident Class	Normal Distribution	Log-Normal Distribution	Too Few Data to Perform Tests
Road Hazard			✓
Vehicle Fire			✓
Hazmat			✓
Weather Related			✓
Cargo Spill			✓
Truck Disablement			✓
Disabled Car in the Travel Lane	YES, [$N(27,19^2)$]	NO	✓
Property Damage, Trucks Involved	NO	Approximate	
Property Damage, Trucks Involved, Wreckers not Used	YES [$N(42,22^2)$]	NO	
Property Damage, Trucks Involved, Wreckers Used			✓
Property Damage, No Truck Involved	NO	NO	
Property Damage, No Truck Involved, 1-3 Cars	NO	NO	
Property Damage, No Truck Involved, 1-3 Cars I, 1 Police Car	YES [$N(38,22^2)$]	NO	
Property Damage, No Truck Involved, 1-3 Cars, 2 Police Cars	YES [$N(46,19^2)$]	NO	
Property Damage, No Truck Involved, 1-3 Cars, 3, 3+ Police Car	YES [$N(65,21^2)$]	NO	
Property Damage, No Truck Involved, 4 or More Cars Involved			✓
Personal Injury, the Whole Data Set	Approximate	NO	
Personal Injury, With 1 or 2 Injuries	YES [$N(49,27^2)$]	NO	
Personal Injury, With 1 or 2 Injuries, 1 Police Car on Scene	YES [$N(42,23^2)$]	NO	
Personal Injury, With 1 or 2 Injuries, 2 Police Cars on Scene	YES [$N(52,28^2)$]	NO	
Personal Injury, With 1 or 2 Injuries, 3 or More Police Cars on Scene			✓
Personal Injury, With 3 or More Injuries			✓

4.2.9 Comparison of our Results with Previous Work

Results of our study are very comparable with previous work conducted by Golob et al (1987) , Giuliano (1989), Jones et al (1991)., and Northwest ADVANCE (1994) Project in many aspects.

- Incident characteristics

  Both our study and previous work have the similar conclusions in terms of the percentages of incidents of different severity. The majority of incidents is comprised of minor incidents, typically, disablement incidents. Severe incidents with property damage and personal injuries make only a small percentage of the overall data set.

- Significant factors affecting incident duration

  Our results show that truck involvement is a major factor affecting incident duration. Generally, vehicle type is more important than number of vehicles involved, this result also conforms with the results of other studies. Among other attributes, number of vehicles involved, mainly for non-truck accidents, number of police vehicles on the scene, emergency equipment on the scene (Ambulances, fire engines, wreckers, etc.), and Hazmat involvement are found to be the significant factor affecting incident duration, which match the results of previous studies. Similar to the conclusions drawn by other studies, inclement weather, rush hour traffic, land use, roadway types are also found to have some impact on incident duration.

- Incident duration

  The incident duration is mainly determined by incident type, severity, and clearance characteristics. The different traffic patterns, demand levels, patrol frequency, response and clearance procedures vary from city to city by making the incident duration somewhat site specific. Nevertheless, the incident durations for accidents of different types in our study are still comparable to the results of previous work conducted in other parts of the Country (Table 4.21).

Table 4.21: Comparison of Incident Durations Obtained in Similar Studies

	This Study	Giuliano's (1989)	Northwest Advance Project(1994)	Goolby&Smith (1971)
Property Damage	43	44	42	45
Personal Injury	51	56	51	*
Disablement	18	*	*	18

* Not Available

- Large variation

  Large variations in incident data making the analysis more difficult and the results less consistent are observed in this study. Previous studies had similar problems with large variations in data.

The following characteristics distinguish our study from previous research.

- The main focus of our study was severer incidents, mainly accidents, represented by property damage and personal injuries, while some of the previous work extensively dealt with minor incidents like ticketing and violation stopping. Our study is important for incident management since it focuses on incidents having major impact on traffic.
- In previous studies, lane closure was found to be a major significant factor on incident duration and often used as the main characteristic for incident classification. However, our study shows that it is not a major significant factor for severe incidents. Lane closure only plays a significant role in disablement incidents in our study.

- As stated in the previous section, the incident duration distribution conforms normal distribution if the data set is homogeneous enough. Log-normal assumptions supported in some studies are rejected in our study.
- Accident frequency study conducted in other work is not within the scope of our research since our major task is the management processes taken after the incidents happen.
-

4.3 Incident Delay Prediction

The second sub-module of the diversion initiation module is the delay prediction sub-module. This sub-module adopts the deterministic queuing approach for delay prediction. Delay prediction using deterministic queuing diagram shown in Figure 4.18 is first proposed by Morales (1988). Different versions of this queuing diagram are proposed by several researchers (Khattak et al. 1994; Janson et al. 1995; Al Deek et al. 1994). In this section, basic and extended versions of this queuing diagram which incorporates the change in traffic demand and freeway capacity are both going to be used to predict the delays. The results of this extended version are also compared with the results of the basic version. An important addition to the delay prediction process is the consideration of the timing of the lane openings. This is first proposed by Mastbrook et al. (1996). The extended version of the deterministic delay prediction approach incorporates this concept of sequential lane opening as the change in supply or in freeway capacity.

Deterministic Queuing Diagram

This simple model based on deterministic queuing is an analytical procedure that, from the graphs, calculates the cumulative vehicle hours of delay. This method takes on a number of assumptions. The first of these being that the demand and capacity are assumed to be constant. The second assumption is that the demand is initially less than capacity. This is not always the case especially in urban areas where many freeways are plagued by recurring congestion.

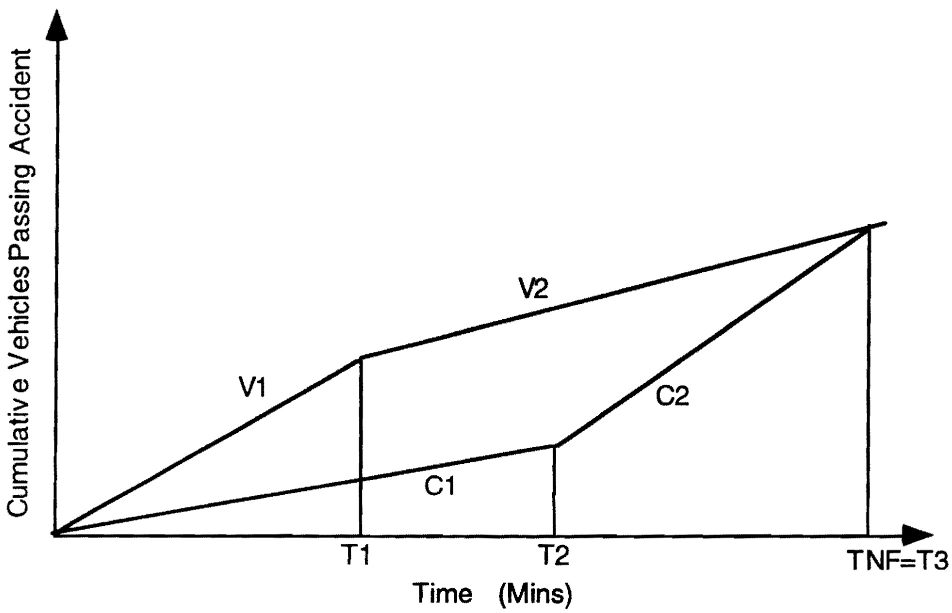


Figure 4.18: Estimation of Vehicle Delays due to Incidents

Two of the outputs from the delay estimation procedure are the TNF, the total time until normal flow is resumed, and the total delay in veh-mins which is the area between the two curves. For the basic version, TNF depends upon several factors:

For the basic version

- C_1, Reduced Capacity due to the Incident , veh/hr
- C_2, Recovered Capacity, veh/hr

- V_1, Initial Demand, veh/hr
- V_2, Adjusted Demand due to the Diversion, veh/hr
- T_1, Duration until first change in Demand, min
- T_2, Duration of Total Closure, min
- T_3, Time to Normal Flow (TNF), min

<u>For the extended version</u>

- C_1, Reduced Capacity due to the Incident , veh/hr
- C_2, Partially Recovered Capacity , veh/hr
- C_5, Fully Recovered Capacity, veh/hr
- V_1, Initial Demand, veh/hr
- V_2, Adjusted Demand due to the Diversion, veh/hr
- T_1, Duration until first change in Demand, min
- T_2, Duration of Partial Closure, min
- T_3, Duration of Total Closure, min
- T_4, Time to Normal Flow (TNF), min

The factors of the basic version are illustrated in the queuing diagram in Figure 4.18. All of these parameters are unknown at the start and during the progress of an incident. However, they can be estimated by using historical and current traffic data. The variables C_1, C_2 , C_2 , T_2 and T_3 are related to changes in the lane blockage situation, and would have to be adjusted by the system when that information about an incident becomes available. The other two variables, V_1 , V_2 , and T_1 are related to the change in demand either due to imposed or natural traffic diversions. During an incident with large delays, some of the traffic may avoid the congestion by taking other routes such as parallel arterials. This decrease in demand is difficult to measure, but should be added as a calibration factor based on historical and existing demands.

4.4 Diversion Initiation Logic

Based on the duration of an incident and the delay that is calculated using the deterministic queuing approach, pre-set thresholds are used to decide if a diversion is needed. This logic constitutes the bridge between the "diversion initiation" module and the "network generator". The Network Generator is activated if it is necessary to initiate diversion.

The diversion initiation logic recommends a diversion based on the severity of the incident. The severity of the incident is measured on three planes:

- Incident Duration
- Incident Delay
- Incident Type

While the planes are definitely interrelated, one may note that the module serves as a prediction module when only incomplete or probabilistic values of these measures are known. Hence to take advantage of the fact that information on one of the planes may be known to a greater degree of accuracy, the decision to divert is based on information on all planes.

The diversion initiation logic will recommend that diversion be initiated if any of the following conditions are true:

Incident Duration > Threshold: If the incident duration lasts longer than a fixed threshold then diversion is recommended. This threshold must be large enough to account for the time necessary to initiate the diversion while being small enough to allow motorists to take advantage of alternate routes that are available to them. The value adopted for the threshold based on consultation with Freeway Incident Management personnel in Northern Virginia is more than 1 hour.

Average Delay to User > Threshold: The average delay experienced by the motorist is an indication of the severity of the incident. The threshold accounts for the fact that for diversion to be initiated the motorist must be sufficiently assured that he will save travel time by using the alternate route. However, note that before the diversion route is established we do not have an idea of the travel time on the alternate route and thus do not have exact knowledge of travel time saved. So the diversion strategy recommendation rules serve as heuristics that recognize that for large enough travel delay on the freeway corridor the motorist is likely to save time by using an alternate route. The value adopted at present for the threshold is 30 minutes average delay.

Incident Type=Pre-defined Incident Type: Sometimes severe incidents require that diversion is necessary irrespective of the fact that the freeway may not have any physical reduction in its capacity. For example, a HAZMAT spill usually requires that diversion be initiated despite the fact that all lanes may not be blocked. Diversion is recommended by the Network Generator, for example, in case of a HAZMAT spill or vehicle fire.

4.5 Case Study

The network and the incident location described in Chapter 3 is chosen as the case study network. An incident which has the characteristics shown in Table 4.22 is used to illustrate the working of duration prediction and delay prediction sub-modules.

Incident Duration Prediction

Although, the incident duration prediction / decision trees are extensively validated in the validation section of this Chapter, one random incident is used to illustrate the idea of using prediction /decision trees for a specific case. Once the incident information is

Table 4.22: Example Incident (Source: Center for Transportation Research Old Survey Forms)

Incident Location	66 East Mile Post 65
Incident Type	Personal Injury - Property Damage
Time of Occurrence	5:35 PM
HAZMAT ?	No
# of Vehicles Involved	2
# of Cars Involved	1
# of Tractor Trailers Involved	1
# of Personal Injuries	1
# of Fatalities	0
# of Cones	6
# of Police Vehicles	2
# of Wreckers	1
# of Fire Engines	2
# of Ambulances	1
Land Use Type	Urban
Location Geometry	Straight / Level
# of Lanes Closed	1 lane and right shoulder
Incident Clearance Time (Recorded)	Between 45 mins and 60 mins
Incident Clearance Time (Predicted)	52 minutes

obtained, the prediction / duration trees can be used to predict the incident duration. Since this incident has personal injury, it can be categorized as the personal injury type of incident. Therefore, we look at the personal injury prediction / decision tree and for 1 injury and 2 police vehicles, the duration is predicted as 52 minutes.

Table 4.23: Two Approaches for Incident Delay Prediction

Basic Deterministic Queuing Approach (BDQ)	Extended Deterministic Queuing Approach (EDQ)
Given	Given
• C_1= 2000 veh/hr • C_2, =4000 veh/hr • V_1=3600 veh/hr • V_2 = 3200 veh/hr • T_1= 25 min • T_2 = 60 min	• C_1= 2000 veh/hr • C_2 =3000 veh/hr • C_5 =4000 veh/hr • V_1 = 3600 veh/hr • V_2= 3200 veh/hr • T_1= 25 min • T_2= 45 min • T_3= 60 min
Calculated	Calculated
• T_3 = TNF, min • Delay=veh-hrs	• T_4, = TNF= min • Delay= veh-hrs

Incident Delay Prediction

Based on the predicted duration, using a simple MATLAB program developed for this delay calculation purpose, the TNF for BDQ is predicted as 167 minutes and for EDQ is predicted as 169. The MATLAB graphs for both methods are also shown in Figures 4.19 and 4.20. Total delays for BDQ and EDQ are predicted as 1958 veh-hours and 1893 veh-hours, respectively. In this simple example, it is shown that the basic version tends to overestimate the delay due to the inability to incorporate the sequential opening of the lanes. This difference in prediction of delays can be larger for freeways with many lanes where the capacity change due to the sequential lane openings during the incident management process can be very substantial.

Based on the prediction of the delay and other incident characteristics, the incident initiation logic decides if a diversion is needed. In this case, the total vehicle delay is quite high and diversion is recommended by the system.

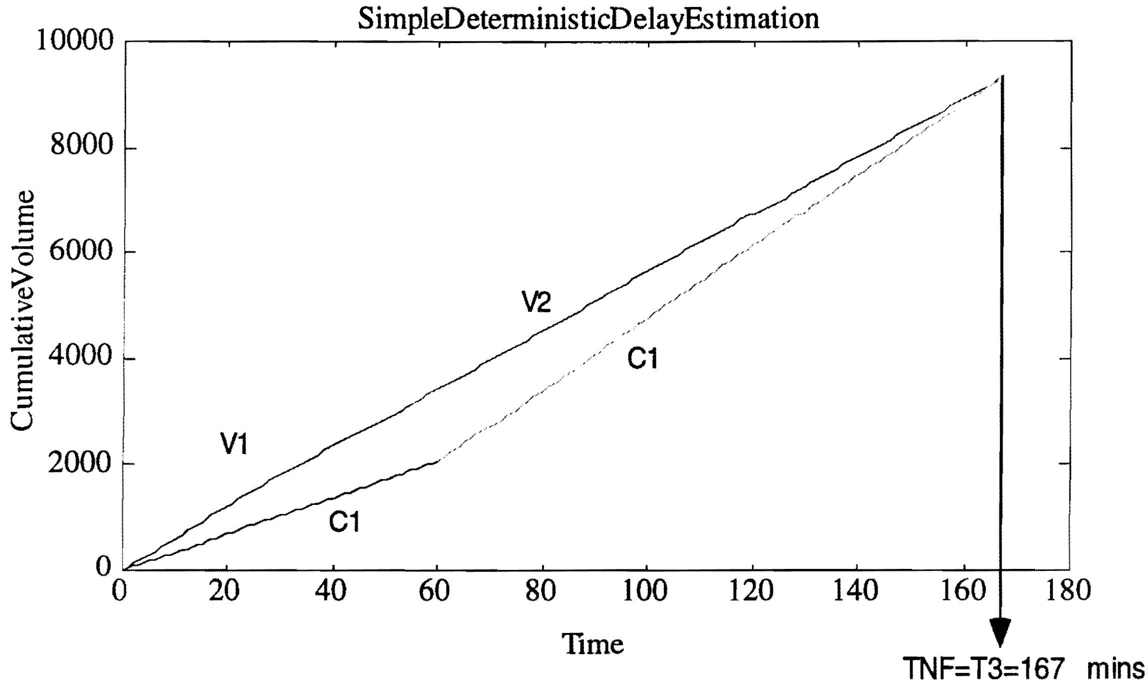


Figure 4.19: Basic Deterministic Delay Estimation

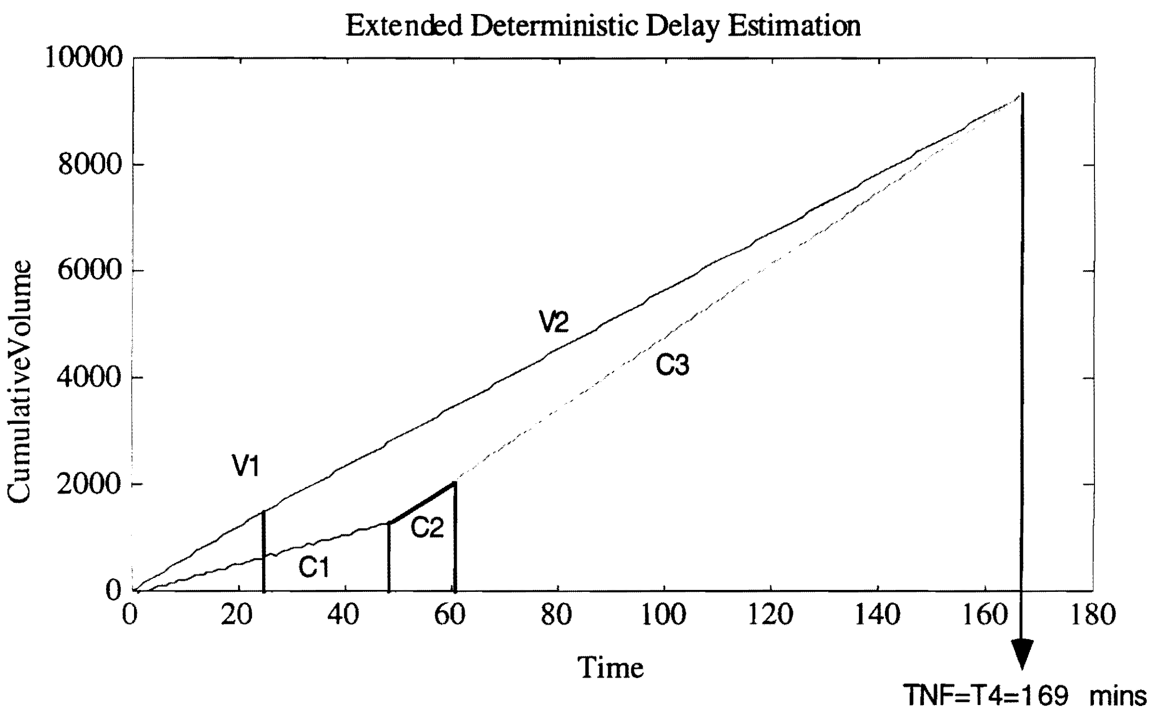


Figure 4.20: Extended Deterministic Delay Estimation

5. Heuristic Network Generator

5.1 Background

Network generator has several functions including incident impact zone determination, link elimination, route generation, and route prioritization functions. The incident impact zone determination and link elimination functions have their roots in the earlier work done by the transportation researchers to reduce the computational complexity of network flow problems. Incident impact zone determination and link elimination aim at reducing the size of the network by focusing on a smaller portion of the network and by eliminating infeasible links in the network. Thus one may broadly say that the network generator also functions as some kind of a network aggregation tool. In general, network generator acts as a flexible, real-time decision support system for supporting the overall real-time diversion process and determining alternate routes for diversion. This goal is achieved by:

1. using heuristic rules to determine incident impact zone and eliminate infeasible links,
2. employing an Expert System shell, NEXPERT-Object which allows very efficient coding of these heuristic rules,
3. using a Geographical Information System (GIS), Arc-Info, to manipulate and process large amounts of network data in real-time, otherwise impossible to handle in real-time,

4. employing state-of-the-art data and command level software bridges between Arc-Info and NEXPERT- Object. These bridges allow the real-time data and command exchange which is vital for the dynamic alternate route generation,
5. providing a unifying implementation framework for other modules developed in this dissertation,
6. supporting real-time communication of involved agencies through the client / server architecture used for its implementation.

Therefore, the software implementation of the network generator is a very important portion of the overall concept of the network generator. This is why part of this Chapter briefly describes software architecture, its implementation and components.

This section first studies the network aggregation models previously developed with the intention of better understanding earlier concepts for link. Network aggregation can be defined as the task of condensing a given network into a smaller one that can be managed efficiently as well as preserving the desired characteristics of the original network. There are two main approaches to this problem:

- 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.

Haghani et al. (1983), present an NEE model that is based on the distribution of traffic on the network after an assignment process. The links which do not carry a significant amount of traffic at the end of the assignment, 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. 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,
- The models employ a static assignment to determine the link volumes instead of a dynamic one. Thus while they can be used for planning purposes their applicability for dynamic conditions is questionable.
- Further dynamic traffic conditions such as incidents or bad weather may require factors other than merely v/c ratios as the governing criterion for network aggregation.
- 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) present a NEA methodology for the Northern Illinois network and also compare regional planning to the sketch planning approach. The characteristics used for the comparison are the total vehicle-miles, vehicle hours, and average speeds. 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 also described 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 of the two techniques. The adjusted values are presented in Table 5.1. The additional intra-zonal trips are assigned on to the same minimum time paths, in the same proportions used in the regional traffic assignment.

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.
- 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.

Table 5.1: 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%)

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 (1978), 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. They also present results of the all-or-nothing and equilibrium assignments based on the sensitivity of link load estimates to the different cases.

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.

This brief review of the network aggregation methods show that they are not suitable for real-time applications. 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 such as, new construction etc., always occurs on the original network. Saving computation time is only a secondary consideration for these studies. For diversion, one is interested in developing a set of routes for a short period (in most cases) in real-time. Therefore, instead of considering volume-capacity ratios alone, other factors are incorporated into the model as expert knowledge.

Moreover, the literature review chapter reveals that currently existing decision support merely select diversion routes after a set of routes are manually input into the system. These systems do not account for dynamic network conditions that may cause the existing set of routes to be infeasible. In such a case new diversion routes need to be generated on-line. The network generator is designed to achieve this using a hybrid system that consists of a GIS and an Expert System that identifies in real-time such routes, by extracting feasible links from the original network. An additional advantage of this hybrid system is the ability to develop diversion plans for a variety of potential incident cases, off-line. The network generator also facilitates the editing, storing and retrieval of these plans.

5.2 Theoretical Modeling of the Network Generator

In this section, we discuss in detail the theoretical modeling of the network generator. For convenience in presentation style, this section begins by defining diversion strategies. This is followed by a discussion of the modules that constitute the network generator.

5.2.1 Elements and Types of Diversion Strategies

A diversion strategy can be completely described with a set of traffic flow variables in conjunction with a set of implementation tools that disseminate the strategy. The elements of a diversion strategy are illustrated below.

Elements of a Diversion Strategy

Traffic Flow Decision Variables:

Diversion Volume(s)

Starting Point(s) for Diversion

Termination Point(s) for Diversion

Diversion Route(s)

Implementation Tools:

VMS

Radio Broadcasts

Detour Signs

In-Vehicle Advisories

Figure 5.1: Elements of a Diversion Strategy

Different values of the different traffic flow variables, lead us to identify different sets of diversion strategies. These strategies may be classified into the following sets:

- No Diversion
- Point Diversion
- Corridor Wide Diversion

- Inter-Freeway Diversion

The no diversion base case, is a strategy that may be appropriate when the incident delay is not significant, or when the alternate routes themselves are very congested. Point diversion is a strategy that diverts the traffic from the exit ramp immediately preceding the incident link, to the entry ramp immediately following the incident link. In cases when this strategy may not be the most appropriate the corridor wide diversion strategy process may be adopted. The corridor wide strategy attempts to divert motorists from a few exits upstream of the incident to a few exits downstream of the incident. In the event of a very major incident that causes several freeway links to be closed for long periods of time, an Inter-Freeway Diversion strategy to divert traffic from one freeway to another, may be appropriate. Of the diversion strategies discussed above the network generator considers all the four options listed above. The diversion strategy thus requires the specification of the diversion volume(s), origin(s) and destination(s) of the diversion and the best route(s) to carry the diversion volume. Note that, it may turn out that the best strategy is one of no diversion.

5.2.2 Incident Impact Area Estimation

5.2.2.1 Incident Impact Area Concept

One of the enhancements carried out to the network generator is the incorporation of a concept called the "incident impact area". The prototype expert system developed in Ozbay et al. (1994) searches for alternate diversion routes within the entire network, irrespective of the severity of the incident. This task is computationally expensive and is unnecessary when the incident is not very severe - a case when we can find diversion routes around the incident area. The enhancement carried out elaborates on this idea, to define an area around the incident called the incident impact area.

The incident impact area is the area around the incident which will form the search space to find alternate routes. For an incident of high severity the most beneficial diversion strategy may call for origin and termination points of diversion that are quite a distance away from the site of the incident. Thus, for such a case, the incident impact area needs to be large enough to be able to permit the flow of very high volumes of traffic from the origin point(s) to the destination point(s) of the diversion. For an incident of lesser severity it may suffice to have a relatively smaller search area, for the diversion strategy.

One can identify four factors that influence the size of the study area. These are:

- Incident Severity
- Traffic Congestion Levels
- Spatial Distribution of Network
- Motorist Behavior

The severity of the incident is a factor that influences the size of the search area. The higher the severity of an incident, the larger needs to be the size of the search area. In addition the spatial distribution of the road network will also influence the size of the search area. In a location where the road network is very sparse, an incident of a given delay may need a larger area for diversion. Also the traffic distribution in a location influences the size of the search area. For example, if an incident of some severity occurs on a freeway link near a highly congested area, then it is likely that a greater search space for diversion is needed than if it had occurred in a location around which the traffic congestion is not very high. However one may not be inclined to have a large search area, since motorists will not be willing to divert to routes that are very far from their original routes. Hence motorist behavior has to be considered in determining search areas.

5.2.2.2 Incident Impact Area Knowledge Representation

While expert systems can represent knowledge in many ways, including rule frames and semantic networks, the knowledge on incident impact areas is represented using production rules. Production rules are statements of the form “If conditions are True, Then consequences are True”. The production rules developed for the impact area module determination should ideally account for all the four factors described above. These heuristic rules use the incident severity to estimate an incident impact area. This set, also includes heuristics relating to traffic congestion levels around the incident area that suggest different impact areas for peak and non peak conditions. This first set of rules, are called “Incident Impact Area - General Rules”, account for the first two factors. There are also site specific rules to account for the next two factors. These rules are called as “Incident Impact Area - Location Specific Rules”.

First, an incident severity index for each incident is defined based on the interviews with the experts and previous studies. The severity index is defined as a function of three important factors namely, incident duration, delay and type, that are also used by the incident initiation logic. The incident severity index is summarized in Table 5.2.

Table 5.2: Incident Severity Index Rules

Incident Severity Index	Rules
Severity Index = 1	If incident_clearance_time ≤ 30 mins and If average_incident_delay ≤ 15 min/veh and If incident_type ≠Pre-specified_type
Severity Index = 2	If incident_clearance_time ≤ 60 mins and If 15 ≤ average_incident_delay ≤ 30 min/veh and If incident_type ≠ Pre-specified_type
Severity Index = 3	If incident_clearance_time ≤ 120 mins and If 30 min/veh ≤ average_incident_delay ≤ 45 min/veh and If incident_type ≠ Pre-specified_type
Severity Index = 4	If incident_clearance_time ≥ 120 mins and If average_incident_delay ≥ 45 min/veh and If incident_type = Pre-specified_ type

For the general rules governing incident impact area, the mapping of the expected delay to the major and minor axis of the incident impact zone are obtained based on the recent incident management literature expert knowledge. An example set of these rules used for the demonstration system is summarized in Table 5.3. However, the threshold values in these rules are location dependent and will vary from one location to another.

Location specific impact area rules are also based on the consultations with the expert. These rules are mainly determined based on the network characteristics and motorist behavior in the area of study. However, in this version of Network generator, these rules are not implemented in the form of rules due to the lack of expertise in determining location specific rules.

Table 5.3: Incident Impact Area - Examples of General Rules

Severity Index	Impact Area Major Axis	Impact Area Minor Axis	Diversion Strategy
1			No Div
2	≤5 miles (peak) ≤4 miles(non-peak)	≤4 miles ≤3 miles	Point Div / Corridor Div
3	≤ 8 miles (peak) ≤ 6 miles (non-peak)	≤ 6 miles ≤ 4 miles	Point Div / Corridor Div
4	≤18 miles (peak) ≤ 15 miles (non-peak)	≤ 9 miles ≤ 6 miles	Corridor Div / Inter-Freeway Div.

5.2.2.3 Determining the Starting and Terminating Points of the Diversion

The network generator identifies the starting and termination points for the diversion based on the expected duration, delay and associated backup due to the incident. The following heuristics are adopted to determine the starting and termination points of the diversion.

- If the incident severity index is equal to 3 or 4 then employ a freeway-freeway diversion strategy.
- If incident is near exit ramp of directed link then divert at least two exits prior to the incident link.

Assuming only entrance exit combination will be employed for the diversion find the route within the search area that is closest to minimizing the sum of arterial and freeway delays.

5.2.2.4 Diversion Volume Estimation

An important element of the diversion strategy that needs to be determined is the volume that is to be diverted. Although this volume is not used for real-time diversion of traffic, it is needed to roughly predict the v/c ratios on the network links after the diversion is initiated. This volume was obtained in the prototype system as a pre-defined fraction of the existing freeway link volume (Ozbay et al., 1994). This approach is simplistic and is refined in this enhanced version of the network generator.

Gupta et al. (1992) approach the problem of estimating the diversion volume by regarding any volume that produces a v/c ratio greater than a pre-specified critical threshold as the potential volume for diversion. Then a decision is made to check if the excess volume when diverted reaches the termination location of the diversion, faster than if they would have not diverted. If so, then the diversion is implemented. Gupta at al. (1992) check whether diversion is beneficial for three such v/c thresholds.

The approach we intend to follow is analogous to the approach of Gupta et al., (1992) but has several important advantages. To illustrate this consider a major incident that occurs making the v/c ratio to be very high (could be much greater than 1). In their system, the option of reducing the v/c ratio to one of the three specified thresholds (all less than 1) would entail the diversion of either very large amounts of link volume or none at all. The approach does not allow for consideration of a larger set of possible diversion volumes from which one may select the most beneficial one. One way to consider a larger set of diversion volumes would be to consider a large set of v/c ratios (both less than and greater than 1). Instead the network generator accomplishes this through a more intuitive approach where the diversion volume is first set to an on-line generated lower bound and is then continually marginally increased until the best diversion strategy is identified. The best strategy is identified in terms of travel time increases due to the extra volume that is going to be diverted.

The diversion volumes have to be bounded between zero and the link volume on the freeway. However, to reduce the search space for the diversion volume we set certain upper and lower bounds for the diversion volume. The value for these bounds will be fixed based on a variety of factors. The factors that need to be considered for the lower bound on the diversion volume are the incident duration and the available capacity. Instead the network generator uses a measure that takes both incident severity and available capacity into account. The upper bound for the diversion volume is also fixed based on the severity of the incident. Also, note that when the arterials are very crowded, or if the incident is not very severe, the upper bound on the diversion volume can be set to be much less than the entire link volume. Thus the network generator employs an iterative step wise search between on-line generated lower bounds and upper bounds for the diversion volume. The lower and upper bounds are expressed as fractions of the freeway volume at the link corresponding to the exit ramp where diversion route begins.

If the most beneficial diversion strategy is found to be at the lower bound then we may decrease the lower bound and continue searching in the direction of the new lower bound. Similarly if the best diversion strategy is found to be at the upper bound then we increase the bound and continue searching in the direction of the new upper bound. Note that this is a quick heuristic procedure to roughly estimate v/c ratios. It is not going to be used for real-time diversion.

5.2.3 Dynamic Link Elimination

5.2.3.1 Dynamic Link Elimination Concept

An urban arterial network typically comprises of a large number of links with different traffic and geometric characteristics. Any route determination model, for diversion or for dynamic traffic assignment, needs to consider all these links to be potential components of the route that is to be determined. The route generation procedures of the diversion or route guidance models often becomes computationally very expensive, because they have to consider all the links within a given search area. Often these route generation models end up considering quite a large number of links that are not a part of the routes they generate. Thus, if one were to have a module that uses some approach to identify and eliminate a set of links that are not likely to be a part of these routes then the computational efficiency of these route generation / route guidance models would increase considerably. We call such a module a "Link Elimination Module", and the concept is called "Link Elimination".

The real-time factors can be traffic conditions, special events, construction and maintenance activities that cause a link not to be a part of a route may vary with time. Thus, the link elimination process has to be dynamic.

In addition to the network aggregation models described in the background section of this Chapter, there has been some research conducted so far that may be applicable to link elimination. The notion that factors other than the v/c ratio have to be considered is expressed in Ketselidou et al. (1989) for diversion route generation. This model also recognizes that the v/c ratios on the alternate routes may vary with time. Although this approach does not directly employ a link elimination module it uses an alternate route selection strategy that uses several criterion other than the v/c ratio to determine the diversion route. This model, in effect, eliminates a route if either a particular link in the route violates some criteria or if the characteristic of the route as a whole (e.g. route length) violates some pre-determined criteria.

The approach proposed in this research is an enhancement over the network aggregation models because it considers dynamic factors for link elimination. In addition, we recognize that the arterial v/c conditions can vary with time and also recognize that the need to consider factors other than the v/c condition in link elimination.

As compared to the approach proposed by Ketselidou (1989), the approach proposed in this Chapter has the following characteristics:

- this approach is much more direct. We first eliminate links and then build routes. In comparison, the approach proposed by Ketselidou et al. (1989) uses pre-defined routes and then eliminates the route if a link violates some condition.

- As explained in the following sections, a larger and more comprehensive set of factors to decide whether a link can be a potential element of a route are used in this research.
- Instead of choosing from a limited set of pre-defined alternate routes, *new diversion routes* are chosen after eliminating infeasible links.

5.2.3.2 Proposed Approach for Link Elimination

The approach adopted for link elimination considers a large set of static and dynamic factors that will decide on the feasibility of a link for diversion/dynamic route guidance. In comparison to the network aggregation models, the analytic processing of the decision factors will be done heuristically and not mathematically. This will make the processing much more computationally efficient, making it much more suitable for real-time implementation.

The heuristics developed examine each link with respect to a set of decision factors. Then, a decision is made on the feasibility of the link. The heuristics are represented as rules in an expert system. The set of production rules employed for the network generator constitute a deterministic set of rules. The antecedents and conclusions of these rules are well defined without being fuzzy. The various decision factors considered, form the antecedents to these rules. The decision factors considered, the heuristics involved, and the heuristic processing mechanisms are described in detail in the sections that follow.

Before the current implementation of the network generator, an earlier version that employed classical Boolean logic production rules for link elimination existed (Ozbay et al., 1994). The major obstacle with the representation of knowledge base using only predicate calculus and Boolean logic is that when the knowledge involves uncertain events (e.g. unknown volumes), the knowledge is seldom of a binary nature. In addition, the early prototype expert system did not recognize that different set of antecedent may lead to different confidences in the conclusions reached. For example, a conclusion to eliminate the link because of a very severe HAZMAT spill has a different confidence level as compared to a conclusion to eliminate the link because of the link being in a bad neighborhood. Simple binary logic systems thus are prone to making false positive and false negative predictions, because of a lack of probabilities being associated with the derived conclusions.

In order to overcome the difficulty of incorporating probabilistic knowledge in binary logic, the enhanced link elimination module uses weights for the degree of belief in the rule. This approach was used with success in the medical diagnostic system MYCIN (Heckerman, 1986), where the production rules are weighted by a certainty factor. Positive certainty factors are a measure of relative support for a hypotheses and negative factors a measure against the hypotheses. Heckerman and Horvitz (1987) show that such an approach is analogous to building a probabilistic expert system except in that it assumes the conditional independence of rules.

5.2.3.3 Factors Influencing Link Elimination

In order to better understand a decision process that will judge whether a link is feasible for diversion or not, we may consider the decision to be an outcome of the interaction of external forces and link attributes. One may view any factor that may cause a link to be eliminated, as an external force. Link attributes are characteristics of the link that deal with these external forces. The decision process then depends on our prediction of the response of the link, which depends on its characteristics, to these external forces. If we believe that the link is capable of handling the external forces well, then we don't eliminate the link, else we eliminate the link. Table 5.4 lists the external factors and link attributes that the network generator will consider in deciding whether a link will be eliminated or not.

A wide array of decision factors can result from a combination of different external forces that can vary in their extent/magnitude and link attributes. The different factors that are considered by the network generator are listed as the antecedents of the heuristics rules developed for "Dynamic Link Elimination". At this junction however, a discussion on each external force and the magnitudes they can assume, followed by a discussion on the link attributes and the values they can take is appropriate. The justification for the choices of external forces and link attributes is sometimes illustrated with an example.

Table 5.4: External Forces on a Link

Traffic Volumes
Weather
Jurisdictional Issues
Special Events
Construction/Maintenance
Bad Neighborhoods
Office/School Times
Incidents on nearby arterials
Adjacency to Hospitals / Metros
Incident Delay

Incident Type

Table 5.5: Link Attributes

Capacity
Geometric Design
Ice/Snow Clearance
Speed Limit
Link Location
Link Type

5.2.3.3.1 External Forces

Incident Severity:

Incident severity affects many of our decisions on using a link. If severity is very high then we may compromise on many other factors in deciding the feasibility of a link. As an illustration, if the incident delay is very high then we may not wish to avoid a link despite its speed limit being low.

Traffic Volumes:

This volume represents the net volume on the link. This includes the volume that normally flows on these links and the additional volume that is added by the diversion. Traffic volumes can be thought of as external forces because they are an input to determining the congestion level of a link. If the congestion level of the link is high then we may need to eliminate this link.

Maintenance

In the event of maintenance activity on some arterial links, the link may be deemed unfit for diversion. In the modeling of the network generator, the operator has to input the links that are currently undergoing maintenance.

Weather Conditions

Weather forces clearly are external forces. These forces can prevent the usage of the link for diversion, even when the link is not congested. While quantifying this external force completely can be a difficult task the heuristics for the network generator will work with descriptive measures for the weather. For example, very icy is a descriptive measure of the weather. Table 5.6 shows the different classifications of weather conditions that the network generator deals with.

Table 5.6: Classification of Weather Conditions

Snow	"Light Snow" "Moderate Snow" "Heavy Snow"
Ice	"Icy" "Very Icy"
Rain	"Light Rain: "Heavy Rain"
Fog	"Heavy Fog"

Jurisdictional Factors

Sometimes jurisdictional factors may not permit some of the links in the arterial network to be used for diversion. The links which have to be eliminated for diversion depend on the location of the incident

Special events

Special events like football games etc., can generate additional traffic on the arterial network. The volumes generated may be so high that it may not permit additional diverted traffic.

School/Office Hours

This external force is similar to special events except in that these have to be dealt with every day.

Arterial Incidents

Incidents on arterial streets are an external force that may cause a link to be eliminated for diversion of traffic from freeways.

Hospitals/Airports/Metros

Often in urban cities, it may not be advisable to divert traffic onto streets that are adjacent to hospitals. This is because the traffic disruptions brought about by the diverted volumes may prevent easy access to these locations. Hence hospitals/airports/metros constitute an external force

5.2.3.4 Link Characteristics

Link Capacity

Link capacity is a very important characteristics that the diversion model needs to consider. The effect of loading a given diversion volume can be determined only with a knowledge of the link capacity.

Geometric Characteristics

Geometric characteristics of major roads is particularly important under bad weather. To illustrate this point consider the example of a road with very steep gradient. It may not be appropriate to divert traffic onto such a road when the weather is icy. Also, some roads can be prone to incidents under heavy congestion due to their geometric characteristics. The values this property may take are listed in the table 5.7.

Table 5.7: Geometric Characteristics

Gradient	Steep
	Level
Curvature	Windy

	Straight

Ice/Snow Clearance

This clearance property of a link is needed to determine its feasibility when the weather is icy/snowy. If the link is periodically cleared then we may not wish to eliminate this link. If the link is not cleared then we may wish to eliminate this link. The values this property may take are listed in Table 5.8.

Table 5.8: Ice/Snow Clearance Measures

Adequately Cleared
Moderately Cleared
Not Cleared

Speed Limit

A low speed limit may cause a link to be eliminated, because drivers may not be willing to divert onto such a link.

Location

The location of a link with respect to bad neighborhoods, office/school zones, special events/ airports influences its elimination.

Link Type

Link type has an influence in our decisions. The conditions under which we eliminate an arterial link may be different from the conditions under which we eliminate a freeway link or entrance/exit ramp.

Table 5.9: Types of Links

Arterial Street
Freeway Link

Entrance Ramp
Exit Ramp
Bridge / Tunnel

5.2.3.5 Rule Base for Dynamic Link Elimination

A rule for dynamic link elimination connects a decision factor with a decision or with a result that could lead to a decision. Before the decision factors are described, the decisions each rule may lead to are described. The table below shows the possible results a rule may lead to. Following the table is our justification for using the decision states described by the table below.

Table 5.10: Decision States

Definitely Eliminate (DEL)
Eliminate (EL)
Don't Eliminate (NE)
Definitely Don't Eliminate (DNE)

5.2.3.6 Decision States

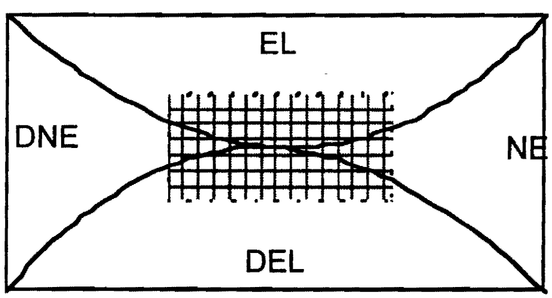
For some cases, the need for eliminating a link which is being can be very strong. In such a case, the link may be considered as to be "definitely eliminated". Such a case may be when safety reasons force the model to definitely eliminate the link. Some times, there may not be enough reason to "definitely eliminate" the link in which case, "eliminate the link" means that the link is rather not used. Links which are eliminated and not definitely eliminated may have to be used in the event that no diversion route is found. Similarly, the "Don't Eliminate" and "Definitely Don't Eliminate" rules represent different levels of decision states. A parallelism may be drawn here between the production rules of the Network Generator and those of MYCIN.

MYCIN's rules serve to diagnose a patients illness. The diagnoses could be any of a set of possible ailments of varying severity. The network generator's link elimination rules could be similarly thought of as diagnostic rules that ascertain whether a particular link is "healthy" for use in diversion. The network generator's diagnoses for a particular link could be a malfunction that cripples the links functioning in which the link is definitely eliminated due to icy conditions on a bridge. Or, it could be a diagnosis that calls for temporary suspension from activity until absolutely necessary, in which case the link is eliminated due to jurisdictional problems. The links functioning may be normal in which case the link is not eliminated, or the link could be diagnosed to have some salient features (a very low v/c) that prompts us to definitely use it.

At this point, a difficulty that may arise in developing a rule processing scheme that will utilize the above decision states must be pointed out. Such a difficulty may arise because two different rules may be applicable and may yield different decision states. Then, we need to resolve the conflict between the decisions a set of rules may yield. Extending the MYCIN analogy further this is similar to a case where a patient has symptoms characteristic of several ailments. In comparison, the prototype expert system had link elimination rules with just one decision state - "Eliminate". If any one rule's antecedent matched the given condition then the link would be eliminated. While this approach does not have any conflict resolution problems, it leads to some other serious drawbacks. Tests revealed that too many links were often being eliminated because the external force and link attribute combination often match at least one rules antecedent (Krishnaswamy V., 1994). In this enhanced model, we aim for more realistic outputs and thus adopt a multi valued decision state for our rules, and also employ a conflict resolution technique.

5.2.3.7 Link Elimination Decision Making

The decision making task may be represented by the Venn diagram below. The diagram shows four mutually exclusive sets representing the four decision states discussed in the previous section. Each set has a set of characteristics that uniquely identify it with that set. For example a v/c ratio of less than 0.3 belongs only to the set "Definitely do Not Eliminate (DNE)". The incidence (external forces + link characteristics) can be represented by the set I. Note that I can share common areas with the other four sets. For example, the incidence could be a low v/c ratio and a HAZMAT spill, in which case the incidence shares common areas with the "Definitely Eliminate (DE)" and "Definitely do Not Eliminate (DNE)" sets.



Legend:
DNE: Definitely do Not Eliminate
NE: Dont Eliminate
EL: ELiminate
DEL: Definitely ELiminate
▬ Incidence

Figure 5.2: Venn Diagram for Decision States

The decision making task for the link elimination module is to map a given incidence set to the most appropriate one of the four decision states. The link elimination decisions are arrived at using a production rule based approach. In order to fully comprehend the decision making process it is important to discuss the link elimination rule structures.

5.2.3.8 Link Elimination Rule Structure

The production rules for the "Link Elimination:" module, as discussed earlier, incorporate probabilistic knowledge in binary logic by weighing each rule with a confidence factor. These rules may be represented by

$$E \Longrightarrow H$$

where H is the hypotheses and E is the evidence relating to the hypotheses. For example a hypotheses might be that safety reasons demand a link to be eliminated and the evidence could be a severe HAZMAT spill around the area.

The enhanced prototype recognizes that relationship between evidences and hypotheses are often uncertain in the extent to which we can depend on the hypotheses. For example a hypotheses to eliminate the link based on safety reasons is better supported by evidence of a HAZMAT spill as compared to the link being in a bad neighborhood. In order to accommodate the degree of beliefs in a hypotheses, the network generator uses certainty factors to weigh each rule. The certainty factor represents change in belief about a hypotheses given some evidence. This may be represented by

$$CF(H,E).\ E \Longrightarrow H$$

Certainty factors range between -1 and 1. If the certainty factor for a particular rule is positive then it means that the evidence increases the belief in the hypotheses of that rule. On the other hand if it is negative it implies that the evidence decreases the belief in the hypotheses of that rule. For example a high volume by capacity ratio increases our belief in the hypotheses that the link need be eliminated, while a low v/c ratio decreases our belief. These two rules may be represented as in Figure 5.3. The figure indicates that the hypotheses to eliminate a link is strengthened by 0.7 if the evidence that v/c > 0.8 is true. On the other hand the belief in this hypotheses is reduced by 0.9 if there is evidence that v/c < 0.3.

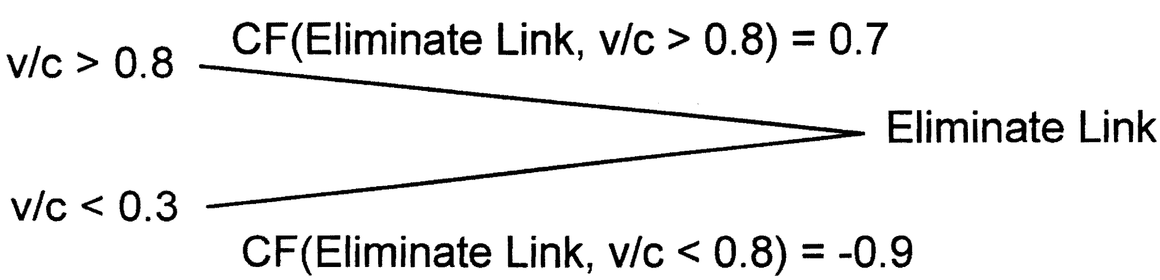


Figure 5.3: Representation of Two V/C Rules

\

5.2.3.9 Link Elimination Decision Process

The decision to eliminate a link or not, is dictated by the decision state reached through the processing of knowledge represented as rules. The rules being adopted follow the structure discussed in the previous section. The present discussion focuses on how the structure of the rules is exploited in order to arrive at one of the four decision states discussed in section 5.2.3.6.

First, it is important to note that there are some conditions in which a single rule, if applicable, governs the decision state. These rules primarily focus on safety issues (very icy conditions) and obvious feasibility issues (very high congestion during peak periods) during diversion. If any of these are applicable then we definitely eliminate the link. If not we continue with our processing to inspect the link, with reference to other decision factors. This concept is illustrated with the following figure:

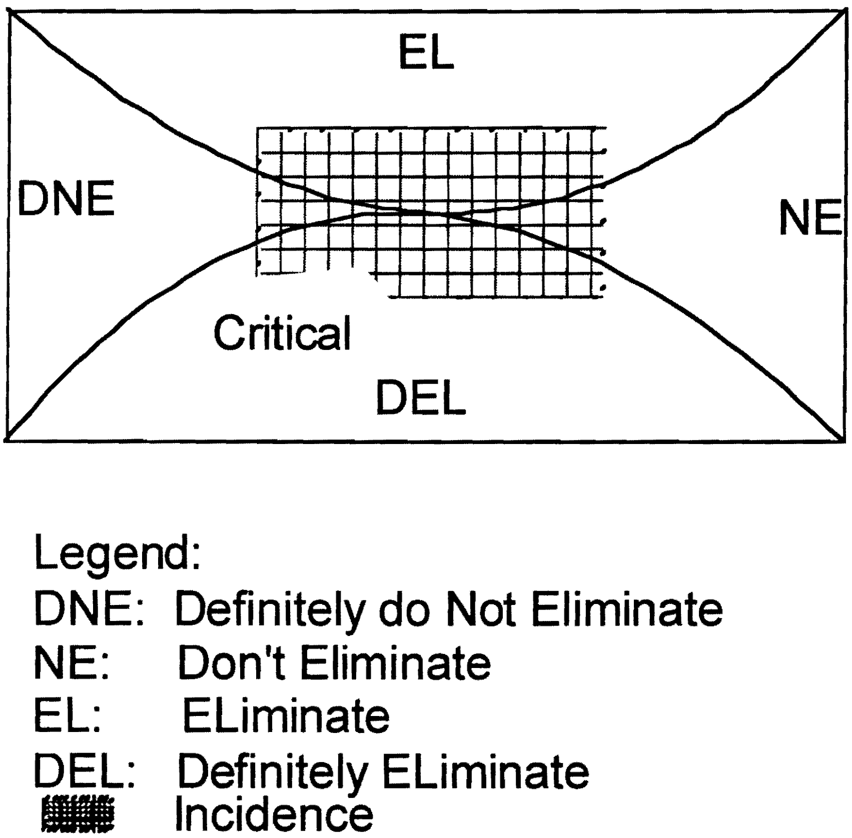


Figure 5.3: Concept of Critical Region

The critical region is a subset of the DEL region. If the incidence presented has any components that have applicable antecedents in the Critical region then the link must be eliminated. The rules that apply for the critical elimination are depicted in the network shown below.

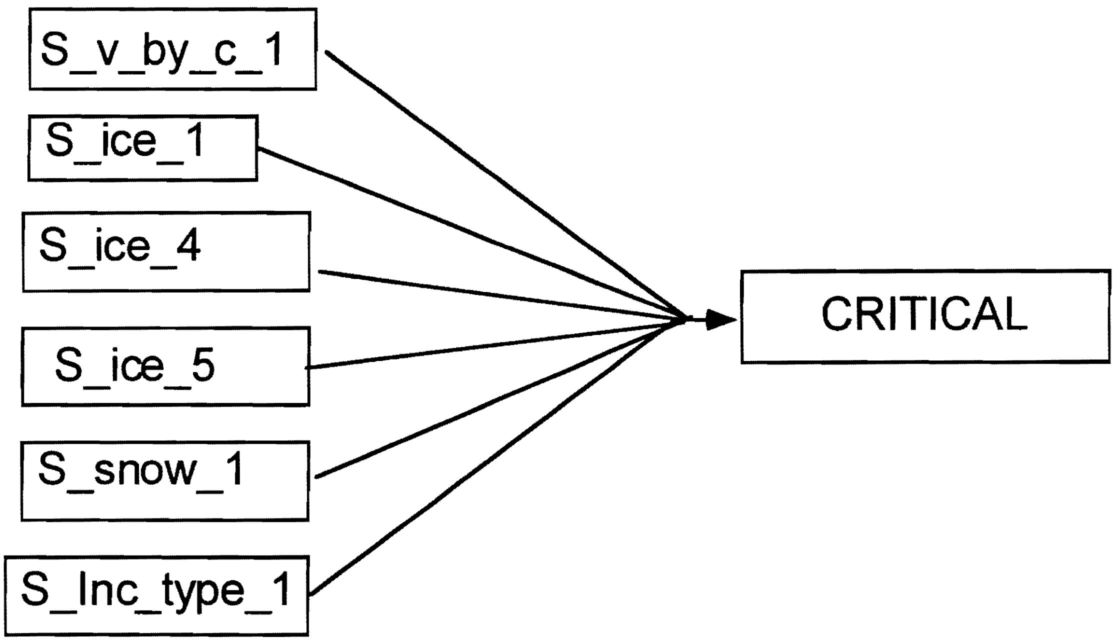


Figure 5.4: Critical Elimination Rule Network

Figure 5.5 shows the rule network for critical elimination. If any of the rules shown and presented in Table 5.10 are applicable then the hypothesis of CRITICAL conditions holds true then the link is definitely eliminated. Once a particular hypotheses is true then we stop processing. This is achieved in NEXPERT using the strategy operator to set exhaustive evaluation to false, locally.

If the incidence does not have a mapping onto the critical region, the rest of the decision making process described below must be carried out. The philosophy of the decision process rests on the observation that the four decision states may be represented as different measures of belief against an assumed hypotheses. To illustrate this further let's choose an assumed hypotheses that the link is to be eliminated. Then, it is possible to write all the rules in such a way that their confidence values are measures of belief against the hypotheses that the link is to be eliminated. The philosophy of the approach is to use the cumulative belief for the incidence, against the hypotheses that the link is to be eliminated, to map the it to the four decision states. This process may be represented by the figure below.

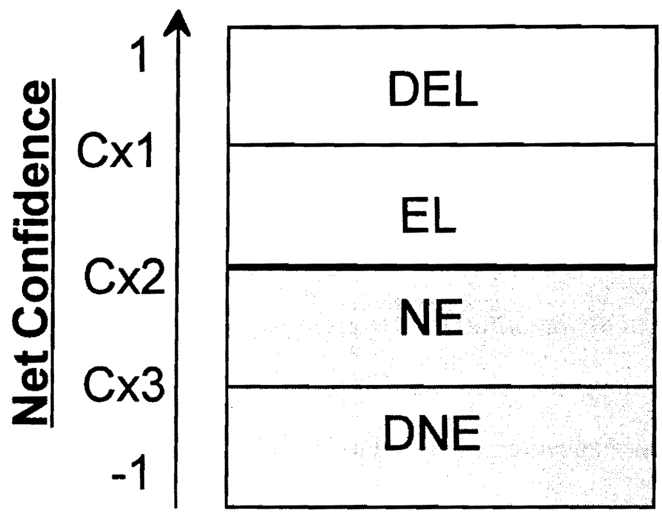
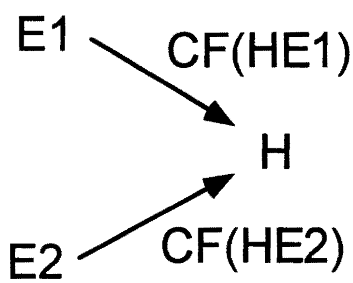


Figure 5.5: Net Confidence

If the net confidence against the hypotheses that the link is to be eliminated varies from Cx1 to 1 then there is good reason to definitely eliminate the link. If it varies from Cx2 to Cx1 then we may merely eliminate the link. If the net confidence turns out to be negative then we either do not eliminate the link or definitely don't eliminate the link. Thus if the net confidence varies from Cx2 to Cx3 we don't eliminate the link and if it is lesser than Cx3 we definitely don't eliminate. Extensive tests have to be done before the thresholds Cx1, Cx2 and Cx3 are determined. Note however that a value of 0 for Cx2 seems a very logical choice.

5.2.3.10 Cumulative Weight Function for Conflict Resolution

The cumulative weight function for conflict resolution used for link elimination is again similar to one used in MYCIN. The cumulative measure of belief against a hypothesis is computed by the following function given by a parallel combination of evidence shown in Figure 5.7.



CF(H,E1,E2) ={ CF(H,E1) + CF(H,E2) }/ (1 - min(CF(H,E1), CF(H,E2)) if CF(H,E1)*CF(H,E2) < 0

CF(H,E1,E2) = CF(H,E1) + CF(H,E2) } + (CF(H,E1)* CF(H,E2)) if CF(H,E1),CF(H,E2) < 0

CF(H,E1,E2) = CF(H,E1) + CF(H,E2) } - (CF(H,E1)* CF(H,E2)) if CF(H,E1),CF(H,E2) > 0

Figure 5.6 : A Parallel Rule Network (Heckerman, 1986)

It is very important to note that the combination function presented above meet the desired condition of not having the cumulative weight being influenced by the order in which the evidence is presented, for parallel combinations.

For sequential combinations of evidence shown in Figure below, that arise in a rule network the combination function is given as

CF(H,E') = CF(H,E)*max(0,(CF(E,E')))

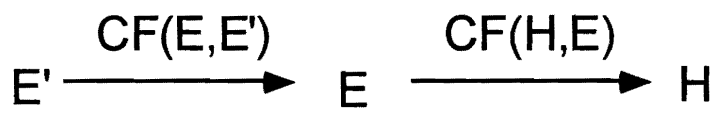


Figure 5.7: Sequential Combination of Evidence

The rule network for decision making is depicted in the Figure 5.9. The decision state chosen finally depends upon the cumulative weight. In order to compute the cumulative weight all possible rule antecedents have to be evaluated. Thus, each rule that is applicable in the left most column is evaluated and the corresponding influence on the cumulative weight is then computed. Once all the rules are evaluated the final decision will be based on the net measure of belief against the hypothesis of not eliminating the link.

Note that the above rule processing procedure seemingly requires that all rules be evaluated. However, such a case may be avoided by evaluating the mutually exclusive antecedents separately. To illustrate this idea, let us consider a case where the weather is icy. Then, we immediately know that snow rules need not be applied. Another illustrative case is the v/c ratio test. Instead of checking the v/c ratio against each of the thresholds, we may instead compute it and immediately set the rules that are irrelevant to false.

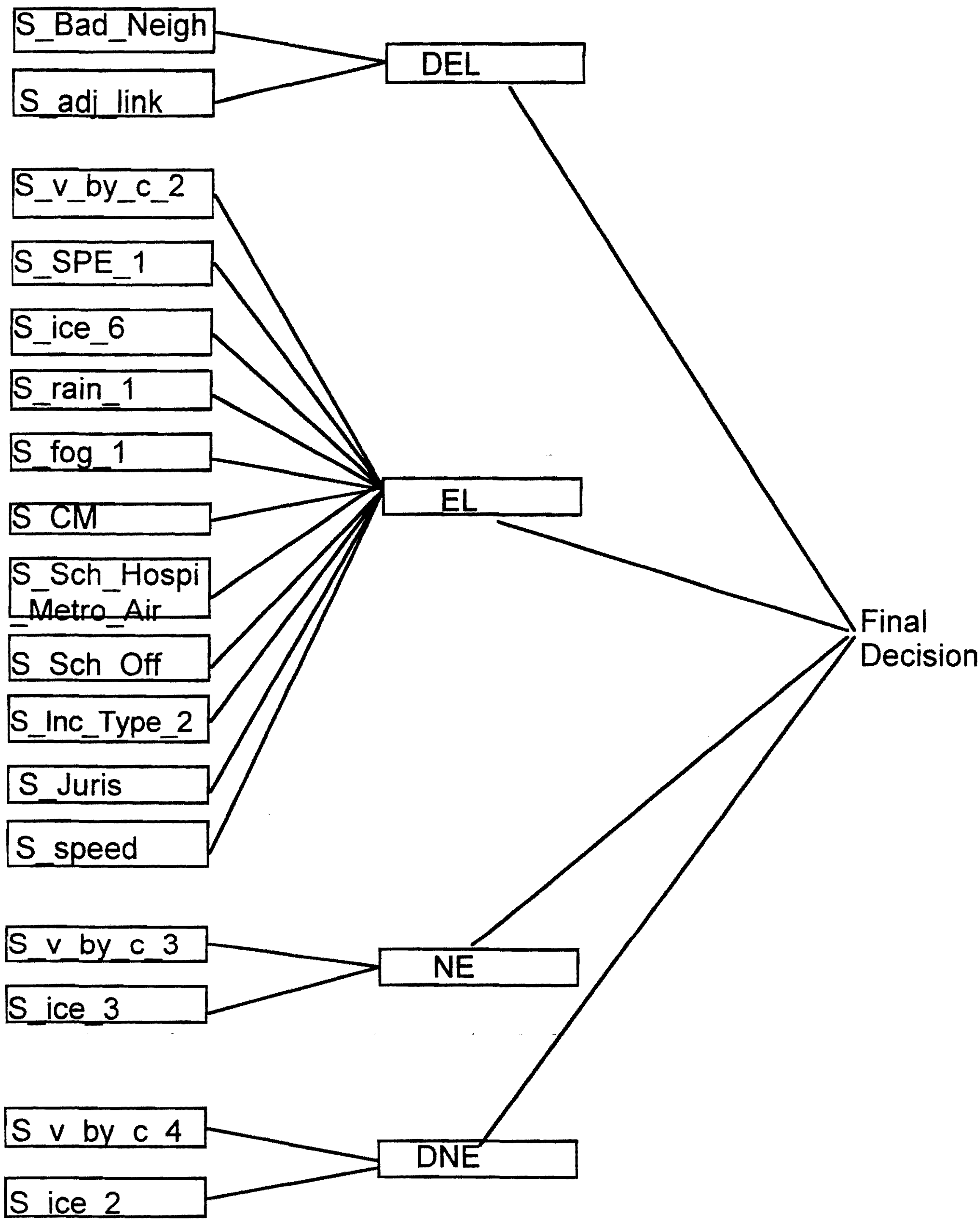


Figure 5.8: Rule Network for Decision Making

5.2.3.11 Rule Antecedents

As explained earlier, a rule antecedent generally comprises of a set of external forces coupled with a set of link attributes. If one were to explicitly enumerate all possible antecedents then a given state of external force and link attribute can lead to only one possible antecedent being applicable and the result of the decision can be immediately obtained. However the number of possible antecedents is numerous and does not permit its explicit enumeration. This would imply the need for some kind of a strategy in which only a limited number of rule antecedents are enumerated and the expert system maps a given combination of external factors and link attributes to several of the enumerated antecedents. If the mapping is unique then we have a decision. If the mapping is not unique and we have several of the enumerated antecedents applicable then we have to use some kind of a conflict resolution mechanism.

5.2.3.12 Link Elimination Rules

This set of rules may be further classified into simple link elimination rules and compound link elimination rules. Simple link elimination rules have antecedents which comprise of one external force and one link attribute. In this section we list out the link elimination rules that the network generator considers. Each rule is followed by an explanation.

Table 5.11: Example Link Elimination Rules

Hypothesis Name	*Antecedent* & Explanation	Decision/ Action
S_v_by_c_1	*if(volume/capacity > threshold_1) Then* Link is too congested.	DEL
S_v_by_c_2	*if(threshold_1 > volume/capacity > threshold_2)* Then Link is congested so eliminate it.	EL
S_v_by_c_3	*if(threshold_2 > volume/capacity > threshold_3) Then* Link is not so congested and we'd like to retain it for diversion purposes.	NE
S_v_by_c_4	*if(threshold_3 > volume_by_capacity) Then* Link offers excellent potential for diversion, as it is not at all congested.	DNE
S_SPE_1	*if((link location = adjacent to special event)&(time=special event beginning or ending time) Then* This rule takes into account the traffic disturbances created by the continuous stream of traffic entering the arterials.	EL
S_SPE_2	*if(time = special event begin time) Then* If we don't have real time data on link volumes then we need to account for the extra volumes generated by the special event.	EL
S_ice_1	*if(weather = icy) and (link type = bridge) Then* Bridges can be very dangerous under icy conditions and heavy traffic.	DEL (CRITICAL)
S_ice_2	*if(weather = icy) and (Ice_Snow clearance = adequate) Then* Not many roads in the arterial network will be adequately cleared by the authorities in charge. Hence it may be appropriate not to eliminate any roads that are well maintained.	DNE
S_ice_3	*if(weather = icy) and (Ice_Snow_Clearance = moderate) Then* If diversion is really needed it might be necessary not to eliminate links that are moderately maintained.	NE

S_ice_4	*if(weather = icy) and (Ice_Snow_Clearance = not cleared) Then* In such conditions the link will be a safety hazard.	DEL (CRITICAL)
S_ice_5	*if(weather = very_icy) Then* In such conditions diversion must not be initiated. This is used as a double check.	DEL (CRITICAL)
S_ice_6	*if(weather = icy) and (gradient = steep) and (Ice_Snow_Clearance = moderate) Then* Steep Slopes are dangerous in icy conditions.	EL
S_snow_1- S_snow_5	Similar to S_ice_1 to S_ice_5	
S_rain_1	*if(rain = heavy_rain) and (road is winding) Then* This situation makes it very difficult for the diverted traffic to flow smoothly).	EL
S_fog_1	*if(fog = heavy) and (road is winding)Then* This situation makes it very difficult for the diverted traffic to flow smoothly.	EL
S_CM	*if(link location = Construction/Maintenance Location) and (time = construction period) Then* The maintenance activity will reduce capacity and also cause additional traffic disturbances.	EL
S_Bad_Neig	*if(link location = Bad Neighborhood Zone) and (Time = Night) Then* We do not wish to divert through a bad neighborhood.	DEL
S_Sch_Off	if(link_location = School_Office_Location) and (time = office_school_open_close_time) Then We may not wish to add further traffic to this link.	EL
S_Sch_Hospi_ Metro_Air	if(link_location = Hospital/Metro/Airport location) Then We don't wish to add more traffic and make access to these sensitive points.	EL
S_Inc_Type_1	if(incident_type = hazmat_severity_1) and (link_location within hazmat_impact_area). Then	DEL (CRITICAL)
S_Inc_Type_2	if(incident_type = hazmat_severity_2) and (link_location within hazmat_impact_area). Then Hazmat accidents.	EL

S_Adj_Bl	if(adjacent links are all DEL) Then if all adjacent links are definitely eliminated, then we eliminate this one too.	DEL (CRITICAL)
S_Juris	If(jurisdictional_factor = yes) Then The reason we don't certainly eliminate the link is that in some cases it may be possible to override jurisdictional issues by the operator contacting the appropriate authorities.	EL
S_speed	if(speed_limit < threshold_6) Then	EL

5.2.4 Route Generation

The network generator has two sets of options for route generation. The first option available to the user is pre-defined diversion plans. These plans could be a set of diversion routes that are currently being employed by the Northern Virginia incident management personnel. If a given incident link has a pre-defined route associated with it then the network generator has the capability to extract this information and display it. If, on the other hand, such a plan is not available, or such a plan is not implementable due to reasons like traffic congestion on the arterials, or construction activity on one of the links in the plan, then the network generator has the capability to dynamically generate diversion routes. The links that are not eliminated are used to form the routes. The route generation is done through use of a static shortest path algorithm that generates a shortest route from a given starting point of the diversion to a given termination point. Although, it is not implemented yet, the route generation function of the network generator will be enhanced by adding:

- static and time-dependent kth shortest path algorithms that have already been programmed in C,

- static and time dependent disjoint pairs of shortest path algorithms that are currently under development (Sherali et al. 1996).

5.2.5 Route Prioritization

Several researchers have estimated probabilistic functions for driver route choice. One way of prioritizing routes is then to use these functions once the routes have been generated based on impedance function criteria such as travel time. However these route choice models consider several variables regarding the class of motorists including household and economic information. While such models are likely to produce accurate predictions on route choice once the data is available, employing these models for real-time diversion is not recommended since this data is usually not known. Hence, we propose an approach in which the route prioritization module of the network generator works by incorporating several travel disutility components in the link impedances used by the shortest path algorithms of the route generation module.

The factors considered in the travel impedance are particularly of interest to diversion. These factors were chosen based on a study conducted by the Texas Transportation Institute to study driver propensity for diversion (Ullman et al., 1994).

- Average distance from the freeway of all links in the route
- Total Arterial Travel Length
- Travel Time on the link

In addition to these factors we can also add:

- Additional Term for links not on familiar diversion routes
- Turn Penalties at each intersection

The factors stated above are combined to give the following form.
Z = k1 (distance from freeway) + k2(arterial travel length) + k3 (Travel Time on Link) + k4 (Penalty for link not being on familiar route). One may normalize the weights above to be of the form of k1+k2+k3+k4=1.

For intersections the turning movements can be represented as links where the only non zero dependent variable in the impedance function is the travel time. This module also has not yet been implemented since the data and multiple route determination capability of network generator is not available either. However, in places such as Northern Virginia, multiple diversion routes are not anyway available in general. Also, the dissemination of multiple route information is quite problematic. Therefore, in this dissertation, although these ideas about route generation and prioritization are discussed, they are not going to be implemented for practical implementation reasons.

5.3 Network Generator Software Development

In this section, the network generator software development efforts are summarized. The software implementation for network generator was a great challenge due to the complexity of the real-time system described above. The software development has been a collective effort of many people involved in the incident management project including the author of this dissertation. The details of this effort is discussed in Ozbay et al. (1996). The aim of this section is to describe the relationship between the modeling and software development and present some of the important features of the network generator.

The network generator is a user friendly, interactive, quick and efficient decision support tool that aids the operator of the transportation system to determine alternate routes for diversion. The essential components of the decision support tool can be characterized by the following features:

- Spatial data management
- Analytical and heuristic reasoning
- Messaging

5.3.1 Spatial Data Management

The first step in the network generators decision making process is a study of the variables of the system. The decision variables for the diversion problem are of two classes - static, and dynamic variables. Static variables are those that do not change on a regular basis, for example node-arc incidence, lane capacities, link length etc. The second class of variables vary with time of day and are called dynamic variables such as, link volumes, weather conditions. The software developed has the capability to handle and manipulate these data. Features that have access to any subset of the data are also in built into the system. The software is specifically designed to handle the large amounts of spatial data and has the capability to handle data on the thousands of links that make up an urban network. The data handling features are in built through the integration of a data base management system with the rest of the software.

Since the data relates to an urban network, the input of the static and dynamic values for the variables, and system output are generated and displayed through the maps to the operator. This objective then requires the software to have built in graphical functions. To accommodate the data handling and display requirements a GIS is made a necessary component of the network generator.

5.3.2 Analytic and Heuristic Reasoning

The data stored in the system is used by the network generator to arrive at decisions necessary for the implementation of a diversion strategy. The network generator employs two styles of reasoning. Knowledge processing or inferencing, is a methodology in which several heuristics are employed to process the state of the system in order to draw inferences on its behavior and control. This type of processing is accomplished through the expert system component of the network generator. The second type of processing employed by the network generator is called procedural processing. Here some mathematical or logical functions are employed to process the data. The network generator uses functions developed using the C programming language for its procedural processing. The combination of procedural and heuristic reasoning capabilities gives the network generator a powerful problem solving framework.

5.3.3 Messaging

The final step in the development of a traffic diversion plan is the dissemination of the route to several agencies that will help coordinate the diversion. These agencies typically include the Department of Transportation, state and local police, fire and rescue, and media. Thus facilities for interagency communication through the software need to be built in. This function is addressed through a client / server architecture as described in later sections of this Chapter. The software components of the network generator are summarized in Table 5.12.

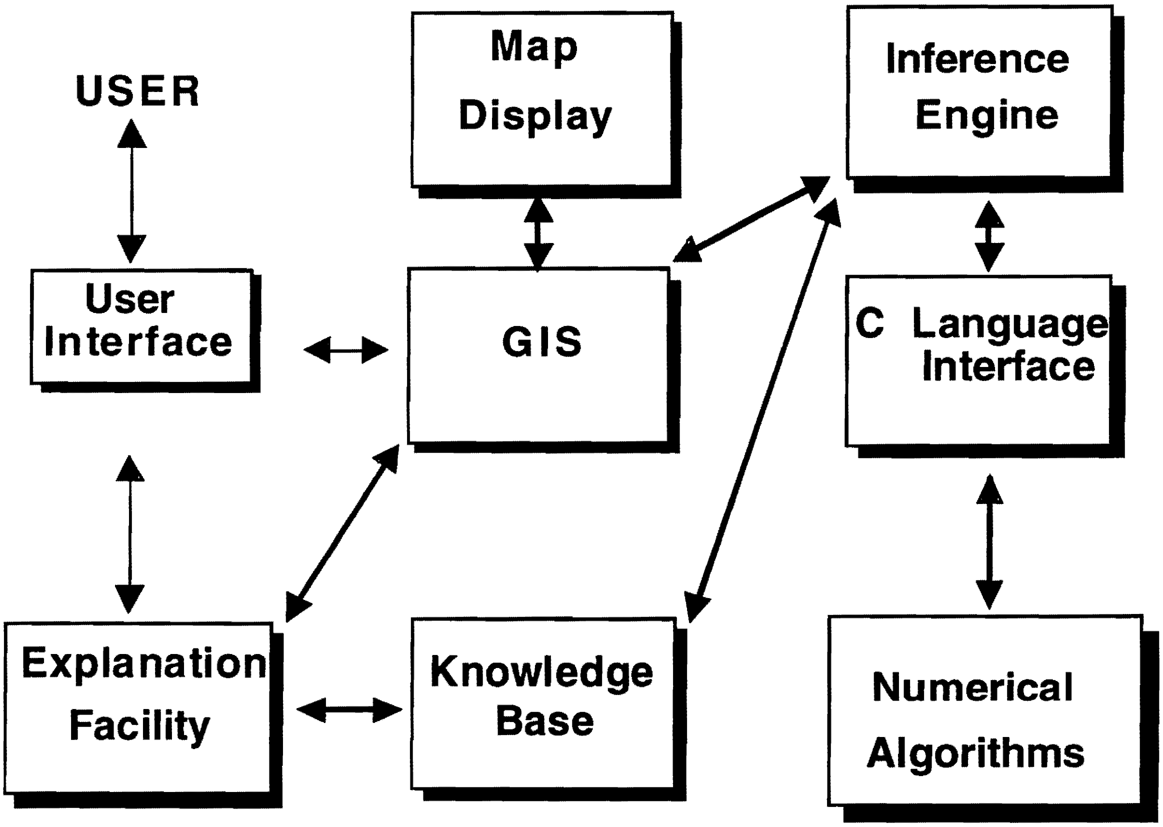


Figure 5.9: Components of the Network Generator

5.3.4 Software Architecture

The current architecture of the network generator is shown in Figure 5.10. There are two essential components of the software. The first component of the software contains the application programs necessary for diversion route generation and dissemination to other agencies that coordinate the diversion. The second component of the software constitutes the communication channel for interagency communication. Thus the architecture of the software is described in this chapter from two perspectives. Presented first are details on the application design. Then presented are the details of the client server architecture itself.

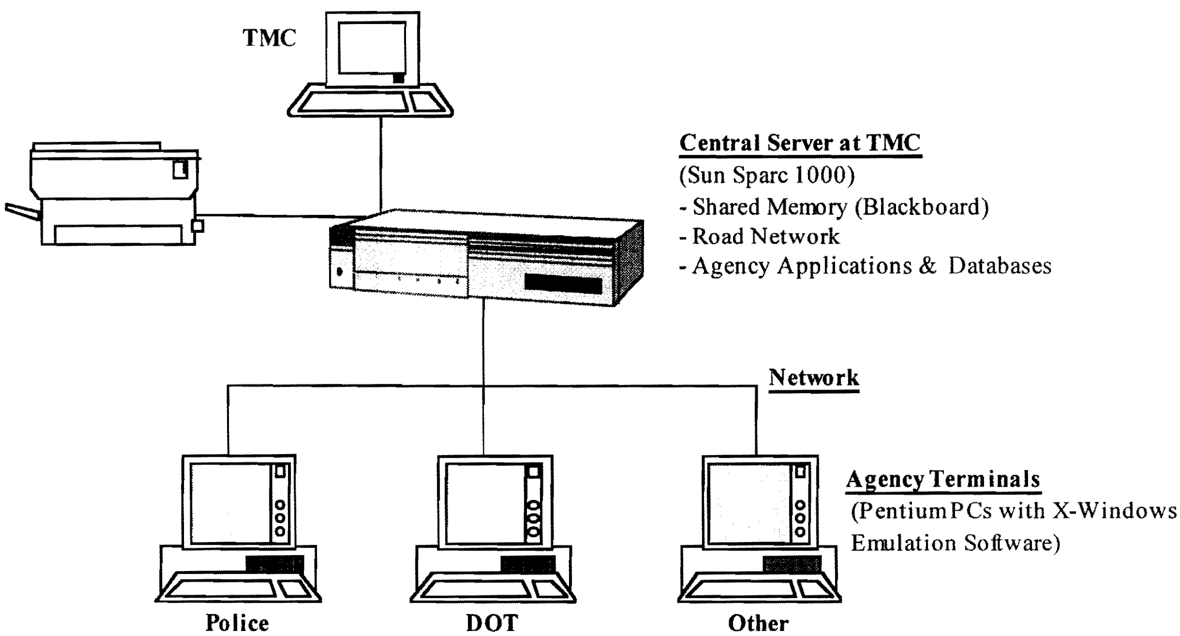


Figure 5.10: Current Implementation Architecture

5.3.5 Application Design

Developed in an UNIX environment on a Sparc-1000 server, Network Generator applications combine the rule-based reasoning capabilities of the Nexpert-Object expert system shell with spatial data handling capabilities of Arc/Info. Linking an expert system with a GIS enables both the expert system and the GIS to perform new tasks, and opens the way for more complex spatial analysis and more flexible querying of GIS databases. Figure 5.11 shows the overall design of the application.

Table 5.12: Network Generator Software Components and Tools Chosen

Functionality	Tool
Data Base Management	Arc/Info
Knowledge Processing/Inferencing	NEXPERT-OBJECT
Graphical and Procedural Programming	C
User Interface	Open Interface Elements

The above design was arrived at after a careful consideration of the available GIS software and expert system shells. Arc/Info, developed and supported by the Environmental Systems Research Institute, Inc. (ESRI), is a powerful toolbox that supports the entire spectrum of GIS applications. Spatial objects in Arc/Info are represented as points, lines or polygons. The locational data associated with these objects are defined through a topological model, while the thematic data are defined using a relational data model in Info. The basic unit of storage in Arc/Info is called a coverage which corresponds to a single layer of a map that contains information about one type of locational feature.

Arc/Info has a layered architecture, the foundation of which is the data engine used to access and manage the geographic database. At the next level, Arc/Info contains a powerful and flexible command language, providing access to sophisticated geoprocessing tools which operate on the various data sources supported by Arc/Info. AML, the ARC Macro Language, provides the development environment in which sophisticated macro procedures can be automated and custom user interfaces can be built. A third method for accessing Arc/Info is through the use of inter-application communications (IAC). IAC tools in AML allow other applications software to execute operations in Arc/Info, thus allowing its use as a GIS data and process server. These capabilities of Arc/Info make it especially suitable for the design of the system described.

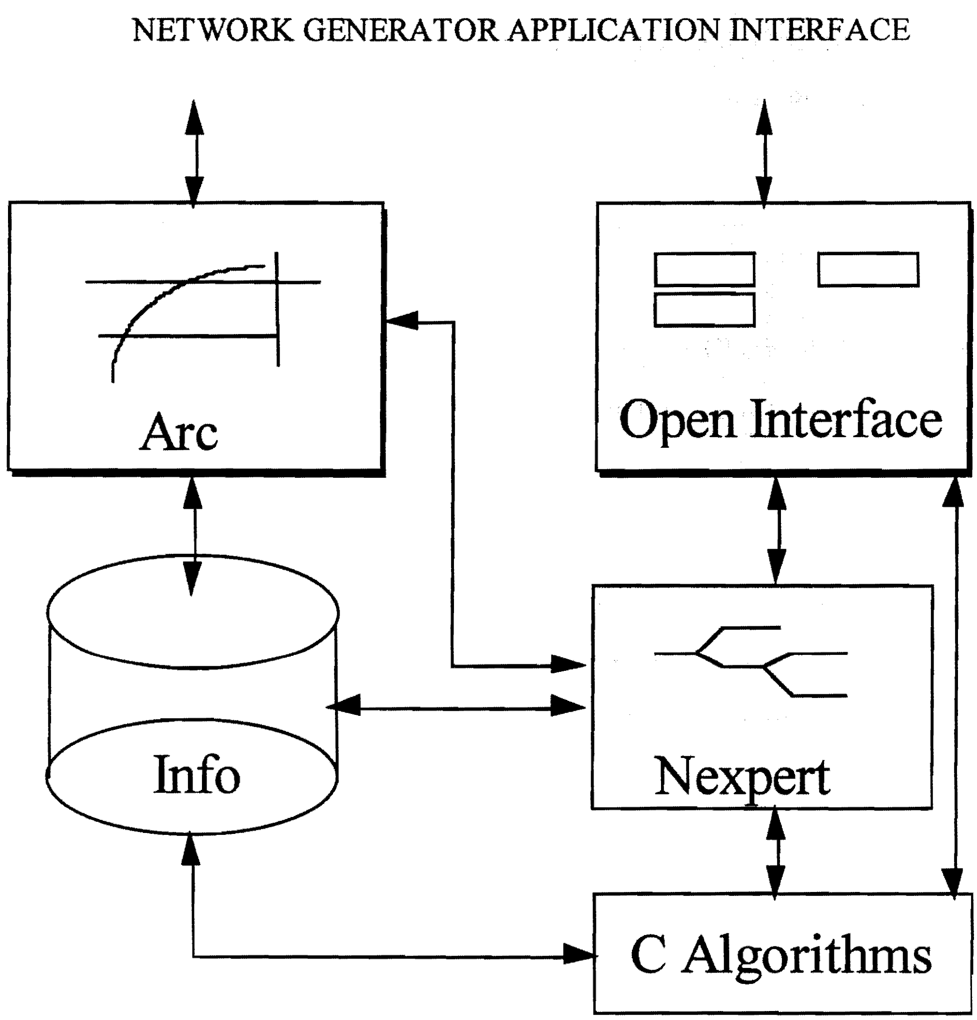


Figure 5.11: Network Generator Application Design

Nexpert-Object, developed and supported by Neuron Data, provides a framework for the construction of rules, and an object-oriented model for representation if the data that the rules act on. A Nexpert rule consists of three parts - a series of conditions, a hypothesis, and a series of actions. If all conditions are true then a hypothesis is true, which can be used to execute a set of actions. Nexpert's rules act on objects which are members of classes and possess properties. Objects and classes are incorporated into the conditions and actions of the rules. Rules can change object class memberships or the values of its properties. Together the rule and the object networks form the knowledge base of the expert system.

Nexpert-Object was chosen for the expert-system side of the Expert-GIS since its application programming interface (API) makes the task of integration relatively easier. In addition, its object oriented data model offers a great deal of flexibility in data representation, and its ability to reason by both forward and backward chaining offers a similar flexibility in logic representation [Site Reference 10]. Although Nexpert has built-in data base conversion modules for several popular data-base systems, it unfortunately does not support Info, the database used by Arc/Info in the UNIX environment.

Neuron Data's Open Interface Elements was chosen to build additional graphical user interfaces that the system needs. Open Interface is a software development environment that permits the development of cross platform applications with native graphical interfaces. Open Interface comprises a set of libraries and a resource editor. It also has a C language API which is a highly modular ANSI C library. The API provides facilities for developing applications with graphical user interfaces for any standard windowing environment.

A key element in the design of the application is bridging the different software environments of Arc/Info and Nexpert-Object and Open Interface to permit easy transfer of data and command level control between the software. The two levels of integration required are - data level integration and command level integration.

Data level integration

An easy approach for data level integration is to employ data files in a format that both Arc/Info and Nexpert-Object can support. However, this would require writing data into a common file format before it can be accessed by either software. This operation is a time consuming process and is not suitable to employ such a method for real-time systems.

In order to permit a more direct data bridging, a C environment was implemented to permit direct access to the Info file structure and Nexpert-Object variables. The Nexpert-Object expert system development package includes an application program interface, which permits developers to access Nexpert variables from C. In addition, Open Interface, the GUI development kit, comes with a tool kit of C libraries that can be used to read and write widget attributes. C functions developed by Todd Stellhorn (ESRI) were used to allow Nexpert to directly read network information from the Info files. This eliminated the otherwise tedious task of writing to a common file format.

Command level Integration

A command level interface permits Nexpert-Object and Arc/Info to issue commands to each other and act upon the command issued by the other program. Although both packages permit the execution of external programs through system calls, only a command interface can provide for the transfer of control between the two programs.

This level of integration was achieved using Arc/Info's Inter Application Communication and by building an environment in C for controlling knowledge processing in Nexpert. The API for Nexpert makes it possible to build a controller for the inferencing and knowledge processing in C which could suggest parameter values and hypothesis directly into the inference engine.

Inter Application Communication (IAC) in Arc/Info enables software applications on remote or local machines to communicate with each other [Site Reference 14]. Based on Open Network Computing's (ONC) remote procedure call (RPC) protocol, IAC provides a way for external applications to request services of an Arc/Info process in server mode (AI-Server) and also makes it possible for an AML application to exploit the capabilities of other applications by being a client of those applications. An important component of IAC are the AI-Server and AI-Client shells. AI-Server provides a shell for creating a server to execute C functions and have these functions directly accessible from Arc/Info's AML-IAC interface. This functionality is used to call Nexpert executables with AML variables as arguments. AI-Client on the other hand provides the ability to create client applications in C that can execute Arc/Info commands. This was used to write C functions that Nexpert-Object could access and execute Arc/Info commands by connecting to a Arc/Info server process and passing requests to that process.

5.3.6 Application Modules

The network generator comprises of a powerful set of independent modules. Each of these modules is designed to carry out a specific step in the process of developing a diversion plan. These modules are listed in Table 5.13 below.

Table 5.13: Application Modules

Function	Tools
Network Static Data Entry	Arc/Info
Network Dynamic Data Entry	Arc/Info
Network Maps	Arc/Info
Data Extraction	Arc/Info-C
Graphical Interaction with System	Arc/Info-C
Search Area Determination	Nexpert - C - Arc/Info
Link Elimination	Nexpert - C - Arc/Info
Route Generation	C-Arc/Info
Route Prioritization	Nexpert-C-Arc/Info
Output Display	Open Interface - Arc/Info - C
Reports	Open-Elements
Inter Agency Communication functions	Arc/Info

5.3.7 Network Static Data Entry

Static data refers to data that does not change on a regular basis. This includes connectivity information, link lengths, number of lanes, capacity/lane, position of traffic signals, historical data - like historic link volumes, accident data etc. The geographic portion of this information is available in available special formats like TIGER that are public domain, or commercial files like ETAK maps. Data from ETAK files are employed by the network generator to form a coverage of the area in Info. Then the coverage can be employed to a wide range of spatial data display and manipulation. To the existing geographic attributes additional attributed for links may be added through an Arc/Info session. The network generator also provides for an interactive tool to edit the link attributes.

5.3.8 Network Dynamic Data Entry

One of the important functions of the network generator, as a decision support system is to be able to handle dynamic conditions on the network. The dynamic data that affect the diversion process may be broadly classified into:

- Incident Data - Location, Time, Day, Date
- Weather - Snow, Ice, Rain Fog
- Network Data - link volumes, capacities, link closures etc., signal plans etc.

The above information is entered by the system operator (the end user of the software) each time a diversion plan is necessary. The data entered by the operator is further filtered by the network generators data processing functions to eliminate errors. These data processing functions further serve to derive additional information from the raw data that the end user enters into the system. Data filtering is done by integrating the GUI widgets of Open Interface Elements with functions written in C. Figure 5.13 shows the data entry window for incident characteristics.

The data entry windows are very user friendly. Choices are presented using pull down menus, choice boxes, check boxes, radio buttons and other graphical widgets. Mouse clicks and key board accelerators can be used to focus on these widgets.

The GIS component of the network generator, permits the user click to the location of the incident and enter the relevant characteristics. Information like dynamic link volumes will be accepted by either reading from user specified files or from a user specified on line data base.

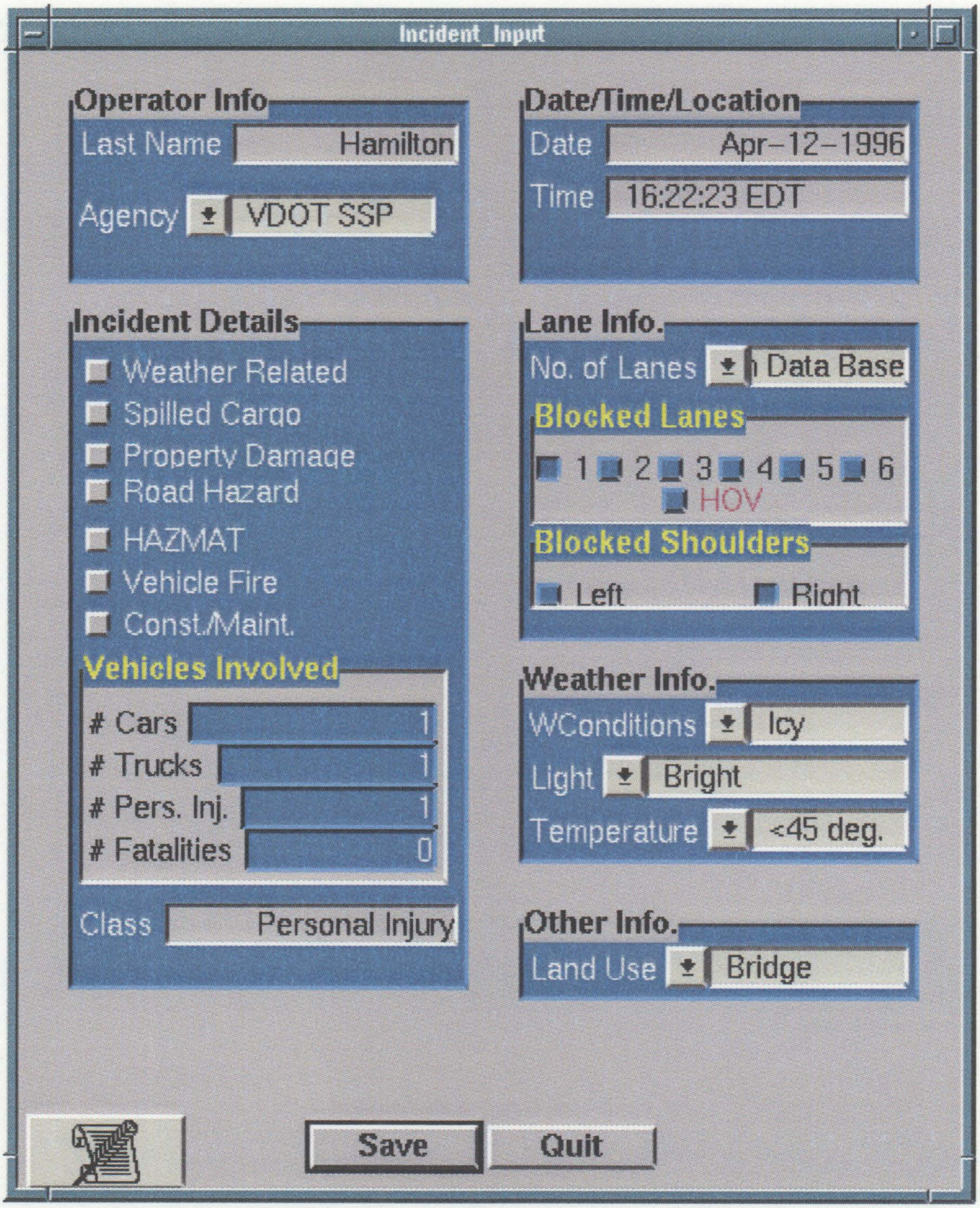


Figure 5.12: Incident Input Menu

5.3.9 Network Maps

The computer generated maps for the area will primarily be done using graphics utilities of Arc, and the data base for the Network as represented using Info.. The software will provide essential features including standard provisions for zooming, panning, selective viewing of transportation infrastructure (highways, interstate, arterials, etc.). Figure 5.14 shows the map display of a portion of the Fairfax County where the incident described in Chapter 3 occurred.

5.3.10 Data Extraction

The data that is required by the expert system may not always be stored in the same form. Hence the data extraction functions of the software are handled by procedural functions written in C in conjunction with retrieval function that AMLs and C interfaces support.

5.3.11 Graphical Interaction with System

The end user of the software will interact with the system using a graphic interface. The graphic interface will constitute the following

1. Computer generated maps
2. Bitmaps
3. Menus
4. Dialogue Boxes
5. Buttons
6. Text Boxes

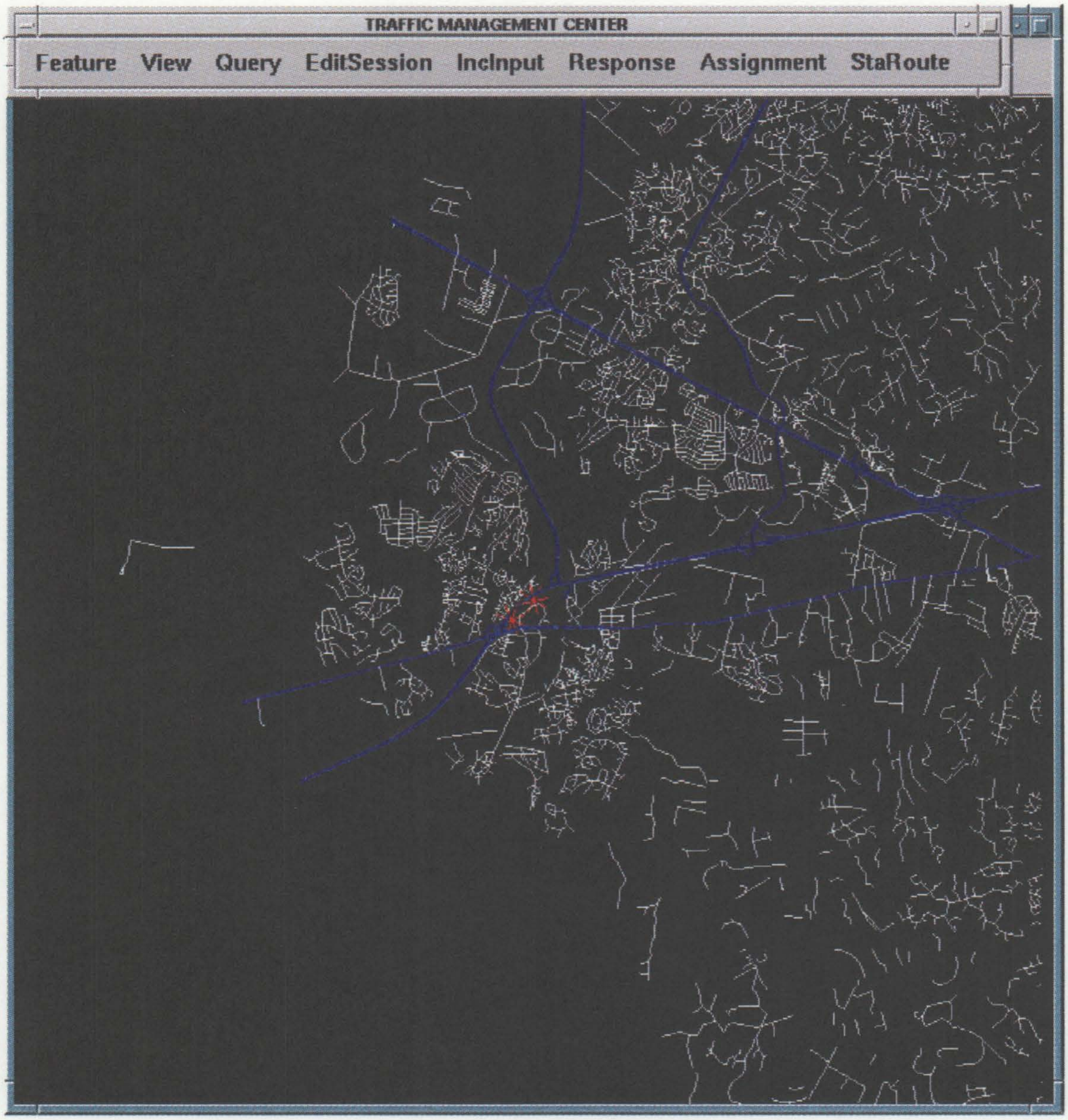


Figure 5.13: Map Display Screen for the Incident

To facilitate the implementation of the above system, the graphic front ends provided by Open Interface Elements are used in conjunction with procedural functions written in C.

5.3.12 Search Area Determination

The area within which the diversion routes is determined is the search area for the network generator. The search area depends on the severity of the incident in the sense that the more severe the incident the larger the size of the search area that would be necessary to efficiently divert the traffic. The network generator uses heuristic rules to determine the size of the search area. These rules are implemented using the NEXPERT-Object expert system shell. The details of the study area will be extracted from the data base.

5.3.13 Link Elimination

There are three steps in this decision making process. The first step is the extraction of all relevant data about the link. This is done, again, using Info and C. The next step is to use this data as input for the heuristics in NEXPERT Object. These heuristics may call upon several procedural functions written in C , to assist in the decision making process. Thus these three components of the software are integrated.

Once a decision is made by the knowledge base the elimination status of each link is transferred to Arc/Info. Then the display is updated and the network generator highlights the eliminated links. The operator is also given the choice of reversing the decision reached by the expert system.

5.3.14 Route Generation, Prioritization

The route generation module of the network generator is a sophisticated module that permits the operator to use pre-defined static routes for diversion. Alternatively, the operator is given the option of using dynamically created routes for the diversion.

5.3.14.1 Static Diversion Routes

Pre-defined diversion plans can be coded into the network generator using the network generator's network editing option developed using Arc/Info. Using this tool one may define a diversion route and associate it with a particular link on the network. For any incident on this link one may display the static diversion plan by choosing the static diversion option. Static routes have two components:

1. the route itself
2. attributes attached to it - such as the location of policemen along the route, location of sign boards etc.

The static route editing option of the network generator allows the user to build the route database and is implemented in ARCEDIT which has functionalities as:

1. Select links
2. Remove or add links to selection
3. Make simple routes
4. Append links to simple route to make complex route
5. Remove route from existing route database.

5.3.14.2 Dynamic Route Generation

The dynamic route generation module of the network generator employs shortest path algorithms to find out dynamic diversion routes that take existing network conditions into effect. The network generator has the ability to incorporate the dynamics of the network as reflected in link volumes, lane closures, and link closures, in determining the diversion routes. If dynamic information on link volumes is not available historic information may be used in generating the diversion route.

5.3.15 Diversion Route Dissemination

The network generator has the ability to disseminate information on the diversion routes to any of the clients of the system. The disseminated route information may be received by each user and his/her display is automatically updated to show the route and its attributes.

5.3.16 Reports

Two kinds of reports are available. A text file is generated that contains explanations of the reasons for elimination. In addition interactively the user can clink on an eliminated link on the map display and find out why it was eliminated.

5.3.17 Case Study for Testing the Network Generator

The Fairfax test network that consists of 57 nodes and 110 links and shown in Chapter 3 is used for testing the Network Generator. The traffic and network data, including the link counts on certain links, is obtained from VDOT. According to the VDOT data, the traffic demand on this network is approximately computed as 12,478 vehicles per hour for 32 origin-destination pairs. The incident that is simulated is also the same incident used in the case study section of the Chapter 4. The effectiveness of Network Generator is tested using the simulation package, INTEGRATION.

INTEGRATION developed by Van Aerde et al. (1986) is capable of analyzing traffic flow and the effects of different ATMIS strategies on the traffic flow. Three scenarios are simulated using INTEGRATION:

1. Static diversion route provided by the Northern Virginia Incident Management manual
2. Diversion route determined by the Network Generator
3. No diversion

The incident location along with the static diversion route used by Northern Virginia incident management personnel and the dynamic diversion route generated by the "Network generator" are all shown in Figure 5.15 .

The incident type and characteristics are summarized in chapter 3. For the simulation purposes, it is assumed that one lane blockage caused 50% capacity reduction. The amount of traffic that is assumed to divert is taken as 40% of the total traffic on I-66. Since INTEGRATION does not have a dynamic traffic assignment algorithm, the amount of the traffic that needs to be diverted has be to be pre-specified. The test case is simulated for different incident durations for each scenario. The simulation results are shown in table 5.14.

Table 5.14: Network-wide Vehicle-Travel Time Increase due to the Incident

Incident Duration (mins)	**Total Network-Wide Vehicle-Travel Time Increase (Veh-Hours)**		
	No Diversion	VDOT Alternate Rte.	NG Alternate Rte.
15	917	0	0
30	2333	200	500
45	3750	2833	1333
60	4500	3500	1000

From the simulation results, it is clear that the alternate route generated by the Network Generator succeeded to reduce the network-wide travel times compared to the VDOT alternate and no diversion scenarios for the given network conditions. Based on these simulation results, we can conclude that network generator is capable of using real-time information and generating routes that can reduce the delays substantially. The fact that VDOT routes are generated fifteen years ago, they are not even able to take into account existing demands and freeway capacities that have definitely changed over the years. Therefore, this might explain the reason for obtaining better results using Network Generator's diversion route. These results make clear that a system such as , Network Generator, that can use the latest information that is available can always outperform a static system such as VDOT's static diversion routes. However, more simulation and on-line studies are needed to assess the real effectiveness of the Network Generator under different conditions.

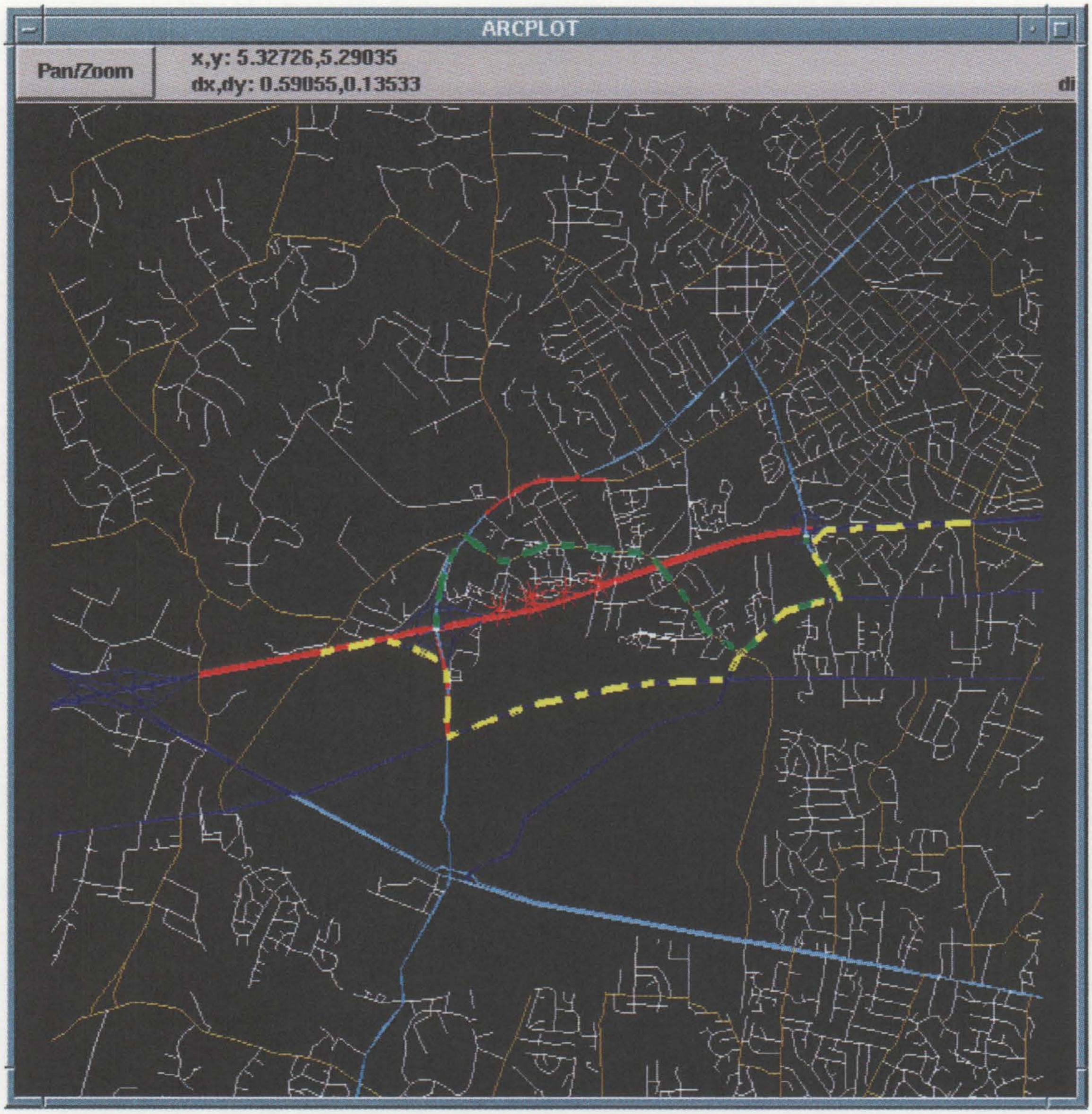


Figure 5.14: Network Generator Output Screen

6. On-Line Traffic Routing / Control Through Feedback Regulation

6.1 Introduction

Dynamic Traffic Routing and Assignment (DTR / DTA) has been an important research topic in transportation engineering. Although there have been several recent attempts to develop real-time Dynamic Traffic Assignment/ Routing (DTA/DTR) algorithms that will perform on-line traffic control, majority of the transportation research has focused on off-line planning problems. With the advent of ITS, the need for dynamic models, capable of working in real-time has become clearer.

The traditional optimization based approaches attempt to solve the DTR/DTA problem by optimizing the objective functions for the nominal model over the "planning horizon". For real time traffic flow control, where on-line sensor information and actuation methods are available, this technique is not very well suited.

On-line traffic management differs from transportation planning in many different ways. A diversion system that will be used for on-line traffic management should make an effective use real-time traffic data obtained from sensors. In other words, it should heavily depend on the data in order to manage the traffic. It should work in real-time and be responsive to the changes in traffic flow. These constraints create a need for a better definition of dynamic traffic routing / control problem. In the light of the facts described above, we can cite the following points for showing that traditional optimization based approaches to solving DTA / DTR fall short of satisfying the requirement of real-time on-line traffic management system:

1. Optimization based approaches take the present state of the transportation system that is modeled and optimized and try to optimize the system for the "planning horizon". In doing so, they predict future states of the system based on the pre-defined assumptions about the dynamics of the system. This approach works well for planning purposes where the time variances of the system do not affect the system too much. But, for on-line control, the classical optimization approach which depends on the present network conditions and predictions to optimize 1 hour (more or less) into the future, is not realistic.
2. Smallest disturbance to the system will completely change the set optimized predictions determined by the optimization approach. This could mean that the extensive computations used to reach the global (or local) optimal solution could be wasted because of a small disturbance.
3. The complexity of the optimization models will make it very difficult to solve these models in real-time. A solution that does not work in-real-time is not acceptable at all for on-line traffic management and control.
4. Existing DTA/DTR models that optimize the traffic system by distributing traffic flows to the alternate routes do not address at all to the problem of effectively using this information in-real time. Once the optimal traffic flows for each time period are determined, the implementation of these results for on-line traffic control is never discussed in detail.

6.2 Feedback Control

Feedback control as the term implies, refers to the control of a system to achieve certain specified objectives by feeding some output signal of the system back into the controller. The feedback mechanism makes the system a closed loop control system. In open loop systems, there is no feedback. This difference can be illustrated by a simple but representative example of driving a vehicle on a curved track by controlling the throttle and steering. An open loop control design would start by mathematically modeling the system.

An optimal control, which is an open loop control, can be designed for optimizing some objective functional based on this mathematical model assuming that the parameters of the model are accurate. For instance, if the model was non-linear, we could use a non-linear numerical technique such as sequential-gradient-restoration algorithm to solve for the control signals. This would give us time varying control signals for the throttle and the steering from initial time to the final time of the planning horizon. Then, this would mean that the vehicle will be driven by these control signals without any real-time input from the environment. This controller is not designed explicitly to deal with the real perturbations to the mathematical model due to the unmodeled dynamics of the actual system. For example, the perturbation could be due to the simplifications during the modeling process or lack of knowledge of some parameters/dynamics and real-time occurrence of an unexpected situation. These perturbations could make the system *at least* sub-optimal, if not unstable, since the optimality was calculated based on the nominal off-line parameters and initial conditions.

Feedback control design on the other hand is based on the real-time feedback information. For instance in our example, we could use sensors which would provide the controller with the real-time information about the coordinates of the vehicle with respect to some convenient frame of reference on the track. The controller will be driven by the difference between the sensed location of the car on the track and the desired location of the car at each sampling instant.

In order to overcome the aforementioned problems, we are proposing a feedback control approach to the DTA / DTR problem. Similar approach has been discussed by Papageorgiou et al. (1983) for simple traffic networks consisting of two alternate routes. At every sampling interval the feedback controller produces control signals reacting to the information coming from the sensors. The control update time in this situation is close to the sensor update time which is of orders of magnitude lower than what would be typically used for a pseudo-on-line optimal control problem. Since, a typical time horizon for optimal control approach is much higher than the sensor sampling update time, the optimal control solution would be sensitive to the short-time traffic flow changes within that time horizon. Feedback control approach does not have this important limitation. In fact, the feedback control can be approximately as responsive as the sensor update sampling time.

Moreover, resources required to solve an optimal control problem is much more than what is needed for feedback control because, optimal control problem has to be solved for the whole planning horizon whereas feedback control responds to the sensed variables almost immediately. In the actual implementation of the feedback control there are minimal calculations required to generate the control signal from the sensed variables. For instance, in a Proportional-Integral-Derivative (PID) control of a single-input-single-output (SISO) system , there are only three additions and one multiplication required to generate the control signal. Compared to this, the optimal control model will require intensive processing prone to errors due to the lack of sampling within the time horizon.

6.3 System Dynamics and Feedback Control Problem Formulations for Real Time Dynamic Traffic Routing / Control

In this section, we discuss the design methodology for feedback control laws for dynamic traffic routing / control problem (DTR). Although the original system is infinite dimensional, the control input is finite dimensional. The feedback process is illustrated in Figure 6.1 shown below. There are essentially three ways the system modeling could be used to design feedback controllers. These are:

1. Distributed Parameter setting, represented by Partial Differential Equations (PDE).
2. Continuous Time Lumped Parameter setting, represented by continuous time Ordinary Differential Equations (ODE).
3. Discrete Time Lumped Parameter setting, represented by continuous time ODEs.

The rest of this chapter is divided into seven sections as follows. Section 6.4 describes the system dynamics derived from fluid flow analogy to traffic, which is the foundation for the control systems. Section 6.5 provides background information on the design of feedback control in the distributed parameter setting. We show results on the control of Burgers' equation, which is similar to traffic flow equation after a change of variables, and normalization. Statements of two types of control problems are given in the appendix for

Burgers' equation, one in which the controller enters through the dynamic equation, and the one where it enters through the boundary condition. Section 6.5.2 gives the space, and time discretization of the system dynamic equations. Section 6.6 gives the feedback control formulations in terms of lumped parameter system in continuous and discrete time for the DTR problem, section 6.7 presents an example of a quantitative control for discrete time lumped parameter setting and the simulation results, and section 6.8 presents an example of a qualitative control using fuzzy logic for discrete time lumped parameter setting and the simulation results.

There are various advantages and disadvantages in designing the feedback controller using any one of the three kinds of models. The original PDE model is derived from the hydrodynamic analogy presented by Lighthill and Whitman (1955). It is however difficult to design a feedback controller directly for a distributed parameter system, and it is an area of active research. By space discretizing the model, we can design a feedback controller in the continuous time domain which can be easier to design. This model obviously will have discretization errors, which however could be reduced by designing a robust controller that would attempt to eliminate these errors. Finally, it is natural to design a controller using a discrete time ODE model of the system for discrete implementation of the control. Again this model would have more discretization errors, and the controller would have to minimize the effect of those.

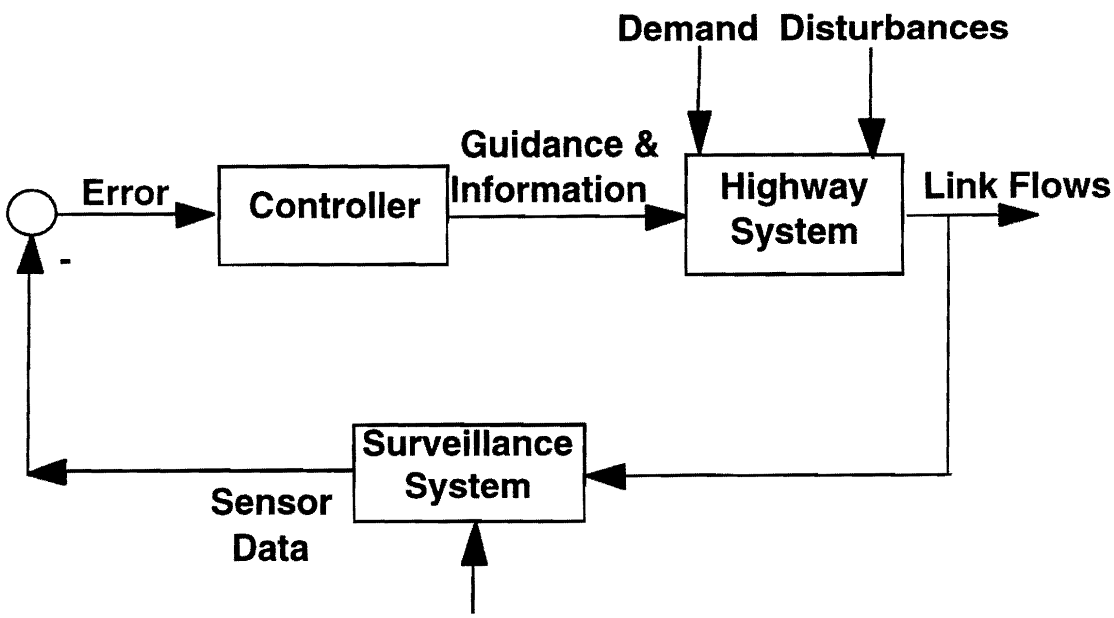


Figure 6.1: Block Diagram for DTR Feedback Control

6.3.1 Preliminary Considerations for Using Feedback Control

Feedback control for DTR can be an effective solution for alleviating traffic congestion during major incidents. However the success of such a system depends on the effective modeling of the system as well as the design of the appropriate control law. The designer of the controller needs to address issues such as controllability and observability of the traffic system, actuation and sensing, robustness, and stability of the closed loop system. *Actuation and sensing* : The actuation of this system can be achieved in many ways, such as Variable Message Sign (VMS), in-vehicle guidance, and Highway Advisory Radio (HAR). State variables such as the traffic density, average traffic speed, etc. can be sensed using various types as traffic sensors such as inductive loops, traffic cameras, transponders, etc. can be used as sensors.

Controllability and Observability : The designer should analyze the system before designing the controller to determine if the system is controllable and observable. Controllability implies that a suitable control law can be devised in order obtain a desired response from the system. Observability implies that the system state variables can be observed from the sensed output. For instance, if the system is not controllable, then we might decide to add more actuation infrastructure such VMS, HAR, etc., and if the system is not observable, we might add more sensors in the system.

Robustness, and stability: : The effectiveness of the control design can be measured in terms of its robustness, stability, and transient characteristics. A robust controller will perform well even in the presence of uncertainties in the nominal model of the system. Models representing traffic systems can not represent the system fully, and therefore there are uncertainties in the system which have to be dealt with. A control law should provide stability to the system and desirable transient response. For instance, a good DTR control law would minimize time for the system to change from a congested state to a normal flow state.

6.4 System Dynamics

The first step in the design of feedback controllers for DTR is to model the system dynamics appropriately. A macroscopic model of the traffic can be effectively used in this context. From the macroscopic perspective, the traffic flow is considered analogous to a fluid flow, which is a distributed parameter system represented by partial differential equations. Mass conservation model of a highway, characterized by $x \in [0, L]$, which is the position on the highway, is given by

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}q(x,t) \qquad (6.1)$$

where $\rho(x,t)$ is the density of the traffic as a function of x, and time t, and $q(x,t)$ is the flow at given x, and t. The flow $q(x,t)$ is a function of $\rho(x,t)$, and the velocity v(x,t), as shown below:

$$q(x,t) = \rho(x,t)v(x,t) \qquad (6.2)$$

This model of a highway section is shown in Figure 2.

Figure 6.2:Highway Model

There are various static and dynamic models which have been used to represent the relationship between v(x,t) and $\rho(x,t)$. One of the most simple models is the one proposed by Greenshield (Gazis D.C., 1974), which hypothesizes a linear relationship between the two variables.

$$v = v_f(1-\frac{\rho}{\rho_{max}}) \qquad (6.3)$$

where v_f is the free flow speed, and ρ_{max} is the jam density.

6.5 Feedback Control for the Traffic as a Distributed Parameter System

The system represented by equations (6.1-6.3) is an infinite dimensional representation of the traffic, since it has infinite state variables. There are two ways to design a DTR feedback controller for such a system. One way is to work in the infinite dimensional domain, and design a controller, which then can be discretized. Another way is to space discretize Partial Differential Equation (PDE) to obtain an Ordinary Differential Equation (ODE) representation of the system. The ODE representation is a standard representation for most of the results available in feedback control theory, and therefore, using that representation many techniques from control theory can be utilized.

The modeling of traffic in the PDE domain provides a reasonably accurate model of the traffic system, especially since phenomena such as shock waves are effectively represented. Hence, it would be highly desirable to design a feedback controller directly utilizing this model. However, it is not trivial to design feedback control laws using this modeling scheme. Fortunately, transforming the traffic model into a more convenient form of Burgers' equation reduces the complexity of the control design. Burgers' equation is just a way of representing the same hydrodynamic traffic flow model in a different form utilizing the diffusion behavior of the traffic. Hence this model is not different in physical interpretation than the classical models which are being widely used in the traffic modeling community.

In the PDE context, researchers have used Burgers' equation to model the traffic flow (Fletcher, 1982, Musha et al., 1978). There has been limited amount of research work conducted on the problem of computing feedback laws for Burgers' equation, which is a nonlinear PDE (Burns et al., 1991a, 1991b). Burgers' equation

$$\frac{\partial}{\partial t}\rho(x,t)+\rho(x,t)\frac{\partial}{\partial x}\rho(x,t)=\varepsilon\frac{\partial^2}{\partial x^2}\rho(x,t) \tag{6.4}$$

was introduced by Burgers (Burgers, 1939, 1948, 1972) as a simple model for turbulence, where ε is a viscosity coefficient. The following is borrowed from (Fletcher, 1982) to show how the traffic problem can be modeled as Burgers' equation.

In order to account for the fact that drivers look ahead and modify their speeds accordingly, (6.3) can be replaced by

$$v = v_f (1 - \frac{\rho}{\rho_{max}}) - D(\frac{\partial \rho}{\partial x}) / \rho \tag{6.5}$$

Using (6.5), relationship (6.2) now can be replaced by

$$q(x,t) = \rho(x,t) v(x,t) - D \frac{\partial}{\partial x} \rho(x,t) \tag{6.6}$$

where D is a diffusion coefficient (Burgers, 1972) given by

$$D = \tau v_r^2 \tag{6.7}$$

where v_r is a random velocity, and τ is the mean collision time for the cars. Many researchers have used the diffusion term in traffic modeling (Fletcher, 1982, Musha et al., 1978, Papageorgiou et al., 1989), and some have used Burgers' equation to represent traffic system (Fletcher, 1982, Musha et al., 1978). Combining equations (6.1) and (6.6) gives

$$[\frac{\partial}{\partial t} \rho(x,t) + v_f \frac{\partial}{\partial x} \rho(x,t)] - 2 \frac{\rho}{\rho_{max}} v_f \frac{\partial}{\partial x} \rho(x,t) - D \frac{\partial^2}{\partial x^2} \rho(x,t) = 0 \tag{6.8}$$

If we introduce a moving reference frame

$$\xi(x,t) = -x + v_f t \tag{6.9}$$

and nondimensionalize $\rho(x,t)$ by $\rho_{max}/2$, and t by t_0, equation (6.7) gets transformed to

$$\frac{\partial}{\partial t} \rho(\xi,t) + \rho \frac{\partial}{\partial \xi} \rho(\xi,t) - \frac{1}{R_e} \frac{\partial^2}{\partial \xi^2} \rho(\xi,t) = 0 \tag{6.10}$$

Here, R_e is a dimensionless constant, and is analogous to the Reynolds number in fluid dynamics. R_e is given by

$$R_e = (\frac{v_f}{v_r})^2 \frac{t_0}{\tau} \tag{6.11}$$

Equation (6.10) shows the Burgers' equation formulation of the traffic flow problem. Many researchers have also worked on the conservation law

$$\frac{\partial}{\partial t}\rho(x,t)+\rho(x,t)\frac{\partial}{\partial x}\rho(x,t)=\varepsilon\frac{\partial^2}{\partial x^2}\rho(x,t) \tag{6.12}$$

with a solution obtained by taking the following limit.

$$\rho(x,t)=\lim_{\varepsilon \to 0}\rho^{\varepsilon}(x,t) \tag{6.13}$$

where $\rho^{\varepsilon}(x,t)$ satisfies (6.4) (Cole, 1951, Glimm, 1970, Hopf, 1950, Lax, 1973, Maslov, 1987a, 1987b). Using this form reduces the Burgers' equation formulation into the classical traffic model with no diffusion. So, even in the PDE domain, we can try to work on the same model as the classical traffic models.

Work on the feedback control of Burgers' equation has been performed [Burns et al., 1991a, 1991b, Curtain, 1984) by several researchers in the past. Curtain (1984) showed using Kielhofer's stability results for semilinear evolution equations (Kielhofer, 1974), that there exists a stabilizing feedback law which can be obtained from the linearized equation, when the domain of the output operator is a certain subspace of L^2 which contains the Sobolev space H_0^1, where for a given domain Ω with boundary $\partial\Omega$, $L^2(\Omega)$ is the space of all measurable functions f such that $\int_\Omega |f(x)|^2 dx < \infty$, and $H_0^1(\Omega)$ is the set of all functions f in $L^2(\Omega)$ such that the derivatives f' (or ∇f) are also in $L^2(\Omega)$ and $f|_{\partial\Omega} = 0$, implying that f=0 on the boundary. Burns et al. (1991a) show the design of Linear Quadratic Regulator (LQR) optimal controller for the linearized equation with bounded input/output, and with unbounded control and observation (1991b). They also study a boundary control problem (1991b).

The DTR problem in the distributed parameter domain is essentially a boundary injection feedback control problem, where the control input enters the traffic system dynamics through the boundary condition. One solution to that relies on converting it into an affine control problem and applying LQ method. Some background development towards that is presented in Kachroo and Ozbay (1996).

6.5.1.1 DTR Formulation

One of the most common diversion scenarios involves the point diversion of traffic in order to route vehicles from one entry point of a congested freeway to an alternate route. This is a problem with major practical and application related implications. This is also relatively easy to implement as the local traffic system can be seen as a decoupled system, and it has limited infrastructure requirements. This sample problem shown in Figure 3, where there are two alternate routes, and the incoming traffic has to be divided between the two, is chosen to illustrate the formulation of the DTR problem in PDE setting.

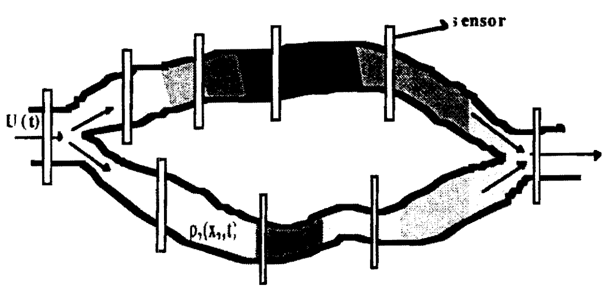


Figure 6.3: Alternate Routes

We can write two Burgers' equations, one for each route. The following are the conservation equations (from which the Burgers' equations are to be derived) for the two routes.

$$\frac{\partial}{\partial t}\rho_1(x_1,t)+\frac{\partial}{\partial x}q_1(x_1,t)=0, \quad 0\le x_1 \le L_1, \tag{6.14}$$

$$\frac{\partial}{\partial t}\rho_2(x_2,t)+\frac{\partial}{\partial x}q_2(x_2,t)=0, \quad 0\le x_2\le L_2, \tag{6.15}$$

with the following boundary conditions at $(x_1,x_2)=(0,0)$.

$$q_1(0,t)=\beta(t)U(t) \quad \text{and} \quad q_2(0,t)=(1-\beta(t))U(t), 0\le\beta(t)\le 1. \tag{6.16}$$

These boundary conditions are written in terms of flow. By substituting the relationship (6.6) in (6.16), we get the boundary conditions as nonlinear relationships as shown below.

$$\rho_1(x_1,t)\,v(x_1,t)-D\frac{\partial}{\partial x_1}\rho_1(x_1,t)=\beta(t)U(t),$$

$$\rho_2(x_2,t)\,v(x_2,t)-D\frac{\partial}{\partial x_2}\rho_2(x_2,t)=(1-\beta(t))U(t). \tag{6.17}$$

The boundary conditions at $(x_1,x_2)=(L_1,L_2)$ should be absorbing boundary conditions, so that the outflow of the sections is dependent on the dynamics of the traffic rather than on the imposed boundary condition. The output bounded equations for this traffic system, based on the sensors used, are:

$$y_i^1(t)=\wp_i^1(q_1(x_1,t)), \quad i=1,2,....,m,$$

$$y_j^2(t)=\wp_j^2(q_2(x_2,t)), \quad j=1,2,...,p, \tag{6.18}$$

where the bounded function $\wp_i^1(.)$ is given by

$$\wp_m^k(q_k(x_k,t))=\frac{1}{2\delta}\int_{\bar{x}_m-\varepsilon}^{\bar{x}_m+\varepsilon} q_k(x_k,t)dx_k. \tag{6.19}$$

The control aim is to solve for user equilibrium, i.e. to equate travel times on the two alternate routes. The nonlinear optimal control problem can be stated as: find $\beta_o(t)$, the optimal $\beta(t)$, which minimizes

$$J(\beta)=\int_0^{t_f}[\int_0^{L_1}\chi(x_1,t)dx_1-\int_0^{L_2}\chi(x_2,t)dx_2]^2dt, \tag{6.20}$$

where $\chi(.,.)$ is the travel time function and t_f is the final time. Note that a feedback solution is needed for the problem, and not an open loop optimal control. Hence, we can either decide the structure of the feedback control such as a PID control with constant gains, and solve numerically for the optimal values of the gains, or we can state the control objective for a standard feedback control problem, such as steady state asymptotic stability

$$Lt_{t\to\infty}[\int_0^{L_1}\chi(x_1,t)dx_1 - \int_0^{L_2}\chi(x_2,t)dx_2] \to 0 \tag{6.21}$$

and some transient behavior characteristics like some specified settling time, percent overshoot, etc. We could also linearize the nonlinear system, and formulate the LQR problem solution of which is a state feedback control, or try some extensions of techniques developed for control of nonlinear ODEs to nonlinear PDEs.

This problem can be generalized to "n" alternate routes as follows:

Problem 6.5.4: Find β_o^i, i=1,2,...,n, which minimize

$$J(\beta_0^i, i=1,2,..n)=\int_0^{t_f}\sum_{k,p}\{\int_0^{L_k}\chi(x_k,t)dx_k - \int_0^{L_p}\chi(x_p,t)dx_p\}^2 dt \tag{6.22}$$

(k=1,2,...,n, p=1,2,...,n, and the summations are taken over total number of combinations of n and p, and not permutations so that (k,p)=(1,2) is considered the same as (k,p)=(2,1), and hence only one of these two will be in the summation), or which guarantee

$$Lt_{t\to\infty}\mathbf{e} \to \mathbf{0}, \tag{6.23}$$

where $\mathbf{e} = [\{\int_0^{L_1}\chi(x_1,t)dx_1 - \int_0^{L_2}\chi(x_2,t)dx_2\},...,\{\int_0^{L_k}\chi(x_k,t)dx_k - \int_0^{L_p}\chi(x_p,t)dx_p\},...]$

with some transient behavior characteristics like some specified settling time, percent overshoot, etc. for the system

$$\frac{\partial}{\partial t}\rho_i(x_i,t) + \frac{\partial}{\partial x}q_i(x_i,t)=0, \quad 0 \le x_i \le L_i, \quad i=1,2,...,n, \tag{6.24}$$

with the boundary conditions

$$\rho_i(x_i,t)v(x_i,t) - D\frac{\partial}{\partial x_i}\rho_i(x_i,t) = \beta^i(t)U(t), \tag{6.25}$$

absorbing boundary conditions on the other boundary, output equations

$$y_i^j(t) = \wp_i^j(q_i(x_i,t)), \quad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,k_j, \tag{6.26}$$

and input constraint

$$\sum_{i=1}^{n}\beta^i = 1. \tag{6.27}$$

The input constraint ensures that the inflow of the traffic is distributed among the candidate alternate routes.

6.5.2 Discretized System Dynamics

Many researchers have studied and designed optimal open loop controllers utilizing space and time discretized models of traffic flow (Tan et al., 1993, Friesz et al., 1989). Some researchers have also designed feedback control laws using similar models (papageorgiou et al., 1991, Papgeorgiou, 1983). The reason for the popularity of these models is that there are many techniques available to deal with discretized systems. The same is also true for feedback control, and hence, in order to utilize the various linear and nonlinear (Kuo, 1987, Mosca, 1995, Isidori, 1989, Slotine, 1991) control techniques available for lumped parameter systems, the distributed parameter model is space discretized Papageorgiou, 1983). For this the highway is subdivided into several sections as shown in Figure 6.4.

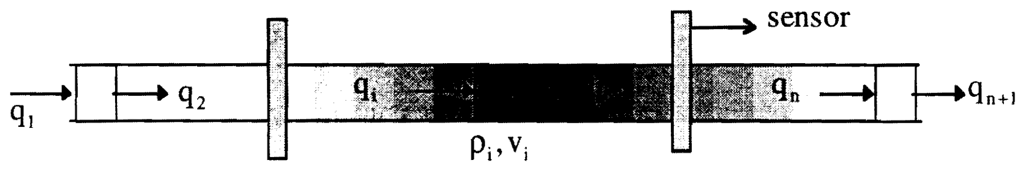


Figure 6.4: Highway Divided into Sections

The space discretized form of (6.1) produces the following n continuous ODEs for the n sections of the highway.

$$\frac{d}{dt}\rho_i = \frac{1}{\delta_i}[q_i(t) - q_{i+1}(t) + r_i(t) - s_i(t)], \quad i = 1,2,...n. \tag{6.28}$$

Here, $r_i(t)$ and $s_i(t)$ terms indicate the on-ramp and off-ramp flows. Equation (6.28) combined with (6.2), (6.3) or (6.4), (6.5) depending on the decision to include the diffusion term), and the output equations (6.29) give the mathematical model for a highway, which can be represented in a standard nonlinear state space form for control design purposes.

$$y_j = g_j(\rho_1, \rho_2, ..., \rho_n), \quad j = 1,2,...,p, \tag{6.29}$$

The standard state space form is

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t)],$$

$$\mathbf{y}(t) = \mathbf{g}[\mathbf{x}(t), \mathbf{u}(t)],$$

$$\mathbf{x}(0) = \mathbf{x}_0, \tag{6.30}$$

where $\mathbf{x} = [\rho_1, \rho_2, ..., \rho_n]^T$ and $\mathbf{u}(t) = q_0(t)$.

There are various other proposed models, which are more detailed in the description of the system dynamics. The phenomenon of shock waves, which is very well represented in the PDE representation of the system, is modeled by expressing the traffic flow between two contiguous sections of the highway, as the weighted sum of the traffic flows in those two sections which correspond to the densities in those two sections (Lighthill et al., 1955, Richards, 1956, Berger et al., 1977). A dynamic relationship instead of a static one like (6.3) has also been proposed by payne (1971) and used successfully.

The model thus obtained can also be time discretized to transform the continuous time model into a discrete time mode. A comprehensive model, which incorporates shock waves, as well as represents the dynamic nature of mean speed propagation, is shown in Papageorgiou (1983) and is reproduced here for completion. The difference equations

$$\rho_j(k+1)=\rho_j(k)+\frac{T}{\delta_j}[q_{j-1}(k)-q_j(k)],$$

$$v_j(k+1)=v_j(k)+\frac{T}{\tau}[v_e(\rho_j(k))-v_j(k)]+\frac{T}{\delta_j}v_j(k)[v_{j-1}(k)-v_j(k)] \qquad (6.31)$$

$$-\frac{vT}{\tau\delta_j}[\frac{\rho_j(k+1)-\rho_j(k)}{\rho_j(k)+\vartheta}]$$

with the relationships

$$q_j(k)=\alpha\rho_j(k)v_j(k)+(1-\alpha)\rho_{j+1}(k)v_{j+1}(k), \quad 0\le\alpha\le 1,$$

$$v_e(\rho)=v_f[1-(\frac{\rho}{\rho_{max}})^l]^m, \qquad (6.32)$$

output measurements of traffic flows q and time mean speeds y, shown as

$$y_j(k)=\gamma v_j(k)+(1-\gamma)v_{j+1}(k), \quad 0\le\gamma\le 1, \qquad (6.33)$$

and the boundary conditions

$$v_0(k)=y_0(k), \qquad (6.34)$$
$$\rho_{n+1}(k)=q_n(k)/y_n(k),$$

gives the discrete system dynamics, which can be represented in the standard nonlinear discrete time form

$$\mathbf{x}(k+1)=\mathbf{f}(\mathbf{x}(k),\mathbf{u}(k)), \qquad (6.35)$$
$$\mathbf{y}(k)=\mathbf{g}(\mathbf{x}(k),\mathbf{u}(k)),$$
$$\mathbf{x}(0)=\mathbf{x}_0,$$

where control $\mathbf{u}(k)$ for a single highway can be the input flow, and mean speeds at the entrance and the exit of the highway.

If the control actuation is discrete, such as the ones implemented by microprocessors and computers, feedback control laws can be designed based on the discrete model (6.35), or can be designed using (6.30) after which the controller can be discretized.

6.6 Feedback Control for the Traffic as a Lumped Parameter System

In the discretized traffic flow model, the freeway is divided into sections with aggregate traffic densities. Sensors are used to measure variables such as densities, traffic flow and traffic average speeds in these sections, which can be used by the feedback controller to give appropriate commands to actuators like VMS, HAR, etc.

There are essentially two ways to design controllers for such nonlinear traffic systems of the form (6.30) and (6.35). One way is to design the controller by linearizing the nonlinear dynamics of the system about its equilibrium, or a trajectory; the other way is to design directly for the nonlinear system. The first way is easier, since immense literature is available describing the various design techniques, especially for LTI systems, but the results are valid only where the linearization is applicable. On the other hand, design of controllers directly for the nonlinear system is much more difficult, but the results are usually global. Some of the linear control techniques are LQR, LQG, PID, H_∞, preview control, etc., and some of the nonlinear ones are describing functions design, feedback linearization, sliding mode control, nonlinear H_∞, etc. Qualitative methods such as fuzzy control and expert systems can also be utilized. Qualitative methods are usually easy to design but difficult to tune and analyze. Some of these issues are discussed by Messmer and Papageorgiou (1994).

6.6.1 DTR Formulation

For two alternate routes problem, as shown in Figure 6.3, the two routes are divided into n_1, and n_2 sections respectively. For simplicity, we are considering static velocity relationship, and ignoring the effect of downstream flow for the model

$$\frac{d}{dt}\rho_{i,j}=\frac{1}{\delta_i}[q_{i,j-1}(t)-q_{i,j}(t)],\quad (i,j)=((1,1),(1,2),...,(1,n_1),(2,1),(2,2),...,(2,n_2)) \tag{6.36}$$

with relationships (2) and (3). The control input is given by

$$\begin{aligned}&\beta(t)U(t)=q_{1,0}(t),\quad 0\leq\beta\leq 1,\\&(1-\beta(t))U(t)=q_{2,0}(t)\end{aligned} \tag{6.37}$$

The flow U(t) is measured as a function of time, and the splitting rate $\beta(t)$ is the control input. The output measurement could be the full state vector, i.e., vector of flows of all the sections, or a subset of that. The control problem can be stated as: find $\beta_o(t)$, the optimal $\beta(t)$, which minimizes

$$J(\beta)=\int_0^{t_f}[\sum_{i=1}^{m}\chi(\rho_i)-\sum_{m+1}^{m+p}\chi(\rho_j)]^2dt \tag{6.38}$$

where $\chi(.,.)$ is the travel time function and t_f is the final time. Note that, just like for the distributed parameter system case, a feedback solution is needed for the problem, and not an open loop optimal control. Hence, here also, we can either decide the structure of the feedback control, such as a PID control with constant gains, and solve numerically for the optimal values of the gains, or we can state the control objective for a standard feedback control problem, such as steady state asymptotic stability

$$Lt_{t\to\infty}[\sum_{i=1}^{m}\chi(\rho_i)-\sum_{m+1}^{m+p}\chi(\rho_j)]\to 0 \tag{6.39}$$

and some transient behavior characteristics like some specified settling time, percent overshoot, etc.

The discrete time versions of (6.36-6.38) are given below.

$$\rho_{i,j}(k+1)=\rho_{i,j}(k)+\frac{T}{\delta_i}[q_{i,j-1}(t)-q_{i,j}(t)],\quad (i,j)=((1,1),...,(1,n_1),(2,1),...,(2,n_2)) \tag{6.40}$$

$$\beta(k)U(k)=q_{1,0}(k),\quad 0\le\beta\le 1,$$

$$(1-\beta(k))U(k)=q_{2,0}(k), \tag{6.41}$$

$$J(\beta)=\sum_{k=1}^{k_f}[\sum_{i=1}^{m}\chi(\rho_i)-\sum_{m+1}^{m+p}\chi(\rho_j)]^2, \tag{6.42}$$

$$Lt_{k_f\to\infty}[\sum_{i=1}^{m}\chi(\rho_i)-\sum_{m+1}^{m+p}\chi(\rho_j)]\to 0. \tag{6.43}$$

For an n alternate route problem, the continuous time formulation is:

Problem 6.6.1a: Find β_o^i, i=1,2,...,n, which minimize

$$J(\beta_0^i, i=1,2,..n)=\int_0^{t_f}[\sum_{k,p}\{\sum_{i=1}^{n_k}\chi(\rho_{i,k})-\sum_{j=1}^{n_p}\chi(\rho_{j,p})\}^2 dt \tag{6.44}$$

(k=1,2,...,n, p=1,2,...,n, and the summations are taken over total number of combinations of n and p, and not permutations so that (k,p)=(1,2) is considered the same as (k,p)=(2,1), and hence only one of these two will be in the summation), or which guarantee

$$Lt_{t\to\infty}\mathbf{e}\to\mathbf{0}, \tag{6.45}$$

$$\text{where } \mathbf{e}=[\{\sum_{i=1}^{n_1}\chi(\rho_{i,1})-\sum_{j=1}^{n_2}\chi(\rho_{j,2})\},...,\{\sum_{i=1}^{n_k}\chi(\rho_{i,k})-\sum_{j=1}^{n_p}\chi(\rho_{j,p})\},...]$$

with some transient behavior characteristics like some specified settling time, percent overshoot, etc. for the system

$$\frac{d}{dt}\rho_{i,j}=\frac{1}{\delta_i}[q_{i,j-1}(t)-q_{i,j}(t)],\quad (i,j)=((1,1),...,(1,n_1),(2,1),...,(2,n_2),...,(n,1),...,(n,n_n)) \tag{6.46}$$

with given full and partial state observation, and input constraints

$$\sum_{i=1}^{n} q_{i,0}(t) = U(t) \text{ and } \sum_{i=1}^{n} \beta^i = 1. \tag{6.47}$$

The corresponding discrete time formulation is:

Problem 6.6.1b: Find β_o^i, i=1,2,...,n, which minimize

$$J(\beta_0^i, i = 1,2,..n) = \sum_{k=1}^{k_f} \sum_{k,p} \{\sum_{i=1}^{n_k} \chi(\rho_{i,k}) - \sum_{j=1}^{n_p} \chi(\rho_{j,p})\}^2 \tag{6.48}$$

(k=1,2,...,n, p=1,2,...,n,)and the summations are taken over total number of combinations of n and p, or which guarantee

$$Lt_{k_f \to \infty} \mathbf{e} \to \mathbf{0}, \tag{6.49}$$

$$\text{where } \mathbf{e} = [\{\sum_{i=1}^{n_1} \chi(\rho_{i,1}) - \sum_{j=1}^{n_2} \chi(\rho_{j,2})\},\dots,\{\sum_{i=1}^{n_k} \chi(\rho_{i,k}) - \sum_{j=1}^{n_p} \chi(\rho_{j,p})\},\dots]$$

with some transient behavior characteristics like some specified settling time, percent overshoot, etc. for the system

$$\rho_{i,j}(k+1) = \rho_{i,j}(k) + \frac{T}{\delta_i}[q_{i,j-1}(k) - q_{i,j}(k)], \quad (i,j) = ((1,n_1),(2,n_2)\dots,(n,n_n)) \tag{6.50}$$

with given full and partial state observation, and input constraints (6.47) after replacing t with k. Based on these standard forms, the appropriate feedback control laws can be designed. The exact methodology of the design will be studied in the future, but the next section elucidates an example to show how the feedback control could be used.

6.7 Sample Problem for Space and Time Discretized Dynamics - Three Alternate Routes Case Description

In this section we show a design of a feedback control system for a simple network consisting of three alternate routes modeled as space and time discretized system (6.40). Similar feedback controllers can be designed for larger and more complex traffic networks. For simplicity, we are assuming that each of the alternate route is just a single discrete section, and also for simplicity we are ignoring the diffusion term in the model.

Although the traffic network presented in this section seems to be simplistic, it represents one of the most realistic settings for real-time traffic routing. The main goal of traffic routing for today's applications is to alleviate non-recurrent traffic congestion on a freeway by diverting freeway traffic to alternate routes. Unfortunately for most of the urban traffic networks, the number of feasible alternate routes is not more than three. Moreover, due to the difficulties related to the real-time control of traffic at multiple locations, point diversion appears to be a practical near-term alternative to the network wide routing where more advanced traffic control as well as more complex dynamic routing algorithms might be needed. Thus, the sample problem presented in this paper describes a very realistic scenario for demonstrating the feasibility and advantages of fuzzy feedback control for dynamic traffic routing. On the other hand, it is quite clear that the design of more advanced controllers that can tackle more complex networks and scenarios will be needed for future applications.

An incident that occurs on a freeway can be used as a perfect scenario to illustrate the realism of the sample problem. Now, let us assume that as a result of the incident, part of freeway traffic will be diverted to the alternate highways. Diversion will then be initiated at a point before the incident location and the traffic will be diverted back to the major freeway at a point past the incident. Every day occurrences of incident situations similar to the above example are abundant in Northern Virginia and other parts of the country. Given the real-time traffic control capabilities of most of the DOTs, a system that will regulate the point diversion using a widely used actuation method such as variable message signs appears to be a viable solution.

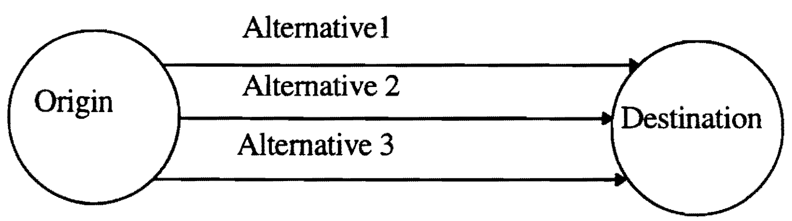


Figure 6.5:Test Network

6.7.1 System Dynamics

The flow equations used for the three alternate routes are:

$$\rho_1(k+1)=\rho_1(k)-\frac{T}{\delta_1}[q_1^{out}(k)-q_1^{in}(k)],$$

$$\rho_2(k+1)=\rho_2(k)-\frac{T}{\delta_2}[q_2^{out}(k)-q_2^{in}(k)],$$

$$\rho_3(k+1)=\rho_3(k)-\frac{T}{\delta_3}[q_3^{out}(k)-q_3^{in}(k)], \qquad (6.65)$$

where the flows are taken to be

$$q_1^{out}(k) = \rho_1(k)v_1(k), \tag{6.66}$$
$$q_2^{out}(k) = \rho_2(k)v_2(k),$$
$$q_3^{out}(k) = \rho_3(k)v_3(k),$$

the average velocities are given by

$$v_1 = v_{f1}(1 - \frac{\rho_1}{\rho_{max1}}), \tag{6.67}$$
$$v_2 = v_{f2}(1 - \frac{\rho_2}{\rho_{max2}}),$$
$$v_3 = v_{f3}(1 - \frac{\rho_3}{\rho_{max3}}),$$

and

$$q_1^{in}(k) = \beta_1(k)U(k), \qquad 0 \le \beta_1(k) \le 1, \tag{6.68}$$
$$q_2^{in}(k) = \beta_2(k)U(k), \qquad 0 \le \beta_2(k) \le 1,$$
$$q_3^{in}(k) = [1 - \beta_1(k) - \beta_2(k)]U(k)$$

The travel times are given by

$$\chi_1(k) = d_1 / [v_{f1}(1 - \frac{\rho_1}{\rho_{max1}})],$$
$$\chi_2(k) = d_2 / [v_{f2}(1 - \frac{\rho_2}{\rho_{max2}})],$$
$$\chi_3(k) = d_3 / [v_{f3}(1 - \frac{\rho_3}{\rho_{max3}})]. \tag{6.69}$$

The state variables are the ρ's for each route. This is a full state measurement problem, assuming the flows are measured, and can be converted into state variable values using the deterministic relationships. The control variables are the splitting rates. For this example, full compliance is assumed. Future work will address the design of controllers for more complete models with partial, time-varying compliances. The overall system in a standard nonlinear state-space form can be written as

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{g}(\mathbf{x}(k))\mathbf{u}(k),$$

$$\mathbf{y}(k) = \mathbf{h}(\mathbf{x}(k)) \tag{6.70}$$

where

$$\mathbf{x}(k) = [\rho_1(k), \rho_2(k), \rho_3(k)]^T, \quad \mathbf{y}(k) = [\chi_1(k), \chi_2(k), \chi_3(k)]^T, \quad \mathbf{u}(k) = [\beta_1, \beta_2]^T,$$

$$\mathbf{f}(\mathbf{x}(k)) = \begin{bmatrix} \rho_1(k) - T\rho_1(k)v_{f1}(1-\rho_1(k)/\rho_{max1})/\delta_1 \\ \rho_2(k) - T\rho_2(k)v_{f2}(1-\rho_2(k)/\rho_{max2})/\delta_2 \\ \rho_3(k) - T\rho_3(k)v_{f3}(1-\rho_3(k)/\rho_{max3})/\delta_3 + U(k) \end{bmatrix},$$

$$\mathbf{g}(\mathbf{x}(k)) = \begin{bmatrix} TU(k)/\delta_1(k) & 0 \\ 0 & TU(k)/\delta_2(k) \\ -TU(k)/\delta_1(k) & -TU(k)/\delta_2(k) \end{bmatrix}, \quad \mathbf{h}(\mathbf{x}(k)) = \begin{bmatrix} d_1/[v_{f1}(1-\frac{\rho_1(k)}{\rho_{max1}})] \\ d_2/[v_{f2}(1-\frac{\rho_2(k)}{\rho_{max2}})] \\ d_3/[v_{f3}(1-\frac{\rho_3(k)}{\rho_{max3}})] \end{bmatrix} \tag{6.71}$$

Here the system is in the local coordinates for the smooth state space manifold M, **f** is the smooth drift vector field on M, **g** is the smooth input vector field on M, and **h** is the smooth output vector field on M.

6.7.2 System Analysis

The nonlinear system (6.65) can be analyzed for controllability and observability. For continuous nonlinear systems, controllability and observability can be studied using Lie brackets and Lie derivatives in terms of accessibility and reachability. These techniques can be extended to be used for discrete system (6.65). For review of these topics please see references (Isidori, 1989, Slotine, 1991, Nijmeijer, 1990). A detailed analysis of this system using feedback linearization technique is given in the next chapter. For this example, we might conjecture that the system is accessible and reachable.

6.7.3 Simple Feedback Control Law

A heuristic controller for (6.65) is designed as a PI (Proportional-Integral) controller with constant feedforward term. As was discussed in section 6.6, there are two ways to design a feedback controller for this problem: one is to solve the optimality and use the feedback solution if available for that problem; the other is to design a feedback controller to satisfy some steady state asymptotic stability or transient behavior. This controller is designed to produce asymptotic stability to the system, so that the error terms defined next tend to zero in time. Since many controllers including LQ are implemented as constant gain (static) controllers, we chose to highlight a design with a constant gain PI controller. The error terms to drive the controller are defined as

$$e_1(k) = \chi_3(k) - \chi_1(k),$$
$$e_2(k) = \chi_3(k) - \chi_2(k). \tag{6.72}$$

The integral term in the PI control is taken as error summation, so that

$$ie_1(k) = \sum_{p=1}^{k} \{\chi_3(p) - \chi_1(p)\},$$
$$ie_2(k) = \sum_{p=1}^{k} \{\chi_3(p) - \chi_2(p)\}. \tag{6.73}$$

The PI controller with constant feedforward shown below has been designed to equilibrate travel times on alternate routes. While integral part of the PI controller affects the steady-state behavior of the system, the proportional part of it combined with the integral part affects the transient behavior. This controller is presented here to show the feasibility of feedback control to meet the objective of achieving equal travel time in all the alternate routes.

$$\beta_1(k) = \max[0, \min\{1, (1/3 + k_1 e_1 + k_{i1} ie_1)\}],$$
$$\beta_2(k) = \max[0, \min\{1 - \beta_1(k), (1/3 + k_2 e_2 + k_{i2} ie_2)\}], \tag{6.74}$$

where $k_1, k_{i1}, k_2, \text{and}\, k_{i2}$ are the controller gains.

As can be seen in section 6.7.4, the results of using this controller are encouraging. However, using PI regulators with input saturation suffers from wind-up phenomenon, which could be countered by using the following form for the control laws.

$$\beta(k) = \beta(k-1) - k_1 e_1(k) - k_2[e_1(k) - e_1(k-1)] \qquad (6.75)$$

where $\beta(k)$ is limited in the range [0,1] and $\beta(k-1)$ takes its actually implemented value (e.g. 0 or 1 if it has been on the bounds). The PI controller described above is shown to be stable using simulations. The results are presented and discussed in the next section.

6.7.4 Description of the Results for Different Scenarios

For the simple example problem and its feedback control solution, we have performed several test runs for different scenarios. The simulation program is developed in Matlab. We have tested the simulation using three scenarios. For each scenario we have considered a simple network that consists of three alternate routes. The splitting decisions are made at one decision point only. Alternate routes have different free flow travel times. The one with the lowest travel time can be assumed to be the freeway and two others with higher travel times can be considered to be highways with lower level of service.

Brief description of each scenario and the corresponding plots are shown below. Each simulation shows two plots showing travel times and splits versus time for all the three routes. The three routes are shown with three different line styles: solid, dashed, and dot-dashed.

First Scenario:

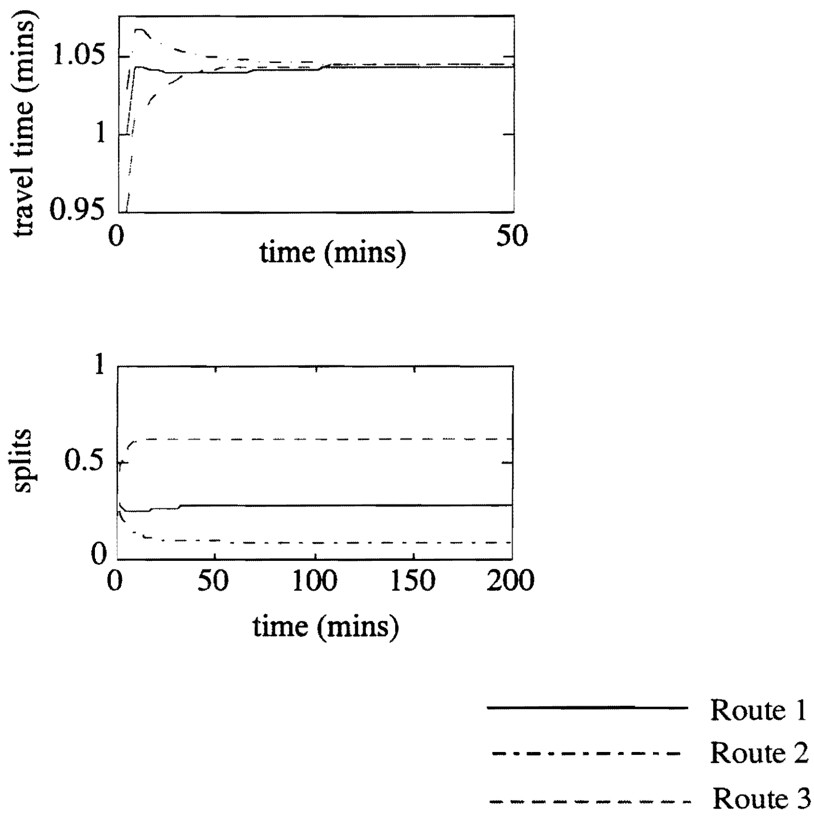


Figure 6.6: Scenario 1

In this scenario, we have constant inflow and no congestion for all the alternate routes. This scenario represents normal traffic conditions. As it can be seen in the first plot, travel times become equal and stay equal until the end of the simulation period since there are no external disturbances.

Second Scenario:

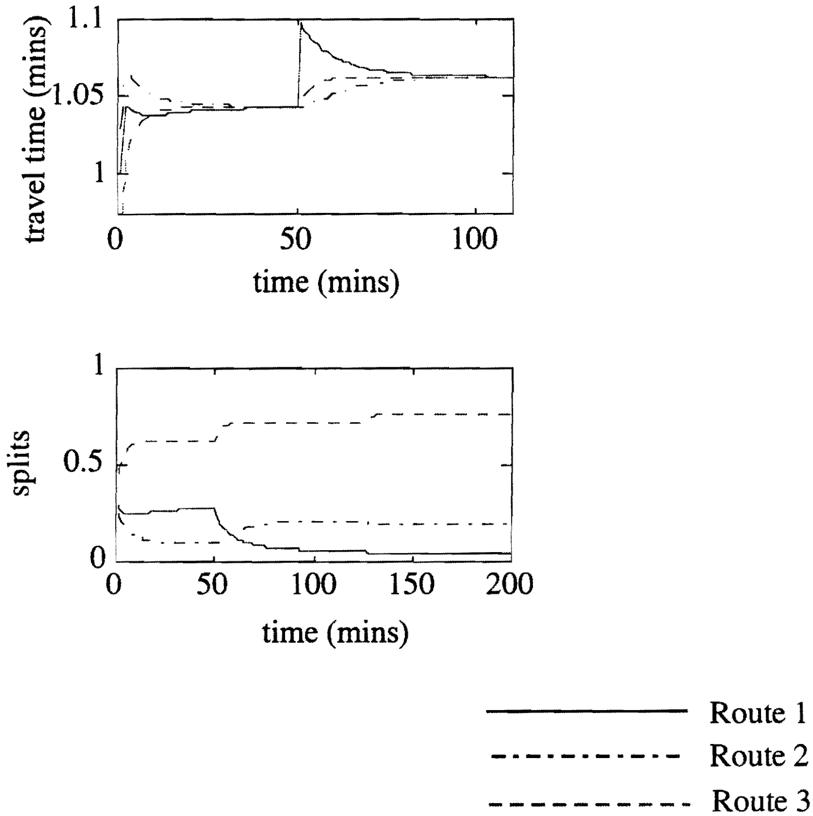


Figure 6.7: Scenario 2

This scenario simulates congestion on route 2 from time 50 to 75, and relief on route 3 from time 125 to 150. The congestion on route 2 may be due to a temporal bottleneck caused by an incident and the relief on route 3 may be due to the clearance of an incident which existed before. Fluctuations in travel times and the response of the controller in order to stabilize the system can be easily seen by studying the plots.

Scenario 3:

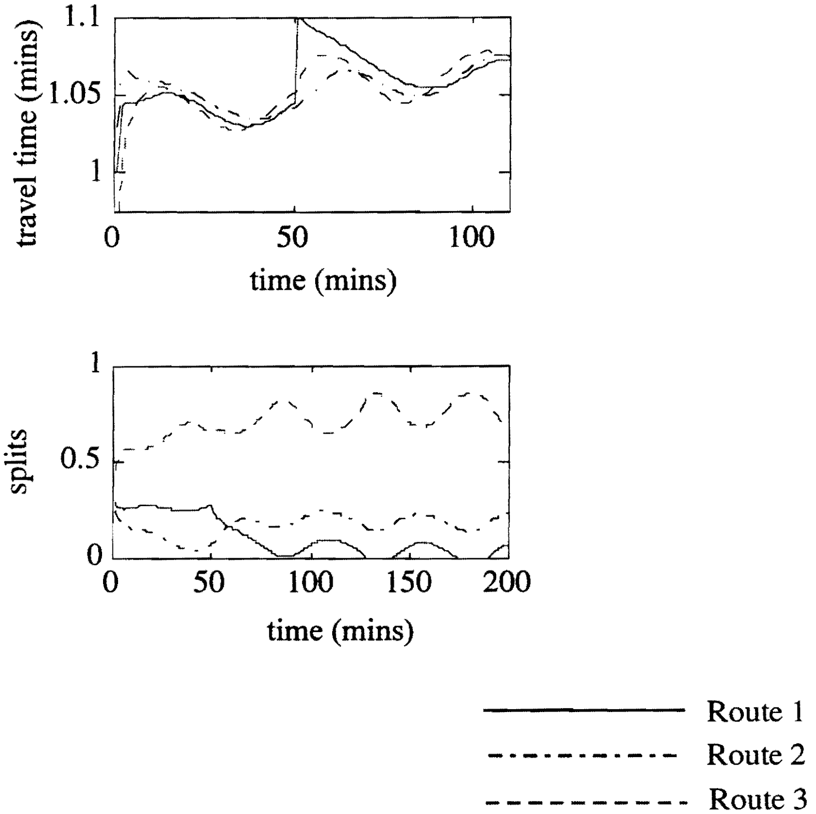


Figure 6.8: Scenario 3

This last scenario has a sinusoidal inflow (demand) function and same traffic patterns as scenario 2. The fluctuations in inflow traffic modeled by the sinusoidal are meant to represent the natural traffic fluctuations that occur in reality. In a fluctuation traffic scenario simulation also, the controller successfully achieves the objective of equating the travel times on the three alternate routes.

6.8 Fuzzy Feedback Control for DTR

In this section, we show the design of a qualitative controller using fuzzy logic. Figure 6.9 shows the block diagram for the fuzz feedback control. Fuzzy logic can be used effectively to deal with uncertainty in decision making processes. Our goal for using the fuzzy logic for designing our controller is to show the applicability of fuzzy logic to DTA / DTR problems. For complex DTA / DTR problems fuzzy controller design can be a very useful tool due to its qualitative nature which facilitates the modeling process. Therefore, for problems where the design of an analytical controller is very difficult or not possible, fuzzy control can be a viable alternative.

Fuzzy control is based on the fuzzy set theory proposed by Zadeh (1965, 1975, 1985). There are three major ways to design fuzzy controllers. In the first method, the controller tries to emulate a human-like control action by transforming linguistic terms into fuzzy variables (Mamdani, 1976, Mamdani et al., 1983, Kickert et al., 1976, Jain et al., 1987). The second method is to develop heuristic based fuzzy controllers. In the third method, the traffic network is represented as a fuzzy system and a control is designed by analyzing the fuzzy model. In this section, we employ the second method.

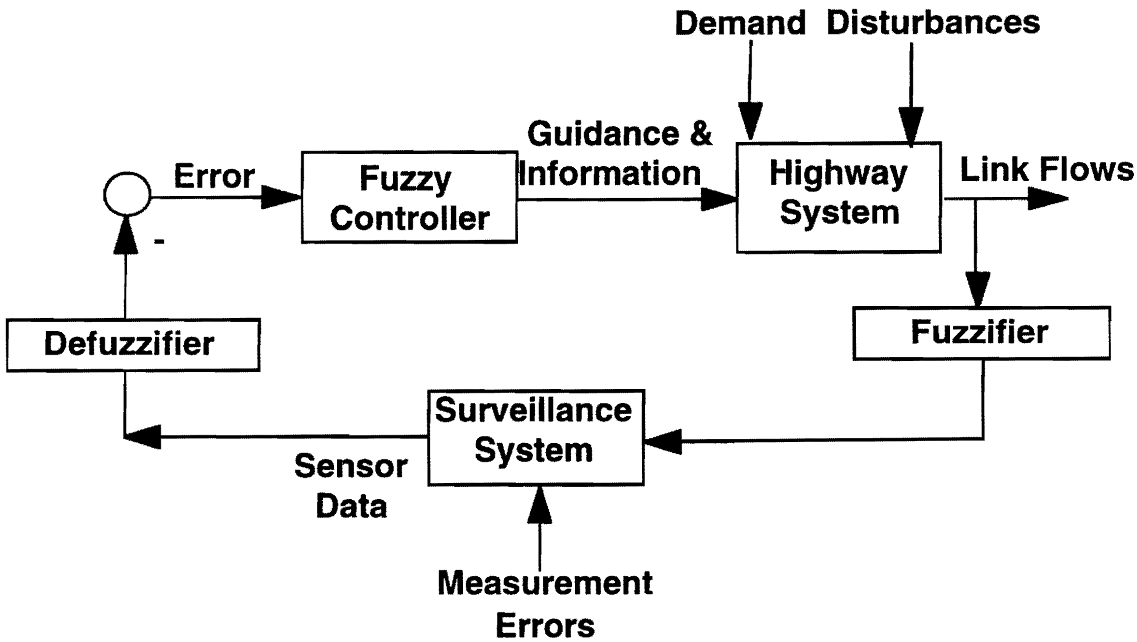


Figure 6.9: Block Diagram for DTR Fuzzy Feedback Control

Recently, fuzzy set theory has been applied to several challenging transportation problems. Lotan and Koutsopulos (1993a, 1993b) presented a modeling framework for route choice in the presence of information based on the concepts from fuzzy set theory, approximate reasoning and fuzzy control. The proposed framework included models for driver's perception of network attributes, attractiveness of alternative routes, as well as models for reaction to information and the route choice mechanism itself. Sasaki and Akiyama (1988) presented a fuzzy reasoning model for on-ramp control. The main objective of this paper was to describe the judgment process of the human operator by using fuzzy logic and to develop a traffic control system for automatic on-ramp control of the expressway. The actual performance of the fuzzy control system was then studied on the Osaka-Sakai route.

Chen, May and Auslander (1990) presented a fuzzy controller for freeway ramp metering on the San Francisco-Oakland Bay Bridge. In order to compare the performance of the fuzzy controller with the performance of the existing ramp controller and an idealized controller, 2 scenarios that contain no incidents and eight scenarios that contain an incident of varying severity and location were employed. In general, the fuzzy controller rapidly and smoothly responded to the incidents by significantly reducing total delays caused by incidents.

Fuzzy logic is essentially a function approximation, and hence, feedback fuzzy control represents the control input as a function of output and time as

$$\mathbf{u}(t) = \hbar(\mathbf{y}(t), t) \tag{6.75}$$

for continuous time feedback, and

$$\mathbf{u}(k) = \lambda\!\!\!^{-}(\mathbf{y}(k), k) \tag{6.76}$$

for discrete time feedback control. In order to apply fuzzy control, we need to convert the output $\mathbf{y}$ into fuzzy variable through fuzzification process, then utilize the fuzzy control to produce the fuzzy control input $\mathbf{u}$, which then can be applied after its conversion to crisp number through defuzzification process. Most of the following discussion is adapted from Mamdani (1995), which is a review paper for application of fuzzy logic for engineering.

The fuzzy logic deals with fuzzy variables, which can be derived from crisp numbers using membership functions m(x). For instance, the variable traffic flow could have membership functions such as little, small, medium, large, and big. There are various types of membership functions proposed, the most common ones being triangular.

In order to manipulate the fuzzy variables, we will use the following set theoretic operations.

$$\mu_{(A \cup B)}(x) = \max[\mu_A(x), \mu_B(x)] \qquad (6.77)$$
$$\mu_{(A \cap B)}(x) = \min[\mu_A(x), \mu_B(x)]$$
$$\mu_{(A')}(x) = 1 - \mu_A(x)$$

We also choose

$$\mu_{p \to q}(x, y) = \min[\mu_p(x), \mu_q(y)] \qquad (6.78)$$
$$\text{or } \mu_{p \to q}(x, y) = \mu_p(x)\mu_q(y)$$

as engineering implications as compared to the traditional prepositional logic implications. The building block of the Fuzzy Logic Inference Engine (FLIE) is Modus Ponens, which states:

Premise1:	x is A
Premise 2:	IF x is A THEN y is B
Consequence:	y is B

This can also be expressed as: $(p \cap (p \to q)) \to q$. In fuzzy logic, this can be constructed by:

$$\mu_{B^*}(y) = \sup_{x \in A^*}[\mu_{A^*}(x) * \mu_{A \to B}(x, y)] \qquad (6.79)$$

In this sup-norm composition, we can use product or minimum as the t-norm for the star composition. Now, if we use a singleton fuzzifier, then function (6.79) becomes

$$\mu_{B^*}(y) = \mu_{A^*}(x') * \mu_{A \to B}(x, y) \qquad (6.80)$$

where $\mu(x') = 1$, at x=x', and zero everywhere else, denoting the use of singleton fuzzification process. Since, the support of A* is x', we obtain

$$\mu_{B^*}(y) = \mu_{A \to B}(x', y) \qquad (6.81)$$

When there are multiple antecedents in a rule, such as, IF u_1 is F_1^ℓ and u_2 is F_2^ℓ and...and u_m is F_m^ℓ, THEN v is G^ℓ, we can write

$$\mu_{A\to B}(\mathbf{x},y) = \mu_{F_1^\ell}(x_1) * \mu_{F_2^\ell}(x_2) * \ldots * \mu_{F_m^\ell}(x_m) \tag{6.82}$$

In order to connect the fuzzy rules, t-conorm, $\oplus$ will be used. An effective t-conorm is the max operation. Hence, the fuzzy output of the FLIE, in which there are p rules, is

$$\mu_{R1}(\mathbf{x},y) \oplus \mu_{R2}(\mathbf{x},y) \oplus \ldots \oplus \mu_{Rp}(\mathbf{x},y) \tag{6.83}$$

We use the singleton fuzzifier method for the fuzzification. A fuzzy singleton has support x_a, i.e., $\mu_{A*}(x_A) = 1$, for $x = x_A$ and $\mu_{A*}(x_A) = 0, \forall\, x \in U, (x \neq x_A)$. Gaussian and triangular membership functions can also be used for the fuzzification process.

There are many ways of performing defuzzification. The following are some of the methods of defuzzification which are adapted from Mendel (1995):

(1) Mean of Maximum Defuzzifier: In this method, the output of the defuzzifier acting on a variable x of a fuzzy set A is the mean of those crisp values of x, which give the maximum value of $\mu_A(x)$.

(2) Centroidal Method: In this method, the output of the defuzzifier acting on a variable x of a fuzzy set A is that crisp value of x, at which the centroid of the area under the membership function curve exists.

6.8.1 Sample Problem - Three Alternate Routes Case Description

In order to illustrate the ideas discussed above, we have designed a fuzzy feedback control system for a simple test network consisting of three alternate routes (Figure 6.10). This example is identical to the example studied in the previous section except the controller designed in this section is a qualitative controller that uses fuzzy logic. Similar feedback controllers can be designed for larger and more complex traffic networks. For simplicity, we are assuming that each of the alternate route is just a single discrete section.

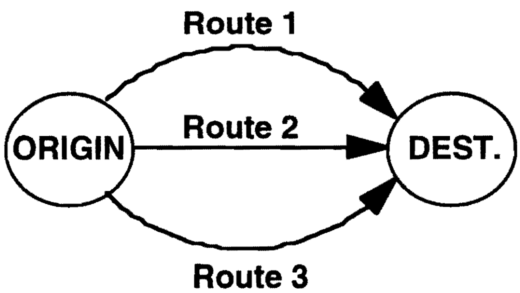


Figure 6.10: Test Network with 3 Alternate Routes

6.8.1.1 System Dynamics

The system dynamics for this sample problem is given in section 6.7.1. The state variables are the ρ' s for each route. This is considered a full state measurement problem, assuming the flows are measured, which can be converted into state variable values using the deterministic relationships. The control variables are the splitting rates. For this example, full compliance of the traffic flow to these split rates is assumed. This assumes that some technique of making the vehicles follow the split rate, such as Variable Message Signs (VMS) or in-vehicle communication is employed The overall system in a standard nonlinear state-space form can be written as shown in (6.71).

6.8.1.2 System Analysis

As it is mentionned in section 6.7, the nonlinear system (6.65) can be analyzed for controllability and observability. However, for the sake of simplicity in this section, we again assume the system to be accessible and reachable.

6.8.1.3 Simple Fuzzy Feedback Control Law

Although the dynamics of the system are described in the previous section, it is difficult to design analytic controllers which are robust to perturbations to a nonlinear plant which represent the real physical system. Fuzzy control provides a design methodology which might prove effective in design of robust controllers for such systems. The details of the fuzzy controller used for this example are shown below.

The error terms to drive the controller are defined as

$$\begin{aligned} e_1(k) &= \chi_3(k) - \chi_1(k) \\ e_2(k) &= \chi_3(k) - \chi_2(k) \end{aligned} \tag{6.84}$$

The integral term in the fuzzy controller is taken as error summation, so that

$$\begin{aligned} ie_1(k) &= \sum_{p=1}^{k} \{\chi_3(p) - \chi_1(p)\} \\ ie_2(k) &= \sum_{p=1}^{k} \{\chi_3(p) - \chi_2(p)\} \end{aligned} \tag{6.85}$$

where p is the dummy time variable used for summation for discrete integration.

The fuzzy control we designed emulates the concept of PI controller. The structure of a PI controller with constant feedforward is shown below. We use fuzzy control, which would try to equilibrate travel times on alternate routes, to perform all the functions of feedback, feedforward, and the saturation. This controller is presented here to show the feasibility of fuzzy feedback control to meet the objective of achieving equal travel time in all the alternate routes, and in general to demonstrate the applicability of fuzzy feedback control to DTR and in general to DTA.

$$\beta_1(k) = \max[0, \min\{1, (1/3 + k_1 e_1 + k_{i1} ie_1)\}] \tag{6.86}$$
$$\beta_2(k) = \max[0, \min\{1 - \beta_1(k), (1/3 + k_2 e_2 + k_{i2} ie_2)\}]$$

where $k_1, k_{i1}, k_2, \text{and } k_{i2}$ are the controller gains.

There are seven membership functions each for $e_1, e_2, ie_1, \text{and } ie_2$, namely, negative high, negative medium, negative low, zero, positive low, positive medium, and positive high (Figures 6.11).

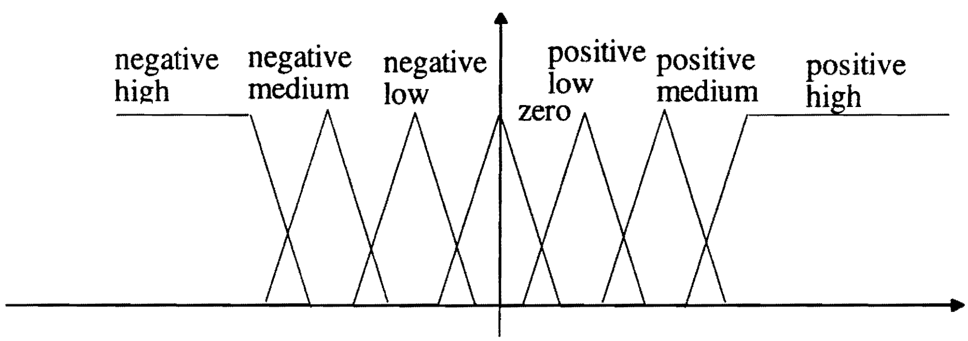


Figure 6.11: Error Membership Functions

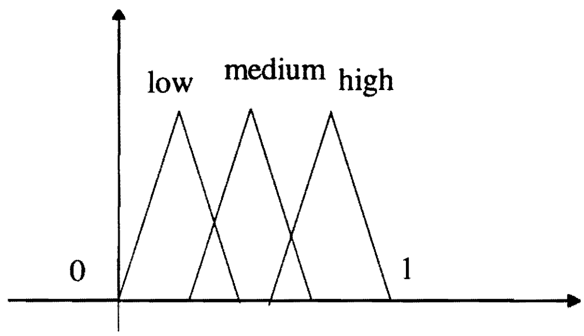


Figure 6.12: Membership Functions for β_1 and β_2.

The support of each triangular membership function is divided into three equal parts. The triangles intersect one third at the bases of each others as shown in Figure 6.12. By putting this constraint, and the constraint that all the membership functions have the same magnitude of angles with the base, leave two degrees of freedom to divide the number line into membership functions. The two degrees of freedom are the magnitude of the base angle of the membership functions 'θ', and the number of membership functions 'n'. For the simulation purposes, we have $\theta = 1.24$ radians and n=7.

For β_1 and β_2, we have three membership functions each, low, medium and high, as shown in Figure 6.12. Here, we use isosceles triangular membership functions covering the entire universe of discourse which is the closed set [0, 1].

For this example, we have used 98 rules utilizing all the possible combinations of fuzzy values of the errors. The rules dealing with e_1 and ie_1 are 49 (7 X 7), and similarly rules dealing with e_2 and ie_2 are 49 too, making the total 98. Note that this is a large set of rules for this choice. The reasons for that are, we have not used a default value; there is region in every membership function which has no overlap; and we have used the min function in combining the membership functions for fuzzy processing. The number of rules can be drastically reduced by changing these options. The following are a few representative rules utilized.

IF e_1 is positive low, AND ie_1 is positive low, THEN β_1 is low.

IF e_2 is positive low, AND ie_2 is positive low, THEN β_2 is low.

IF e_1 is negative high, AND ie_1 is negative low, THEN β_1 is low.

IF e_2 is negative high, AND ie_2 is negative low, THEN β_2 is low.

IF e_1 is positive high, AND ie_1 is negative low, THEN β_1 is medium.

IF e_2 is positive high, AND ie_2 is negative low, THEN β_2 is medium.

IF e_1 is positive high, AND ie_1 is negative medium, THEN β_1 is low.

IF e_2 is positive high, AND ie_2 is negative medium, THEN β_2 is low.

IF e_1 is positive medium, AND ie_1 is positive medium, THEN β_1 is medium.

IF e_2 is positive medium, AND ie_2 is positive medium, THEN β_2 is medium.

We use singleton fuzzification, and centroid defuzzification for crisp-fuzzy and fuzzy-crisp conversions in this example.

6.8.1.4 Results & Description For Different Scenarios

For the simple example problem and its feedback control solution, we have performed several test runs for different scenarios. The simulation program is developed in Matlab. We have tested the simulation using three scenarios. For each scenario, we have considered a simple network that consists of three alternate routes. The splitting decisions are made at one decision point only. Alternate routes have different free flow travel times. The one with the lowest travel time can be assumed to be the freeway and two others with higher travel times can be considered to be highways with lower level of service. Brief description of each scenario and the corresponding plots are shown below.

In this scenario, we have constant inflow and no congestion for all the alternate routes. This scenario represents normal traffic conditions (Figure 6.13). As it can be seen in the first plot, travel times become equal with some steady state error and some steady state oscillations. These errors can be further reduced by retuning the fuzzy control in order to increase the resolution.

First Scenario

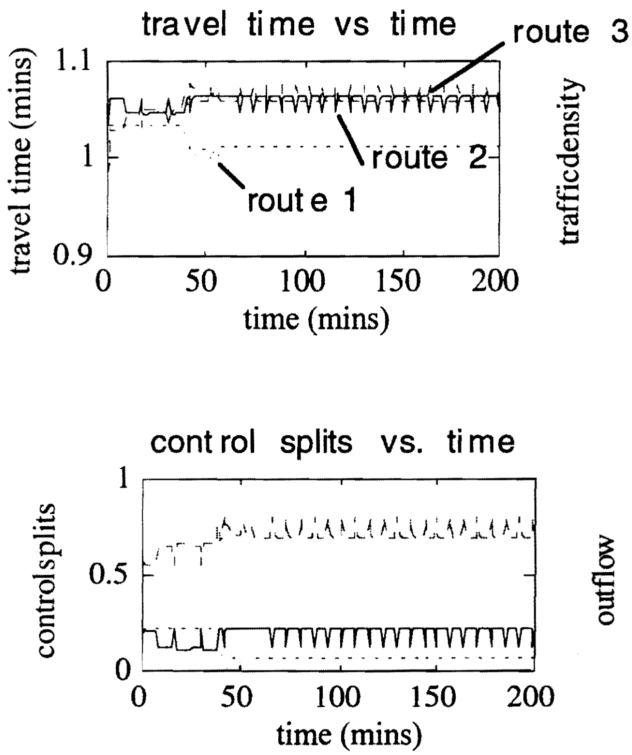


Figure 6.13: Scenario 1

Second Scenario

This scenario simulates congestion on route 2 from time 50 to 75, and relief on route 3 from time 125 to 150 (Figure 6.14). The congestion on route 2 may be due to a temporal bottleneck caused by an incident and the relief on route 3 may be due to the clearance of an incident which existed before. This fuzzy control is not very responsive to the change in congestion level. It can also be improved by further tuning of the control parameters.

Third Scenario

This last scenario has a sinusoidal inflow (demand) function and same traffic patterns as scenario 2 (Figure 6.15). The fluctuations in inflow traffic modeled by the sinusoidal are

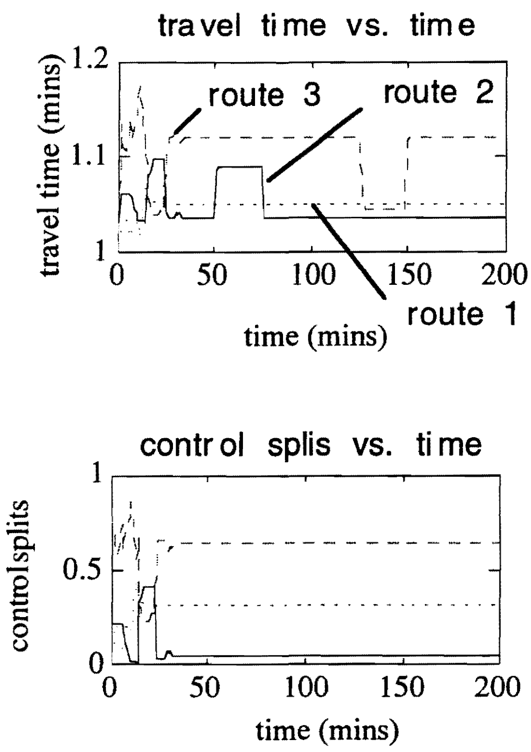


Figure 6.14: Scenario 2

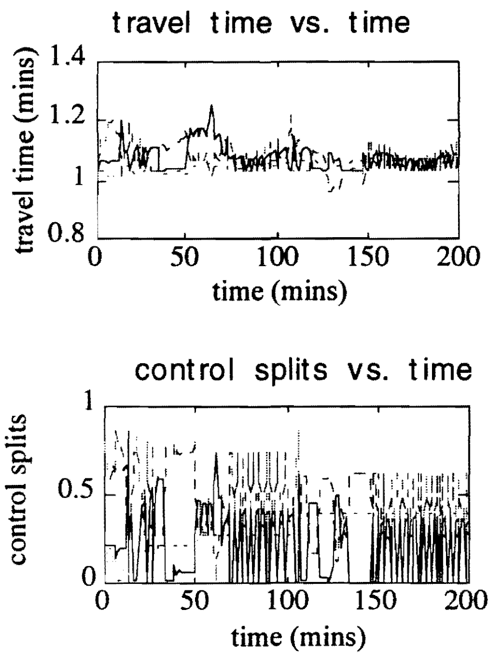


Figure 6.15: Scenario 3

meant to represent the natural traffic fluctuations. Here, the steady state errors are reduced, but there is high control activity, which can be filtered out.

6.9 Conclusions

In this Chapter, we have formulated feedback control problems in the distributed parameter setting as well as in continuous-time and discrete-time lumped parameter setting for DTR. Brief discussions on applicable real-time feedback control methods were provided. In the simulation example, we have shown that feedback control is viable and attractive solution to the on-line dynamic traffic control/routing problem. In order to run the feedback control, we obtained desired states that the controller will track from the traffic sensors. In this control problem, the travel time to be tracked for each route was the travel time of the next route. This solution worked nicely for this simple problem under different demand and traffic conditions. We can try to generalize this concept for more complex networks but, more research needs to be done in this area in order to design for these states to be tracked by the controller.

Finally, as a qualitative method for designing a controller, fuzzy logic is chosen. First, a discussion on applicable fuzzy feedback control methods was provided. A simple but illustrative software simulation example was also shown. In the simulation example, we have shown that fuzzy feedback control is viable and attractive solution to the on-line dynamic traffic control/routing problem. In this control problem too, the travel time to be tracked for each route was the travel time of the next route. This solution also worked nicely for this simple problem under different demand and traffic conditions. This concept also can be generalized for more complex networks.

This chapter should provide an incentive to further the development of feedback control design and application for traffic problems, especially DTA and DTR problems. Feedback control models, by making use of the sensor data and minimizing the need for expensive computational requirements of classical optimization techniques, provides a viable alternative in the context of ITS. Since, most of the routing decisions are local and not network-wide, the solution for the type of DTR problem discussed in this chapetr can be promising under incident conditions which do not require network-wide approaches.

7. Solution to the User Equilibrium Dynamic Traffic Routing / Control Problem using Feedback Linearization

7.1 Feedback Linearization Technique

In the discretized traffic flow model described in section 6.5.2 of Chapter 6, the freeway is divided into sections with aggregate traffic densities. Sensors are used to measure variables such as densities, traffic flow and traffic average speeds in these sections, which can be used by the feedback controller to give appropriate commands to actuators like VMS, HAR, etc. In the following sections, three different controllers for three different versions of the same user-equilibrium formulation of the DTR problem for the discretized traffic flow model given in section 6.7.1 are developed using the feedback linearization technique.

Feedback linearizaton is an appropriate technique for developing feedback controllers for nonlinear systems similar to the DTR model described above. The feedback linearization technique is applicable to an input affine square multiple input multiple output (MIMO), system. The details on exact nonlinear decoupling technique (feedback linearization) can be found in Isidori (1989), Slotine (1991), and Godbole et al. (1995), and is briefly summarized here for the DTR application. Let us consider the following square MIMO system:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + \sum_{i=1}^{p} g_i(\mathbf{x})u_i \qquad (7.1)$$

$$y_j = h_j(\mathbf{x}) \quad j = 1,2,\ldots,p$$

This can be written in a compact form as

$$\Sigma: \begin{array}{c} \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\mathbf{u} \\ \mathbf{y} = \mathbf{h}(\mathbf{x}) \end{array} \qquad (7.2)$$

where, $\mathbf{x} \in R^n, \mathbf{f}(\mathbf{x}): R^n \rightarrow R^n$, $\mathbf{g}(\mathbf{x}): R^p \rightarrow R^n$, $\mathbf{u} \in R^p$, and $\mathbf{y} \in R^p$. The vector fields of $\mathbf{f}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are analytic functions.

It is assumed that for the system Σ, each output y_j has a defined relative degree γ_j. The concept of relative degree implies that if the output is differentiated with respect to time γ_j times, then the control input appears in the equation. This can be succinctly represented using Lie derivatives. Definition of a Lie derivative is given below, after which the definition of relative degree in terms of Lie derivatives is stated.

Definition (Lie Derivative): Lie derivative of a smooth scalar function $h: R^n \rightarrow R$ with respect to a smooth vector field $\mathbf{f}: R^n \rightarrow R^n$ is given by $L_{\mathbf{f}}h = \frac{\partial h}{\partial \mathbf{x}}\mathbf{f}$. Here, $L_{\mathbf{f}}h$ denotes the Lie derivative of order zero. Higher order Lie derivatives are given by $L_{\mathbf{f}}^{i}h = L_{\mathbf{f}}(L_{\mathbf{f}}^{i-1}h)$.

Definition (Relative Degree): The output y_j of the system Σ has a relative degree γ_j if, $\exists$ an integer, s.t. $L_{g_i}L_{\mathbf{f}}^{\ell}h(\mathbf{x}) \equiv 0 \quad \forall \ell < \gamma_j - 1, \forall 1 \le i \le p, \forall x \in U$, and $L_{g_i}L_{\mathbf{f}}^{\gamma_j - 1}h(\mathbf{x}) \neq 0$. $U \subset R^n$ which is in a given neighborhood of the equilibrium point of the system Σ. The total relative degree of the system r, is defined to be the sum of the relative degrees of all the output variables, i.e., $r = \sum_{j=1}^{p} \gamma_j$.

By successively taking the Lie derivatives of each of the output variables up to their respective relative degrees, we obtain

$$\begin{bmatrix} y_1^{\gamma_1} \\ y_2^{\gamma_2} \\ \cdot \\ \cdot \\ \cdot \\ y_1^{\gamma_p} \end{bmatrix} = \begin{bmatrix} L_f^{\gamma_1} h_1(\mathbf{x}) \\ L_f^{\gamma_2} h_2(\mathbf{x}) \\ \cdot \\ \cdot \\ \cdot \\ L_f^{\gamma_p} h_p(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} L_{g_1} L_f^{\gamma_1 - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_1 - 1} h_1(\mathbf{x}) \\ L_{g_p} L_f^{\gamma_2 - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_2 - 1} h_1(\mathbf{x}) \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ L_{g_1} L_f^{\gamma_p - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_p - 1} h_1(\mathbf{x}) \end{bmatrix} \mathbf{u} \tag{7.3}$$

This can be written as

$$\mathbf{y}^{\gamma} = \mathbf{A}(\mathbf{x}) + \mathbf{B}(\mathbf{x})\mathbf{u} \tag{7.4}$$

where

$$\mathbf{y}^{\gamma} = \begin{bmatrix} y_1^{\gamma_1} & y_2^{\gamma_2} & \ldots & y_1^{\gamma_p} \end{bmatrix}^T \tag{7.5}$$

$$\mathbf{A}(\mathbf{x}) = \begin{bmatrix} L_f^{\gamma_1} h_1(\mathbf{x}) & L_f^{\gamma_2} h_2(\mathbf{x}) & \ldots & L_f^{\gamma_p} h_p(\mathbf{x}) \end{bmatrix}^T \tag{7.6}$$

$$\mathbf{B}(\mathbf{x}) = \begin{bmatrix} L_{g_1} L_f^{\gamma_1 - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_1 - 1} h_1(\mathbf{x}) \\ L_{g_p} L_f^{\gamma_2 - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_2 - 1} h_1(\mathbf{x}) \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ L_{g_1} L_f^{\gamma_p - 1} h_1(\mathbf{x}) & \ldots & L_{g_p} L_f^{\gamma_p - 1} h_1(\mathbf{x}) \end{bmatrix} \tag{7.7}$$

If the decoupling matrix $\mathbf{B}(\mathbf{x})$ is invertible, then we can use the feedback control law (7.8) to obtain the decoupled dynamics (7.9).

$$\mathbf{u} = (\mathbf{B}(\mathbf{x}))^{-1}[-\mathbf{A}(\mathbf{x}) + \mathbf{v}] \tag{7.8}$$

$$\mathbf{y}^{\gamma} = \mathbf{v} \tag{7.9}$$

where $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \ldots & v_p \end{bmatrix}^T$

The vector $\mathbf{v}$ can be chosen to render the decoupled system (7.4) stable with desired transient behavior. Now, if the relative degree of the system r, is less than the order of the system n, then the closed loop system should also have stable internal dynamics. In order to study that, one can define state variables $\eta_i(\mathbf{x}), i = 1,2...,n-r$, which are independent of the state variables r related to the output of the system, and also independent of each other. The internal dynamics of the system then can be written as:

$$\dot{\eta} = \mathbf{w}(\zeta,\eta) + \mathbf{P}(\zeta,\eta)\mathbf{u} \qquad (7.10)$$

with (k=1,2,...,n-r) and (i=1,2,...,p)

$$w_k(\zeta,\eta) = L_f \eta_k(\mathbf{x}) \qquad (7.11)$$

$$P_{ki}(\zeta,\eta) = L_{g_i} \eta_k(\mathbf{x}) \qquad (7.12)$$

The feedback controller designed for (7.4) using the feedback linearization technique should also guarantee the stability of the internal dynamics described in (7.10). Note that for a single input case we could use the fact that $L_g \eta_k(\mathbf{x}) = 0$ to choose the independent internal dynamics state variables, but for multiple input case, this condition is not valid, unless the vectors of **g** are involutive.

7.2 Sample Problem (Two alternate routes with one section)

In order to illustrate the ideas discussed above, we have designed a feedback control law for two alternate routes problem with single section each. The control is based on feedback linearization technique for nonlinear systems. The technique is based on defining a diffeomorphism and performing the transformation on the state variables in order to convert them into the canonical form. If the relative degree of the system is less than the system order, then the internal dynamics are studied to ensure that it is stable.

The details of this technique are given in Isidori (1989). In this problem, the system order is two and the relative degree is one.

The space discretized flow equations used for the two alternate routes are:

$$\dot{\rho}_1 = -\frac{1}{\delta_1}\left[v_{f1}\,\rho_1(1-\frac{\rho_1}{\rho_{m1}}) - \beta u \right], \tag{7.13}$$

$$\dot{\rho}_2 = -\frac{1}{\delta_2}\left[v_{f2}\,\rho_2(1-\frac{\rho_2}{\rho_{m2}}) + \beta u - u \right], \tag{7.14}$$

We have considered a simple first order travel time function, which is obtained by dividing the length of a section by average velocity of vehicles on it. According to that, the travel time can be calculated as

$$\chi_1(k) = d_1 /[v_{f1}(1-\frac{\rho_1}{\rho_{m1}})], \tag{7.15}$$

$$\chi_2(k) = d_2 /[v_{f2}(1-\frac{\rho_2}{\rho_{m2}})], \tag{7.16}$$

where, d_1 and d_2 are section lengths, v_{f1} and v_{f2} are the free flow speeds of each section, and ρ_{m1} and ρ_{m2} are the maximum (jam) densities of each section. Since we need to equate the travel times according to the UE DTR formulation discussed in the previous section, we take the new transformed state variable y as the difference in travel times. Differentiating the equation representing y in terms of the state variables introduces the input split factor into the dynamic equation. Therefore, this transformed equation can be used to design the input that cancels the nonlinearities of the system and introduce a design input ν, which can be used to place the poles of the error equation for asymptotic stability. These steps are shown below:

The variable y is equal to the difference in the travel time on the two sections.

$$y = \frac{k_1}{(k_2 - \rho_1)} - \frac{k_3}{(k_4 - \rho_2)} \tag{7.17}$$

where $k_1 = \frac{d_1 \cdot \rho_{m1}}{v_{f1}}, k_2 = \rho_{m1}, k_3 = \frac{d_2 \cdot \rho_{m2}}{v_{f2}}, k_4 = \rho_{m2}$.

This equation can be differentiated with respect to time to give the travel time difference dynamics.

$$\dot{y} = \frac{k_1 \dot{\rho}_1}{(k_2 - \rho_1)^2} - \frac{k_3 \dot{\rho}_2}{(k_4 - \rho_2)^2} \tag{7.18}$$

By substituting (7.15) and (7.16) in (7.18) we obtain

$$\dot{y} = -\frac{k_1\left(v_{f1}\rho_1(1 - \frac{\rho_1}{\rho_{m1}}) - \beta u\right)}{\delta_1(k_2 - \rho_1)^2} + \frac{k_3\left(v_{f2}\rho_2(1 - \frac{\rho_2}{\rho_{m2}}) + \beta u - u\right)}{\delta_2(k_4 - \rho_2)^2} \tag{7.19}$$

This equation can be rewritten in the following form.

$$\dot{y} = F + G\beta \tag{7.20}$$

where

$$F = \left[-\frac{k_1 v_{f1} \rho_1}{\delta_1(k_2 - \rho_1)^2}(1 - \frac{\rho_1}{\rho_{m1}}) + \frac{k_3}{\delta_2(k_4 - \rho_2)^2}\left((1 - \frac{\rho_2}{\rho_{m2}})v_{f2}\rho_2 - u\right)\right] \tag{7.21}$$

$$G = \left(\frac{k_1}{\delta_1(k_2 - \rho_1)^2} + \frac{k_3}{\delta_2(k_4 - \rho_2)^2}\right)u \tag{7.22}$$

Hence a feedback linearization control law can be designed to cancel the nonlinearities and provide the desired error dynamics. The feedback control law given in (7.8) is used:

$$\beta = G^{-1}(-F + \nu) \tag{7.23}$$

which gives the closed loop dynamics as

$$\dot{y} = \nu \tag{7.24}$$

As was mentioned earlier, since the relative degree of the system is one, and the system order is two, we need to test the stability or boundedness of the second transformed state variable given by

$$\eta = \delta_1\rho_1 + \delta_2\rho_2 \tag{7.25}$$

The state variable η is bounded since the densities on the sections can not exceed the corresponding jam densities.

$$\eta \leq \delta_1\rho_{m1} + \delta_2\rho_{m2} \tag{7.26}$$

and hence the overall system is exponentially stable $(y \to 0)$ if we choose $v = -Ky, \quad K > 0$, and y asymptotically goes to zero as $y(t) = y(0)e^{-Kt}$. This implies that when a splitting value based on (7.23) is utilized, the difference in travel time of two alternate routes will go to zero at an exponential rate. Hence, the closed loop traffic system controlled by the proposed feedback linearization law is exponentially stable and has desired transient behavior.

7.3 Sample Problem (Two alternate routes with two sections)

Now, we extend the above problem to two sections case and follow the same steps for designing a new controller for this extended system. The space discretized flow equations used for the two alternate routes are:

$$\dot{\rho}_{11} = -\frac{1}{\delta_{11}}\left[v_{f11}\,\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}}) - \beta u \right], \tag{7.21}$$

$$\dot{\rho}_{12} = -\frac{1}{\delta_{12}}\left[v_{f12}\,\rho_{12}(1-\frac{\rho_{12}}{\rho_{m12}}) - v_{f11}\,\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}}) \right], \tag{7.22}$$

$$\dot{\rho}_{21} = -\frac{1}{\delta_{21}}\left[v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}}) + \beta u - u \right] \quad (7.23)$$

$$\dot{\rho}_{22} = -\frac{1}{\delta_{12}}\left[v_{f22}\rho_{22}(1-\frac{\rho_{22}}{\rho_{m22}}) - v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}}) \right] \quad (7.24)$$

We have considered a simple first order travel time function, which is obtained by dividing the length of a section by average velocity of vehicles on it. According to that, we approximate travel time as

$$\chi_1(t) = \frac{d_{11}}{v_{f11}(1-\frac{\rho_{11}}{\rho_{m11}})} + \frac{d_{12}}{v_{f12}(1-\frac{\rho_{12}}{\rho_{m12}})}, \quad (7.25)$$

$$\chi_2(t) = \frac{d_{21}}{v_{f21}(1-\frac{\rho_{21}}{\rho_{m21}})} + \frac{d_{22}}{v_{f22}(1-\frac{\rho_{22}}{\rho_{m22}})}, \quad (7.26)$$

where, d_1 and d_2 are section lengths, v_{f1} and v_{f2} are the free flow speeds of each section, and ρ_{m1} and ρ_{m2} are the maximum (jam) densities of each section.

The system can be written in the standard nonlinear input affine form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t) + \mathbf{g}(\mathbf{x},t)\mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x},t) \quad (7.27)$$

where

$$\mathbf{x} = [\rho_{11}\ \rho_{12}\ \rho_{21}\ \rho_{22}]' \quad ,\mathbf{u}(t) = \beta \quad (7.28)$$

$$\mathbf{f} = \begin{bmatrix} -\frac{1}{\delta_{11}}\left[v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}}) \right] \\ -\frac{1}{\delta_{12}}\left[v_{f12}\rho_{12}(1-\frac{\rho_{12}}{\rho_{m12}}) - v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}}) \right] \\ -\frac{1}{\delta_{21}}\left[v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}}) - u \right] \\ -\frac{1}{\delta_{12}}\left[v_{f22}\rho_{22}(1-\frac{\rho_{22}}{\rho_{m22}}) - v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}}) \right] \end{bmatrix} \tag{7.29}$$

$$\mathbf{g} = \begin{bmatrix} u \\ 0 \\ -u \\ 0 \end{bmatrix} \tag{7.30}$$

Since we need to equate the travel times on alternate routes according to the UE DTR problem formulation presented in the previous section, we take the new transformed state variable y as the difference in travel times. Differentiating the equation representing y in terms of the state variables introduces the input split factor into the dynamic equation. Therefore, that transformed equation can be used to design the input that cancels the nonlinearities of the system and introduce a design input v, which can be used to place the poles of the error equation for asymptotic stability. These steps are shown below:

The variable y is equal to the difference in the travel time on the two sections.

$$y(t) = \left[\frac{d_{11}}{v_{f11}(1-\frac{\rho_{11}}{\rho_{m11}})} + \frac{d_{12}}{v_{f12}(1-\frac{\rho_{12}}{\rho_{m12}})}\right] - \left[\frac{d_{21}}{v_{f21}(1-\frac{\rho_{21}}{\rho_{m21}})} + \frac{d_{22}}{v_{f22}(1-\frac{\rho_{22}}{\rho_{m22}})}\right] \tag{7.31}$$

This equation can be differentiated with respect to time to give the travel time difference dynamics.

$$\dot{y}=\frac{k_1\dot{\rho}_{11}}{(k_2-\rho_{11})^2}+\frac{k_3\dot{\rho}_{12}}{(k_4-\rho_{12})^2}-\frac{k_5\dot{\rho}_{21}}{(k_6-\rho_{21})^2}-\frac{k_7\dot{\rho}_{22}}{(k_8-\rho_{22})^2} \tag{7.32}$$

By substituting (7.21)- (7.24) in (7.32) we obtain

$$\dot{y}=-\frac{k_1\frac{1}{\delta_{11}}\left[v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}})-\beta u\right]}{(k_2-\rho_{11})^2}-\frac{k_3\frac{1}{\delta_{12}}\left[v_{f12}\rho_{12}(1-\frac{\rho_{12}}{\rho_{m12}})-v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}})\right]}{(k_4-\rho_{12})^2}$$
$$+\frac{k_5\frac{1}{\delta_{21}}\left[v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}})+\beta u-u\right]}{(k_6-\rho_{21})^2}+\frac{k_7\frac{1}{\delta_{12}}\left[v_{f22}\rho_{22}(1-\frac{\rho_{22}}{\rho_{m22}})-v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}})\right]}{(k_8-\rho_{22})^2} \tag{7.33}$$

This equation can be rewritten in the following form.

$$\dot{y}=F+G\beta \tag{7.34}$$

where

$$F=-\frac{k_1\frac{1}{\delta_{11}}\left[v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}})\right]}{(k_2-\rho_{11})^2}-\frac{k_3\frac{1}{\delta_{12}}\left[v_{f12}\rho_{12}(1-\frac{\rho_{12}}{\rho_{m12}})-v_{f11}\rho_{11}(1-\frac{\rho_{11}}{\rho_{m11}})\right]}{(k_4-\rho_{12})^2}$$
$$+\frac{k_5\frac{1}{\delta_{21}}\left[v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}})-u\right]}{(k_6-\rho_{21})^2}+\frac{k_7\frac{1}{\delta_{22}}\left[v_{f22}\rho_{22}(1-\frac{\rho_{22}}{\rho_{m22}})-v_{f21}\rho_{21}(1-\frac{\rho_{21}}{\rho_{m21}})\right]}{(k_8-\rho_{22})^2} \tag{7.35}$$

$$G=[\frac{k_1\frac{1}{\delta_{11}}}{(k_2-\rho_{11})^2}+\frac{k_5\frac{1}{\delta_{21}}}{(k_6-\rho_{21})^2}]u \tag{7.36}$$

Hence a feedback linearization control law similar to the one given by (7.8) can be designed to cancel the nonlinearities and provide the desired error dynamics. The law used is

$$\beta=G^{-1}(-F+v) \tag{7.37}$$

which gives the closed loop dynamics as

$$\dot{y}=v \tag{7.38}$$

The relative degree of the system is one, and the system order is four, we need to test the stability or boundedness of the three internal states. Now, our task is to obtain the other three independent state variables. Since this is a single input system, to obtain these state variables we should satisfy $L_g\eta_k(\mathbf{x}) = 0$ as follows:

$$\frac{\partial\eta_i}{\partial\rho_{11}} = \frac{\partial\eta_i}{\partial\rho_{21}}, \; i = 1,2,3 \tag{7.39}$$

$$\eta_1 = \rho_{12}$$

$$\eta_2 = \rho_{22}$$

$$\eta_3 = \rho_{11} + \rho_{21}$$

The state variable η is bounded since the densities on the sections can not exceed the corresponding jam densities.

$$\begin{aligned} \eta_1 &\le \rho_{m12} \\ \eta_2 &\le \rho_{m22} \\ \eta_3 &\le \rho_{m11} + \rho_{m12} \end{aligned} \tag{7.40}$$

and hence the overall system is exponentially stable $(y \to 0)$ if we choose $v = -Ky, \; K > 0$, and y asymptotically goes to zero as $y(t) = y(0)e^{-Kt}$. This implies that when a splitting value based on (7.37) is utilized, the difference in travel time of two alternate routes will go to zero at an exponential rate. Hence, the closed loop traffic system controlled by the proposed feedback linearization law is exponentially stable and has desired transient behavior.

7.4 Solution for the Generalized DTR Problem for Multiple Routes with Multiple Sections

In this section, we give a generalize solution for the n alternate route DTR problem described in section 6.6.1 of Chapter 6. The space discretized flow equations used for the n alternate routes and n sections are given by (6.46) and (6.47). Number of sections for each alternate route I is denoted by n_i. We are considering full state observation, which is used for estimating (sensing) the travel times on the various alternate routes. The dynamics can be written as

$$\dot{\rho}_{ij} = \frac{1}{\delta_{ij}}\left[v_{fij-1}\rho_{ij-1}(1-\frac{\rho_{ij-1}}{\rho_{mij-1}}) - v_{fij}\rho_{ij}(1-\frac{\rho_{ij}}{\rho_{mij}}) \right] \tag{7.41}$$

$$\text{when}\,(i,j) = ((1,2),...,(1,n_1),(2,2),...,(2,n_2),...,(n,2),...,(n,n_n))$$

$$\dot{\rho}_{ij} = \frac{1}{\delta_{ij}}\left[\beta_i u - v_{fij}\rho_{ij}(1-\frac{\rho_{ij}}{\rho_{mij}}) \right] \text{ when } (i,j) = ((1,1),(2,1)...(n,1)) \tag{7.42}$$

We have considered a simple first order travel time function, which is obtained by dividing the length of a section by average velocity of vehicles on it. According to that, we approximate travel time for a route as

$$\chi_i(t) = \sum_{j=1}^{n_i} \frac{d_{ij}}{v_{fij}(1-\frac{\rho_{ij}}{\rho_{mij}})} \tag{7.43}$$

The system can be written in the standard nonlinear input affine form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x},t) + \mathbf{g}(\mathbf{x},t)\mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x},t) \tag{7.44}$$

where

$$\mathbf{x} = [\rho_{11}...\rho_{1n_1}\cdots\rho_{n1}\cdots\rho_{nn_n}]' \quad ,\mathbf{u}(t) = [\beta_1...\beta_{n-1}] \tag{7.45}$$

The output vector is denoted by **y**, and is given by:

$$\mathbf{y} = [y_1 \; y_2 \cdots y_i \cdots y_{n-1}] \qquad (7.46)$$

where,

$$y_i = \chi_{i+1}(t) - \chi_i(t) \qquad (7.47)$$

This equation can be differentiated with respect to time to give the travel time difference dynamics.

$$\dot{y}_i = \sum_{j=1}^{n_{i+1}} \frac{k_{i+1j1}\dot{\rho}_{i+1j}}{k_{i+1j2}(1-\frac{\rho_{i+1j}}{\rho_{mi+1j}})^2} - \sum_{j=1}^{n_i} \frac{k_{ij1}\dot{\rho}_{ij}}{k_{ij2}(1-\frac{\rho_{ij}}{\rho_{mij}})^2} \qquad (7.48)$$

where k_{ijp} denotes a constant p=1 or 2 that belongs to section j of route i, similar to the constant k described for (7.17). The system (7.48) is in the form of (7.3) and can be represented in the form of (7.4) by finding out the values of **A**(**x**) and **B**(**x**). The control law (7.8) provides us with user equilibrium for DTR problem.

The input appears in all the output equations after differentiating them one time. Hence the relative degree of the system is n-1. The order of the system is $\sum_{i=1}^{n} n_i$. Since the densities on the sections are bounded by jam density values, the independent state variables $\eta_i, i = 1,2,....\sum_{i=1}^{n} n_i + 1 - n$ are also bounded.

7.5 Simulation

Several simulation studies are performed to demonstrate the working of the feedback linearization technique presented in this paper. The test network which consists of two alternate routes is shown in Figure 7.1. Three different simulation scenarios that are chosen are:

1. Model without any parametric uncertainties and full user compliance
2. Model with parametric uncertainties and full user compliance
3. Model with parametric uncertainties and partial user compliance

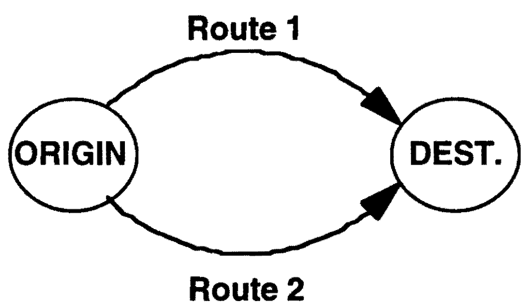


Figure 7.1: Sample Network

The input function is assumed to be a sinusoidal function which reaches a pre-defined peak value and then settles at a constant value for the rest of the simulation period. This function emulates the peak hour demand that reaches its maximum value at a certain time, and then settles at a constant value when the peak period is over. In this specific simulation study, the peak period is assumed to be 1 hour.

7.5.1 Scenario 1: Model with Full User Compliance and without any Uncertainties

In this scenario, the controller has perfect knowledge of the parameters of the traffic model. Also, full compliance of the users to the diversion commands is assumed by the controller. The system dynamics model also simulates full compliance of the users to the controller's diversion commands. In this case, since the controller has the complete knowledge of the system dynamics, it is able to perform exact cancellation of the system nonlinearities and attain exponential error convergence. This result is shown in Figure 7.2. Figure 7.3 shows split factors.

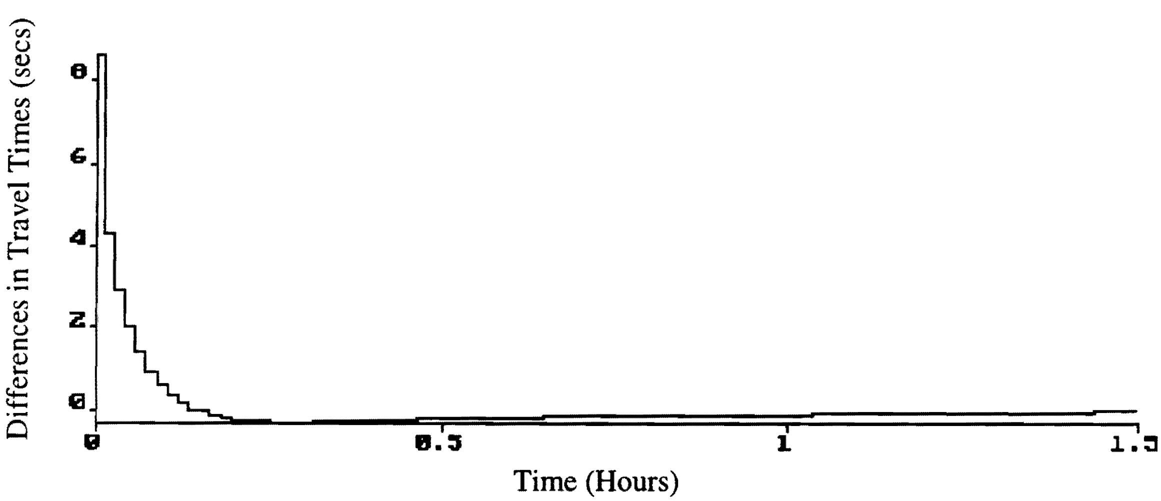


Figure 7.2: Differences in Travel Times (Secs) for Scenario 1

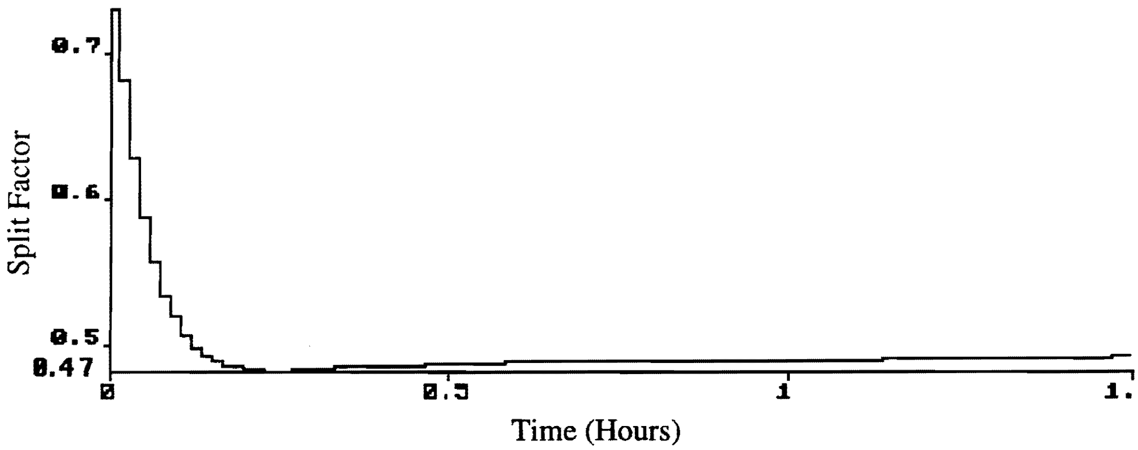


Figure 7.3: Split Factors for Scenario 1

7.5.2 Scenario 2: Model with Full User Compliance and Uncertainties

In this scenario, the controller does not have perfect knowledge of the parameters of the traffic model. In order to simulate the effects of uncertainties, ± 30% errors have been assumed. However, full compliance of the users to the diversion commands is assumed by the controller and also simulated by the system model. In this case, since the controller does not have the complete knowledge of the system dynamics, it is not able to perform exact cancellation of the system nonlinearities and attain exponential error convergence. However, as it can be seen Figures 7.4 and 7.5, the results obtained by using this controller even with such relatively large parametric uncertainty, are highly encouraging.

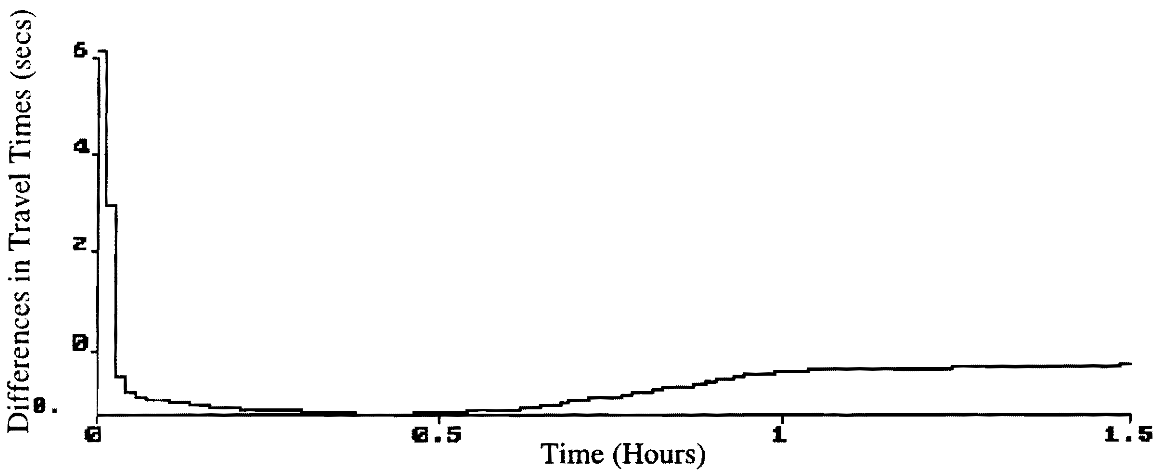


Figure 7.4: Differences in Travel Times (Secs) for Scenario 2

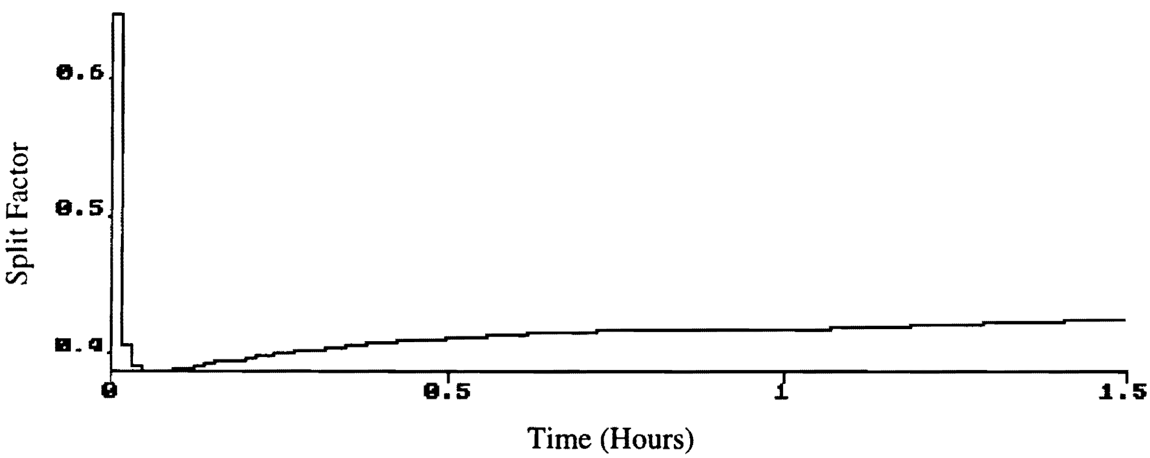


Figure 7.5: Split Factors for Scenario 2

7.5.3 Scenario 3: Model with Partial User Compliance and Uncertainties

In this scenario, we assume both partial user compliance (90%), and existence of parametric errors in the model. As it can be seen in Figure 7.6, the fluctuations of differences in travel time are much higher than the previous scenarios and it takes the controller longer time to attain error convergence. However, even with partial user compliance and fairly large parametric uncertainties, the system stabilizes and the differences in travel times asymptotically converge at a desirable rate.

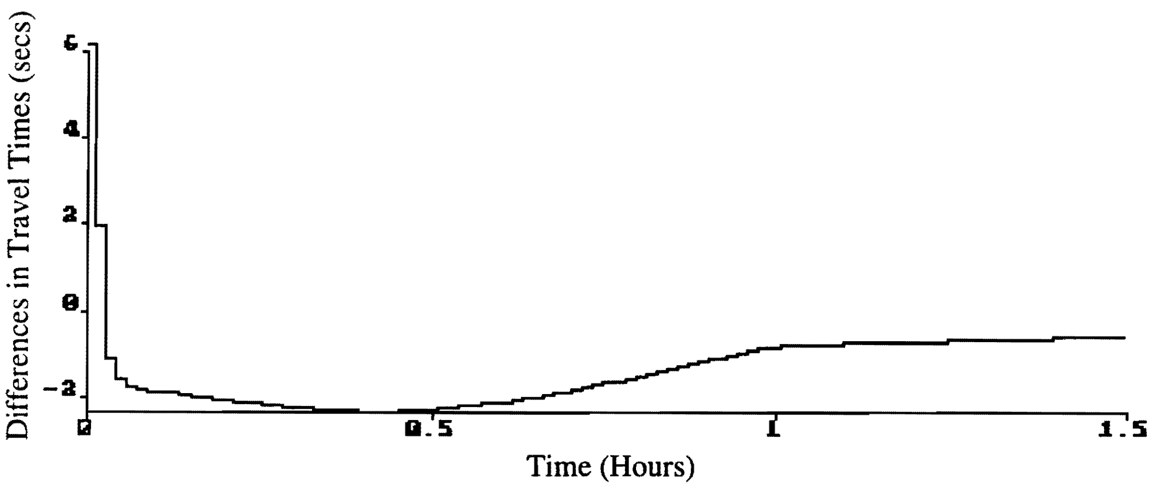


Figure 7.6: Differences in Travel Times (Secs) for Scenario 3

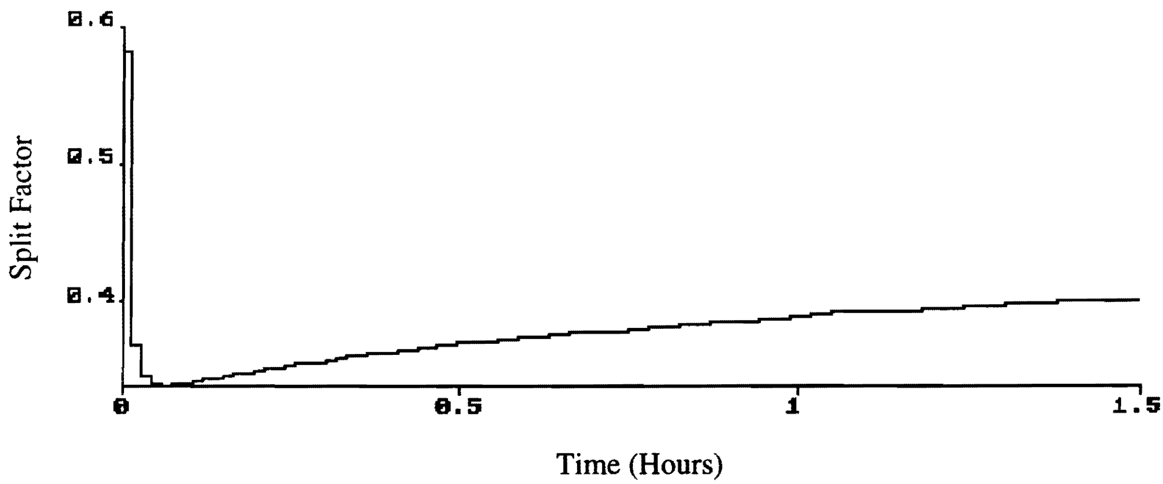


Figure 7.7: Split Factors for Scenario 3

7.6 Conclusions

In this Chapter, we have addressed the real-time traffic control problem for point diversion. A feedback model is developed for control purposes, and feedback linearization technique is used to design this feedback controller. First, the simplest case, with two alternate routes consisted of a single section each, is studied and a feedback controller using feedback linearization technique is developed. Second, the case with two alternate routes with two discrete sections, is analyzed and a feedback controller using feedback linearization technique is also developed. Finally, the general case with multiple alternate routes divided into multiple sections is analyzed and a general solution is proposed. To illustrate the above models, simulation runs are performed for three different scenarios for a network topology of two alternate routes. The feedback controller developed for this test network performed fairly well for all three scenarios. An important finding of this simulation study was the robustness of the controller even for situations where parametric uncertainties exist and partial user compliance is employed. Therefore, we can conclude that feedback linearization is an effective method for designing real-time traffic control systems.

8. Conclusions and Future Research

This chapter summarizes the research conducted in this dissertation and provides recommendations and avenues for future research.

This dissertation provides the following.

1. A practical and efficient framework for real-time diversion during non-recurrent congestion.
2. A prediction / decision tree model for predicting incident clearance duration.
3. A knowledge based expert system (KBES) that determines the incident impact zone, and diversion route(s) during an incident, using simple expert rules in conjunction with static and dynamic network and traffic conditions.
4. Several feedback control model formulations and solution techniques for controlling the point diversion of traffic in real-time.

Chapter 3 describes the proposes real-time diversion framework, its components, and the interaction among these components. Chapter 4 proposes a methodology for predicting incident clearance duration, analyzes the incident clearance duration data collected by the Northern Virginia incident management agencies, determines the major factors affecting incident clearance duration, and develops and validates prediction / decision trees for forecasting incident durations. Chapter 4 also describes a simple deterministic queuing approach along with a diversion initiation logic to decide if a diversion is necessary. Chapter 5 describes in detail a knowledge based expert system, "Network Generator", that determines the incident impact zone, eliminates links for diversion based on simple rules and historical and real-time data, and generates alternate route(s) for diversion.

Chapter 5 also describes the software implementation of this expert system as a hybrid Geographical Information System (GIS) and KBES, and discusses several implementation issues addressed by this software architecture adopted in this research. The software architecture, integration of GIS and KBES, and the relevance of the software architecture and components for real-time applications, are also briefly discussed. The prediction / decision trees are validated using 73 incident cases that were not used for the development of the trees. Prediction / decision trees successfully predicted incident durations for most of the cases. This shows us that our approach is well suited for predicting incident durations. A personal injury traffic accident on Interstate 66, Fairfax Virginia, is also used as an example to illustrate the working and effectiveness of the models and solution algorithms presented in Chapters 4 and 5. The simulation runs were conducted using INTEGRATION simulation software. Simulations results showed that the alternate routes obtained using "Network Generator" provided the most efficient diversion strategy in terms of network-wide delays.

Chapter 6 formulates the dynamic traffic routing problem for real-time point diversion as a feedback control problem both in distributed and lumped parameter settings. Chapter 5 also provides two heuristic controller designs for the discretized version of the lumped parameter formulation of this problem. The first one is a PI Proportional-Integral (PI) controller with a constant feedforward term. The second is a qualitative controller that uses fuzzy logic. The simulation results for a simple network which consists of three alternate routes showed that both controllers produced asymptotic stability to the system and the defined error terms tended to zero as time progressed.

Chapter 7 provides an analytical solution to the Dynamic Traffic Routing (DTR) problem for point diversion. This problem is formulated as a feedback control problem that determines the time-dependent split parameters at the diversion point for routing the incoming traffic flow onto the alternate routes, in order to achieve a user-equilibrium (UE) traffic pattern. The feedback linearization technique is used to solve this specific UE formulation of the Dynamic Traffic Routing (DTR) problem. The control input is the split factor which is obtained by canceling the nonlinearities of the dynamics of the system after transforming the original dynamics of the system into canonical form. Simulation results for a simple network which consists of two alternate routes show that the performance of this controller on a test network is quite promising.

Many improvements can be made to the models and algorithms developed in this dissertation. These are as follows.

1. New incident data can be used to enhance the incident prediction / decision trees. The data set that is used in this research does not have sufficient number of certain types of rare incidents such as HAZMAT and weather related incidents. Moreover, the data collection can be computerized in order to prevent data entry errors that have corrupted some of the data in this research.
2. The developed Network Generator is still an enhanced prototype and needs to be validated for a variety of situations using both simulation and on-line data. As a result of these studies, more location and incident specific rules can be incorporated into the expert system.
3. Feedback control models and solution algorithms developed in this dissertation are only applicable to the point diversion problem. More general feedback models and solution algorithms for network-wide problems can be developed. Also, the models presented in this dissertation are all user equilibrium models. A system optimal

feedback control formulation can be also developed and compared to these user equilibrium models that are developed in this dissertation.

4. Finally, feedback control models can be tested and enhanced using microscopic simulation models that can simulate user behavior as well traffic flow in order to understand the effects of user behavior in achieving a stable system.

References

1. Banos J.C.M., M., Papageorgiou, C., Schaffner. Integrated Optimal Flow Control in Traffic Networks. *European Journal of Operation Research,* Vol. 71, pp. 317-323, 1993.
2. Ben-Akiva M., A., DePalma and P., Kanaraglou. Dynamic Model of Peak-Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science*, Vol. 20, No. 2, 1986.
3. Ben-Akiva M., M., Cyna, A., DePalma. Dynamic Model of Peak Period Congestion. *Transportation Research,* Vol.18B, 1984, pp.339-355.
4. Berger, C. R., R.L., Gordon, and P.E., Young. Single Point Diversion of Freeway Traffic. *Transportation Research Record,* Vol. 601, pp. 10-17, 1976.
5. Berger, C. R., and L., Shaw. Discrete Event Simulation of Freeway Traffic. Simulation Council Proc., Series 7, 1977, 85-93.
6. Bernstein, D., T.L,. Friesz, R.L., Tobin, and B.W., Wie. A Variational Control Formulation of the Simultaneous Route and Departure Choice Equilibrium Problem. Transportation and Traffic Theory, *Proceedings of the 12th International Synposium on the Theory of Traffic Flow and Transportation*, Berkeley, California, USA, pp. 107-126, 1993.
7. Birge, J.R., and J.K., Ho. Optimal Flows in Stochastic Dynamic Networks with Congestion. Operations Research, Vol. 4, No. 1, pp. 203-216, 1993.
8. Bovy, PH.L., and G.R.M., Jansen. Spatial Aggregation in Traffic Assignment: Effects of Zone Size and Network Detail upon All-or-Nothing Assignment Results. Paper Presented at PTRC - Summer Annual meeting, Warwick, England, 1978.

9. Breiman, L., J.H., Friedman, R.A., Olshen, C.J., Stone. Classification and Regression Trees. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, California, 1984.

10. Burns, J. A., and S., Kang. A Control Problem for Burgers Equation with Bounded Input / Output. Nonlinear Dynamics 2, pp.235-262, 1991.

11. Burns, J. A., and S., Kang. A Stabilization Problem for Burgers Equation with Unbounded Control and Observation. Intl. Series of Numerical Math., 100, 51-72, 1991.

12. Burgers, J. M. Mathematical Examples Illustrating Relations Occuring in the Theory of Turbulent Fluid Motion. Trans. Roy. Neth Acad. Sci. 17, Amsterdam, pp. 1-53, 1939.

13. Burgers, J. M. Mathematical Model Illustrating the Theory of Turbulence. Adv. in Appl. Mech. 1, 171-199, 1948.

14. Burgers, J. M. Statistical Problems Connected with Asymptotic Solution of One-Dimensional Nonlinear Diffusion Equation. In M. Rosenblatt and C. Van Atta (eds.), *Statistical Models and Turbulence*, Springer, Berlin, 41-60, 1972.

15. Cambridge Systematics. Incident Management. Final Report Prepared for Trucking Research Institute, ATA Foundation Inc., 1990.

16. Carey, M. Nonconvexity of the Dynamic Traffic Assignment Problem. Transportation Research- B, Vol. 26B, No. 2, pp.12-133, 1992.

17. Carey, M. Extending and Solving a Multiperiod Congested Network Flow Model. Computers Opns Research, Vol.17, No. 5, pp.495-50, 1990.

18. Carey, M. Optimal Time-Varying Flows on Congested Networks Operation Research, Vol. 35, 1987 pp 58-63, 1987.

19. Carey, M. A Constraint qualification for a Dynamic Traffic Assignment Model, Transportation Science 20(1), pp.55-58, 1986.

20. Cascetta E., A., Nuzzolo, and L., Biggiero. Analysis and Modeling of Commuters' Departure Time and Route Choices in Urban Networks, Second International Capri Seminar on Urban Traffic Networks, July 1992.

21. Cremer M., S., Fleischman. Traffic Responsive Control of Freeway Networks by a State Feedback Approach. Transportation and Traffic Theory. N.H. Garthner and N.H.M. Wilson, editors, pp.357-377, 1987.

22. Cole, J. D. On a Quasi-linear Parabolic Equation Occuring in Aerodynamics. *Quart. Appl. Math,*. IX, 225-236, 1951.

23. Curtain, R. F. Stability of Semilinear Evolution Equations in Hilbert Space. *J. Math. Pures et Appl.* 63, 121-128, 1984.

24. Daganzo, C., Y., Sheffi. On Stochastic Models of Traffic Assignment. Transportation Science, 11, pp. 253-274, 1977

25. Daganzo, C. Properties of Link Travel Time Functions under Dynamic Loads. Transportation Research-B, Vol 29B, No. 2, pp. 95-98, 1995.

26. Daniels, E., M. Leven, and J., McDermott. Improving Commercial Radio Traffic Reports in the Chicago area. *Transportation Research Record 600*, Transportation Research Board, Washington D.C., pp. 52-57, 1976.

27. Davis, G., J., Kang. Estimating Destination-Specific Traffic Densities on Urban Freeways for Advanced Traffic Management. *Transportation Research Board*, Washington D.C., 1994.

28. Drissi-Kaitouni, O., A., Hamed-Benchekroun. A Dynamic Traffic Assignment Model and a Solution Algorithm. *Transportation Science*, Vol 26(2), pp 119-128, 1992.

29. Eash R. W., K. S., Chon., Y. J., Lee, and D. E., Boyce.. Equilibrium Traffic Assignment on an Aggregated Highway Network for Sketch Planning. Chicago Area Transportation Study, 1983.

30. Ford L.R. and D.R., Fulkerson. Flows in Networks. Princeton University Press, Princeton New Jersey, 1962.

31. Friesz, T.L., D. Bernstein , T.E., Smith, B.W., Wie. A Variational Inequality Formulation of the Dynamic Network User Equilibrium Model. Operations Reseach, Vol. 42, No. 1, pp. 179-191, 1993.

32. Friesz, T.L., F.J., Luque, R.L., Tobin, B.W., Wie. Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem, *Operation Research,* 37(6) pp 893-901, 1989.

33. Fletcher, C., A. J., Burgers equation: a Model for all Reasons. In J. Noye (ed.), *Numerical Solution of Partial Differential Equations*, North-Holland, 139-225, 1982.

34. Garrison, D., and F., Mannering. Assessing the Traffic Impacts of Freeway Incident and Driver Information. ITE Journal 60 (8), pp. 19-23,1990.

35. Gartner, N. H., and R.A., Reiss. Congestion Control in Freeway Corridors: The IMIS System. *NATO ASI Series*, Vol. F38, Flow Control of Congested Networks', Springer-Verlag, 1987.

36. Glimm, J., and P., Lax. Decay of Solutions of Systems of Nonlinear Hyperbolic Conservation Laws. Amer. Math. Soc. Memoir 101, A.M.S., Providence, 1970.

37. Golob, T., W., Recker, and J., Leonard. An analysis of the Severity and Incident Duration of Truck Involved Freeway Incidents. Accident Analysis and Prevention, Vol. 19, No. 5, pp. 375-395, 1987.

38. Goolsby, M. Influence of Incident on Freeway Quality of Service. *Highway Research Board*, HRR 349, Washington D.C, pp. 41-46., 1971

39. Gupta, A., V., Maslanka, and G., Spring. Development of a Prototype KBES in the Management of Congestion. Preprint No. 920302, Transportation Research Board, Washington D.C., 1992.

40. Giuliano, G. Incident Characteristics, Frequency, And Duration on a High Volume

Urban Freeway. *Transportation Research-A*, Vol. 23 A, No. 5, pp. 387-396, 1989.

41. Haghani, A. E., and M.S., Daskin. Network Design Application of an Extraction Algorithm for Network Aggregation. *Transportation Research Record 944*, 1983.

42. Hall, M.D., D., Van Vliet, L.G., Willumsen. SATURN-A Simulation Assignment Model for the Evaluation of Traffic Management Schemes,Traffic Engineering and Control, 21, pp.168-176, 1980.

43. Hall, W.H. Research on Route Diversion Parameters. Final Report prepared for FHWA, December 1974.

44. Hammerslag, R. Dynamic Assignment in the Three-Dimensional Time Space, *Transportation Research Board,* 68th Annual Meeting, January 22-26, 1988, Washington, DC.

45. Hearn D. W. Practical and Theoretical Aspects of Aggregation Planning in Transportation Planning Models. *Transportation Planning Models*, pp. 258-287, Elsevier Science Publishers, 1984.

46. Heathington K., R., Worrall., and G., Hoff. Attitudes and Behavior of Drivers Regarding Route Diversion. *Highway Research Record,* 3633, Highway Research Board, Washington D.C., 1971, pp. 18-23.

47. Heckerman, D. Probabilistic Interpretations of MYCIN's Certainty Factors. Uncertainty in Artificial Intelligence. Kanal, L., and J., Lemmer, Eds, pp.9-20, North Holland, 1986.

48. Hendrickson C., G., Kocur. Scheduled Delay and Departure Time Decisions in a Deterministic Model. *Transportation Science* 15 (1), pp.62-77, 1977.

49. Hendrickson C., D., Nagin, E., Plank. 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, pp. 321-347, 1983.

50. Ho, J.K. A Successive Linear Optimization Approach to the Dynamic Traffic Assignment, *Transportation Science*, Vol. 14, pp 295-305, 1980.

51. Hobeika, A. G., R., Sivanandan, S., Subramaniam, and K., Ozbay. Real-Time Traffic Diversion Model: Conceptual Approach. *ASCE Journal of Transportation Engineering*, Vol. 119, No. 4, pp. 515-534, July/Aug. 1993.

52. Hopf, E. The Partial Differential Equation $u_t + uu_x = \mu u_{xx}$. *Comm. Pure and Appl. Math 3*, pp. 201-230, 1950.

53. Institute of Transportation Engimeers. *Mobility Facts*. Washington D.C., 1992.

54. Isidori, A., *Nonlinear Control Systems*, Springer-Verlag, 1989.

55. Janson, B.N. A Convergent Algorithm for Dynamic Traffic Assignment, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6366, 1990.

56. Janson, B.N. Estimating Departure Times for Traffic Counts Using Dynamic Assignment, to be presented at the International Conference on Dynamic Traffic Behavior Analysis, Japan, July 18-19, 1991.

57. Jones, B., L., Janssen, and F., Mannering. Analysis of the Frequency and Duration of Freeway Accidents in Seattle. Accident Analysis and Prevention, Vol. 23, No. 4, pp. 239-255, 1991.

58. Kachroo P. and K. Ozbay., " Solving the Dynamic Traffic Routing Problem using Feedback Linearization", Submitted to Transportation Research Part B, 1996.

59. Kachroo P., K. Ozbay. "*Modeling and Fuzzy Feedback Control for Real-Time Traffic Routing.*" Paper Presented at the 1996 TRB Annual Conference at Washington D.C. and Forthcoming in Transportation Research Record, 1996.

60. Kaufman, D.E., K.E., Wunderlich, R.L., Smith. An Iterative Routing / Assignment Method for Anticipatory Real-Time Route Guidance, IVHS Technical Report 91-02, May 1991.

61. Ketselidou, Z. Potential Use of Knowledge-Based Expert Systems for Post-Incident

Traffic Control. Ph.D. Thesis, University of Massachusetts at Amherst, February 1989.

62. Khattak A. J., F.S., Koppelman, and J.L., Schofer. Stated Preferences for Investigating Commuters' Diversion Propensity. 71st TRB Annual Meeting, Washington D.C., 1992.

63. Kielhofer, H. Stability and Semilinear Evolution Equations in Hilbert Space. *Arch. Rat. Mech. Anal.* 57, pp.150-165, 1974.

64. Koutsopoulos, H.N., Xu H. An Information Discounting Routing Strategy for Advanced Traveler Information Systems, presented at the 71st TRB Conference, Washington D.C., 1992.

65. Koutsopoulos, H.N., Lotan, T. Role of Motorist Information Systems in Reducing Traffic Congestion. Final Report to University for Transportation Studies, January 1990, MIT.

66. Kuo, B. C., *Automatic Control Systems*, Prentice Hall, Inc., 1987.

67. LaFortune, S., Sengupta, R., Kaufman D.E., Smith R.L. A Dynamical System Model for Traffic Assignment in Networks. IVHS Technical Report 91-02, May 1991.

68. Lax, P. D., *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves.* CBMS-NSF Regional Conference Series in Applied Mathematics 11, SIAM, 1973.

69. Leboeuf, J., T., Tajima, and J., Dawson. A Magnetohyrodynamic Particle Code for Fluid Simulation of Plasmas. *Journal of Computational Physics*, Vol.31, No 3, June pp. 379-408, 1979.

70. Leonard, D.R., J.B., Though, and P.C., Baguley. CONTRAM: A Traffic Assignment Model for Predicting Flows and Queues During Peak Periods, TRRL Report LR 841, Transport and Road Research Laboratory, Crowthorne, 1978.

71. Lighthill, M. J., and G.B., Whitham. On kinematic Waves II. A Theory of Traffic

Flow on Long Crowded Roads', Proc. of the Royal Society of London, Series A 229, 317-345, 1955.

72. Lindley, J. Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions: An Update. ITE Journal, 1987.
73. Lindley, J. Urban Freeway Congestion Problems and Solutions: An Update. ITE Journal, 1989.
74. Lindley, J. Quantification of Urban Freeway Congestion and Analysis of Remedial Measures. FHWA Staff Report, 1986.
75. Lisco L. Procedure for Predicting Queues and Delays on Expressways in Urban Core Area, *Trans. Research Record 944*, pp.148-154, 1983.
76. Mamdani, E.H.Advances in the Linguistic Synthesis of Fuzzy Controls. *Int. Journal of Man-Machine Studies*, 8, pp.669-678, 1976.
77. Mahmassani, H.S., Ta-Yin Hu, S. Peeta, and A. Ziliaskopoulos. Development and Testing of Dynamic Traffic Assignment and SimulationProcedures for ATIS / ATMS Applications. Technical Report DTFH61-90-R-00074-FG, June 1994.
78. Mahmassani, H., C., Caplice, and C., Walton. Characteristics of Urban Commuter Behavior: Switching propensity and Use of Information. Preprint No. 890738, Transportation Research Board, Washington D.C., 1990.
79. Mahmassani, H.S., G.L., Chang, and R., Herman. Macroparticle Traffic Simulation to Investigate Peak-Period Commuter Decision Dynamics. *Transportation Research Rec. 1005*, 1984, pp.107-120, 1985.
80. Maslov, V. P., On a New Principle of Superposition for Optimization Problems', *Uspekhi Mat. Nauk* 42, 39-48, 1987.
81. Maslov, V. P. A New Approach to Generalized Solutions of Nonlinear Systems. *Soviet Math. Dokl* 35, pp. 29-33, 1987.
82. Mastbrook, S. A Framework for Predicting Urban Freeway Incident Delay in Real

Time. Virginia Tech Master Thesis, 1996.

83. Merchant, D.K., G.L., Nemhauser. A Model and an Algorithm for the Dynamic Traffic Assignment Problems. *Transportation Science*, Vol. 12, pp. 183-199, 1978.

84. Merchant, D.K., G.L., Nemhauser. Optimality Conditions for a Dynamic Traffic Assignment. *Transportation Science*, Vol. 12, pp 200-207, 1978.

85. Messmer, A., M., Papageorgiou. Route Diversion Control in Motorway Networks via Nonlinear Optimization. *IEEE Transactions on Control Systems Technology.* Vol. 3, No. 1. March 1995.

86. Messmer, A., M., Papageorgiou. Optimal Route Control Via Route Reommendation. *1994 Vehicle Navigation and Information Systems Conference Proceedings*, pp.298-303, 1994.

87. Meyer, M. A Toolbox for Alleviating Traffic Congestion. Institute of Transportation Engineers, Washington D.C., 1989.

88. Morales, J.M. Analytical Models for Estimating Freeway Traffic Congestion. *Public Roads*, Vol.50, No.2, pp.55-61, 1986.

89. Mosca, E. *Optimal Predictive and Adaptive Control*, Prentice Hall, 1995.

90. Musha, T., and Higuchi, H. Traffic Current Fluctuation and the Burgers Equation. *Japanese Journal of Applied Physics*, 1978, 17, 5, 811-816.

91. Nijmeijer, H., and Schaft, A. van der, *Nonlinear Dynamical Control Systems*, Springer-Verlag, 1990.

92. Ozbay, K., A.G., Hobeika, S., Subramaniam, and V., Krishnaswamy. A Heuristic Network Generator for Traffic Diversion During Non-Recurrent Congestion. Transportation Research Board, 73rd Annual Meeting, Washington, D.C., Jan. 1994.

93. Papageorgiou, M., and A., Messmer. Dynamic Network Traffic Assignment and Route Guidance via Feedback Regulation. Transportation Research Board, Washington, D.C., Jan. 1991.

94. Papageorgiou, M. Dynamic Modeling, Assignment and Route Guidance Traffic Networks. *Transportation Research-B*, Vol. 24B, No. 6, pp.471-495, 1990.

95. Papageorgiou, M., J.M., Blosseville, and H. Hadj-Salem. Modeling and Real-Time Control of Traffic on the Southern Part of Boulevard Peripherique in Paris: Part 1: Modeling. Transportation Research-A, Vol. 24A, No. 5, pp.345-359, 1990.

96. Papageorgiou, M., J.M., Blosseville, and H. Hadj-Salem. Modeling and Real-Time Control of Traffic on the Southern Part of Boulevard Peripherique in Paris: Part 2: Coordinated on Ramp Metering. Transportation Research-A, Vol. 24A, No. 5, pp.361-370, 1990.

97. Papageorgiou, M. Macroscopic Modeling of Traffic Flow on the Boulevard Peripherique in Paris. *Transportation Research-B*, Vol. 23B, No. 1, pp.29-47, 1989.

98. Papageorgiou, M. *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control*, Springer-Verlag, 1983.

99. Payne, H. J. Models of Freeway Traffic and Control. Simulation Council Proceedings, pp.51-61, 1971.

100. Ran, B., H., Lo, and B., Hongola. Multi-class Dynamic Traffic Model: Formulation and Computational Experiences. Presented at the 75th TRB Annual Meeting, January 7-11, 1996, Washington, D.C.

101. Ran, B. H., Lo, and B., Hongola. Multi-class Dynamic Traffic Assignment, Part I: A Variational Inequality Formulation. Submitted to Transportation Research, 1996.

102. Ran, B. H., Lo and B., Hongola. Multi-class Dynamic Traffic Assignment, Part II Development of Relaxation Algorithm. Submitted to Transportation Research, 1996.

103. Ran, B. and D., Boyce. Dynamic Urban Transportation Network Models: Theory and Implications for Intelligent Vehicle Highway Systems. No. 417 Lecture Notes in economics and Mathematical Systems, Spinger Verlag, Heidelberg, 1994.

104. Ran B., D., Boyce D.E., LeBlanc L.J. Solving an Instantaneous Dynamic User

Optimal Route Traffic Assignment Model. *Transportation Science*, Vol. 29, No.2, pp. 128-142, 1995.

105. Ran B., Boyce D.E., LeBlanc, L.J. A New Class of Instantaneous Dynamic User Optimal Route Traffic Assignment Model. Operations Research 41, pp. 192-202, 1993

106. Ran B., Boyce, D.E., L.J., LeBlanc. Dynamic User Optimal Time and Route Model: A Bilevel, Optimal Control Formulation. Paper Submitted for Consideration for Publication in the special issue of the Annals of Operations Research: Advances in Equilibrium Modeling, Analysis and Computation, 1991 (1992).

107. Ran B., D.E., Boyce and, L.J., LeBlanc. Dynamic User Optimal Time Route ChoiceBased on Route Travel Times. Paper Submitted for Consideration for Publication in *Transportation Research,* 1991 (1992).

108. Recker, W. An Analysis of the Characteristics and Congestion Impacts of Truck-Involved Freeway Accidents. Final Report prepared for CalTrans, Institute of Transportation Studies, University of California, Irvine, 1988.

109. Reiss, R.A., N.H., Gartner, and S., Cohen. Dynamic Control and Traffic Performance in a Freeway Corridor: A Simulation Study. *Transportation Research-A*, Vol. 25 A, No. 5, pp.267-276, 1991.

110. Richards, P. I. Shock waves on the Highway *Oper. Res*. pp. 42-51, 1956.

111. Ritchie G. R. A Knowledge-Based Decision Support Architecture for Advanced Traffic Management. *Transportation Research*, Vol. 24A, No. 1, January 1990.

112. Ritchie, G. R., N.A., Prosser. Real-Time Expert System Approach to Freeway Incident Management. *Transportation Research Record*, No. 1320, 1991.

113. Ritchie, S., and N.A., Prosser. 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.

114. Ritchie S. and H., Zhang. Real-Time Decision-Support System for Freeway Management and Control. *ASCE Journal of Computing in Civil Engineering*, Vol. 8 No. 1, January 1994.

115. Stamatiadis C. and L.W., Taylor. 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.

116. Stephanedes Y., E., Kwon, P., Michalapouplos. 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.

117. Sheffi, Y. Urban Transportation Networks, Prentice Hall, Inc., Englewood Cliffs, New Jersey 07632, 1985.

118. Sherali, H.D., K., Ozbay, K., and S., Subramaniam. Time-Dependent Shortest Pair of Disjoint Paths. Working Paper to be Submitted for Publication, 1996.

119. Sherali, H.D., N., Arora, and A.G., Hobeika. Parameter Optimization Methods for Estimating Dynamic Origin-Destination Trip Tables. To Appear in Transportation Research-B, 1996.

120. Sherali, H.D., R., Sivanandan , and A.G., Hobeika. A Linear Programming Approach for Synthesizing Origin-Destination Trip Tables from Link Traffic Volumes. *Transportation Research-B*, Vol. 28B, No. 3, pp. 213-233, 1994.

121. Siegfried, R. H., N., Vaidya. Automated Incident Management Plan using Geographic Information Systems Technology for Traffic Management Centers. Texas Transportation Institute, Research Report 1928-1F, July 1993.

122. Slotine, J. J. E., and Li, Weiping. *Applied Nonlinear Control.* Prentice Hall, 1991.

123. Smith, M.J. A New Dynamic Traffic Model and the Existence and Calculation of Dynamic User Equilibria on Congested Capacity Constrained Road Networks,

Presented at the 71st TRB meeting, Washington D.C., 1992.

124. Smith, M.J. A New Dynamic Traffic Model and The Existence and Calculation of Dynamic User Equilibria on Congested Capacity Constrained Road Networks. Transportation Researc-B, Vol. 2B, pp.49-63, 1993.

125. Smith, M.J., T.V., Vuren. Traffic Equilibria with Responsive Traffic Control. Transportation Science, Vol. 27, No. 2, pp. 118-132, 1993.

126. Srinivas, P., H.S., Mahmassani. Multiple User ClassesReal-Time Traffic Assignment for On-Line ATIS/ATMS Operations: A Rolling Horizon Solution Framewok. Presented at the 74th TRB Annual Meeting, January 22-28, 1995, Washington, D.C.

127. Stewart, J.R., Applications of Classification and Regression Tree Methods in Roadway safety Studies. Transportation Research Board 75th Annual Meeting, Washington, D.C., January 7-11, 1996.

128. Subramaniam, S., Y., Zhang, Y. Wide-Area Incident ManagementExpert-GIS System Development. Project Progress Report submitted to FHWA, July 1994.

129. Subramaniam, S., V., Krishnaswamy, and Y., Zhang. Wide-Area Incident ManagementExpert-GIS System Development. Report of Activities submitted to FHWA, July 1993.

130. 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.

131. Ullman, G.L., C.L. Dudek, and K.N. Balke. Effectof Freeway Corridor Attributes upon Motorist Diversion Responses to Travel Time Information. Presented at the 73rd TRB Annual Meeting, Washington D.C., January, 1994.

132. Van Aerde, M., and S. Yager. Modeling Dynamic Integrate Freeway / Traffic Signal Networks: A Proposed Routing Based Approach. Transportation Research A,

22A, no.6, 1988.

133. Van Vliet, D. SATURN: A Modern Assignment Model. Traffic Engineering and Control, Vol. 23, pp. 578-581, 1982.

134. Virginia Department of Transportation. Norhern Virginia Freeway Management Team: Operating Manual, 1990.

135. Wang, Mu-Han. Modeling Freeway Incident Clearance Times, Unpublished MS thesis, Civil Engineering Department, Northwestern University, Evanston, Illinois, 1991.

136. Wie, B.W., R.L. Tobin, T.L , Friesz. The Augmented Lagrangian Method for Solving Dynamic Network Traffic AssignmentModels in Discrete Time. *Transportation Science,* Vol. 28, No.3, pp. 204-220, 1993.

137. Wie, B.W. Dynamic Analysis of User Optimized Network Flows with Elastic Travel Demand. Presented at the 70th Annual Meeting of the Transportation Research Board, Washington DC, January 1991.

138. Wie, B.W., T.L., Friesz, and R.L. Tobin. Dynamic User Optimal Traffic Assignment on Congested Mulidestination Networks. *Transportation Research-B,* Vol. 24B, No.6, pp. 431-442, 1989.

139. Zawack, D.J., and G.L., Thompson. A Dynamic Space-Time Network Flow Model for City Traffic Congestion. *Transportation Science 21*, pp.153-162, 1987.

Appendix 1: ANOVA Results

A-0: Significance of Incident Type on Clearance Times (CLT)

Significance of Incident Type on Clearance Times
(The whole data set)

Anova: Single Factor

SUMMARY

Incident Type	*Count*	*Sum*	*Average*	*Variance*
Road Hazard	9	242	26.88889	424.8611
Property Damage	401	17086	42.60848	711.0588
Personal Injury	173	8789	50.80347	914.5425
Disabled Truck	14	844	60.28571	6349.604
Vehicle Fire	7	298	42.57143	442.9524
Hazmat	3	732	244	22953
Weather Related	2	165	82.5	9112.5
Disabled Car in Travel Lane	41	1114	27.17073	227.9951

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	149032.8	7	21290.4	22.99285	6.81E-28	2.023825
Within Groups	594464.6	642	925.9573			
Total	743497.4	649				

A-1: Significance of Truck Involvement on CLT (Property Damage Incidents)

Significance of Truck Involvement on CLT
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

Truck Involved?	*Count*	*Sum*	*Average*	*Variance*
No	357	14888	41.82022	631.1563
Yes	44	2183	50.76744	1305.087

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3071.295	1	3071.295	4.372237	0.037164	3.864983
Within Groups	278874.2	399	702.4538			
Total	281945.5	400				

A-2: Significance of No. of Vehicles in Incident on CLT for Incidents without Truck Involvement (Property Damage Incidents)

Significance of No. of Vehicles in Incident on CLT
for Incidents without Truck Involvement
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

No. of Veh.	*Count*	*Sum*	*Average*	*Variance*
1	44	2067	46.97727	632.8134
2	261	10236	39.21839	475.1406
3	40	1775	44.375	252.8045
4,4+	12	817	68.08333	4835.538

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11471.73	3	3823.909	6.313627	0.000351	2.630202
Within Groups	213797.8	353	605.6595			
Total	225269.5	356				

A-3: Significance of No. of Trucks on CLT For Incidents with Truck Involvement (Property Damage Incidents)

Significance of No. of Trucks on CLT
For Incidents with Truck Involvement
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

No_Truck	*Count*	*Sum*	*Average*	*Variance*
1	41	2001	48.80488	1248.611
2	3	190	63.33333	3033.333

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	590.0534	1	590.0534	0.442452	0.509578	4.07266
Within Groups	56011.11	42	1333.598			
Total	56601.16	43				

A-4: Significance of No. of Vehicles on Incident CLT for Property Damage Incidents Involving Trucks (Property Damage Incidents)

Significance of No. of Vehicles on Incident CLT for Property Damage Incidents Involving Trucks
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

Veh. Involved	*Count*	*Sum*	*Average*	*Variance*
1	3	110	55	50
2	32	1573	49.15625	608.5877
3, or 3+	6	140	28	120

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2079.217	2	1039.609	1.929547	0.159938	3.259444
Within Groups	19396.22	36	538.7839			
Total	21475.44	38				

A-5: Significance of No. of Police Vehicles Involved on CLT For Property Damage Incidents without Truck and 1-3 Cars Involved

Significance of No. of Police Vehicles Involved on CLT For Property Damage Incidents without Truck and 1-3 cars involved
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

POLICEVEH	*Count*	*Sum*	*Average*	*Variance*
1	267	10062	37.68539	403.9232
2	54	2466	45.66667	295.5849
3, 3+	24	1550	64.58333	995.471

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	17444.58	2	8722.291	20.43091	4.15E-09	3.022123
Within Groups	146005.4	342	426.9164			
Total	163450	344				

A-6: Significance of No. of Police Vehicles Involved on Incident CLT For Property Damage Incidents without Truck and 4 or more Cars Involved

Significance of No. of Police Vehicles Involved on Incident CLT For Property Damage Incidents without Truck and 4 or more Cars Involved
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

POLICEVEH	*Count*	*Sum*	*Average*	*Variance*
1	6	235	39.16667	734.1667
2, or 2+	6	582	97	7897.2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	10034.08	1	10034.08	2.325028	0.15829	4.964591
Within Groups	43156.83	10	4315.683			
Total	53190.92	11				

A-7: Impact of Wrecker Involvement on Incident CLT for 1-3 Cars Involved and 1 Police Vehicle (Property Damage Incidents)

Impact of Wrecker Involvement on CLT
For 1-3 Cars Involved and 1 Police Vehicle
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

Wreckers Used	*Count*	*Sum*	*Average*	*Variance*
No	190	6664	35.07368	352.9681
Yes	77	3398	44.12987	476.825

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4493.903	1	4493.903	11.56764	0.000775	3.876792
Within Groups	102949.7	265	388.4893			
Total	107443.6	266				

A-8: Impact of Wrecker Involvement on Incident CLT For 1-3 Cars Involved and 2 Police Vehicles (Property Damage Incidents)

Impact of Wrecker Involvement on Incident CLT
For 1-3 Cars Involved and 2 Police Vehicles
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

Wreckers Used	*Count*	*Sum*	*Average*	*Variance*
No	30	1269	42.3	243.5276
Yes	24	1197	49.875	340.8098

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	765.075	1	765.075	2.669895	0.108303	4.026631
Within Groups	14900.93	52	286.5563			
Total	15666	53				

A-9: Impact of Wrecker Involvement on Incident CLT with Truck Involvement (Property Damage Incidents)

Impact of Wrecker Involvement on CLT with Truck involvement
(Property Damage Incidents)

Anova: Single Factor

SUMMARY

Wreckers Used	*Count*	*Sum*	*Average*	*Variance*
No	31	1288	41.54839	426.6559
Yes	33	2220	67.27273	1522.142

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	10577.53	1	10577.53	10.6621	0.001783	3.995893
Within Groups	61508.22	62	992.0681			
Total	72085.75	63				

A-10: Significance of Truck Involvement on Incident CLT for Personal Injury Incidents

Significance of Truck Involvement on Incident CLT for Personal Injury Incidents

Anova: Single Factor

SUMMARY

Truck Involved?	*Count*	*Sum*	*Average*	*Variance*
No	161	8024	49.83851	844.2738
Yes	12	765	63.75	1823.295

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2161.267	1	2161.267	2.382213	0.124572	3.896417
Within Groups	155140.1	171	907.2518			
Total	157301.3	172				

A-11: Significance of No. of Vehicles Involved on Incident CLT for Personal Injury Incidents

Significance of No. of Vehicles Involved on Incident CLT for Personal Injury Incidents

Anova: Single Factor

SUMMARY

No. of Veh. Involved	*Count*	*Sum*	*Average*	*Variance*
1	19	1223	64.36842	1922.023
2	109	5159	47.33028	759.7973
3, or 3+	45	2407	53.48889	807.0737

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5135.542	2	2567.771	2.868721	0.059524	3.049152
Within Groups	152165.8	170	895.0928			
Total	157301.3	172				

A-12: Significance of No. of Injuries on Incident CLT for Personal Injury Incidents

Significance of No. of Injuries on Incident CLT for Personal Injury Incidents

Anova: Single Factor

SUMMARY

NINJUR	*Count*	*Sum*	*Average*	*Variance*
1 or 2	131	6378	48.68702	769.109
3, or 3+	21	1486	70.7619	1744.69

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	8819.496	1	8819.496	9.808306	0.002089	3.904205
Within Groups	134878	150	899.1865			
Total	143697.5	151				

A-13: Impact of No. of Police Vehicles on Incident CLT for Personal Injury Incidents with 1 or 2 Injuries

Impact of No. of Police Vehicles on Incident CLT for Personal Injury Incidents with 1 or 2 Injuries

Anova: Single Factor

SUMMARY

POLICEVEH	*Count*	*Sum*	*Average*	*Variance*
1	73	3070	42.05479	545.4414
2	61	3154	51.70492	756.6448
3, 3+	18	1079	59.94444	1307.703

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5985.264	2	2992.632	4.171153	0.017274	3.056783
Within Groups	106901.4	149	717.4592			
Total	112886.7	151				

A-14: Impact of No. of Police Vehicles on Incident CLT for Personal Injury Incidents with 3 or more Injuries

Impact of No. of Police Vehicles on Incident CLT for Personal Injury Incidents with 3 or more Injuries

Anova: Single Factor

SUMMARY

POLICEVEH	*Count*	*Sum*	*Average*	*Variance*
1 or 2	15	845	56.33333	540.9524
3, or 3+	6	641	106.8333	3278.167

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	10929.64	1	10929.64	8.665572	0.008336	4.380752
Within Groups	23964.17	19	1261.272			
Total	34893.81	20				

A-15: Impact of Total Number of Wreckers and Ambulances on Incident CLT for Personal Injury Incident

Impact of total Number of Wreckers and Ambulances on Incident CLT for Personal Injury Incident

Anova: Single Factor

SUMMARY

WRECKER+AMBULANCE	*Count*	*Sum*	*Average*	*Variance*
0-1	39	1347	34.53846	394.1498
2, or 2+	33	1693	51.30303	597.5928

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5023.782	1	5023.782	10.31255	0.001997	3.977789
Within Groups	34100.66	70	487.1523			
Total	39124.44	71				

A-16: Impact of Total Number of Wreckers and Ambulances on Incident CLT for Personal Injury Incidents with 1 or 2 Injuries and 2 Police Vehicles Involved

Impact of Total Number of Wreckers and Ambulances on Incident CLT for Personal Injury Incident with 1 or 2 Injuries and 2 Police Vehicles Involved

Anova: Single Factor

SUMMARY

WRECKER+ AMBULANCE	*Count*	*Sum*	*Average*	*Variance*
0-2	34	1744	51.29412	984.8806
3, or 3+	27	1410	52.22222	495.5641

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	12.96303	1	12.96303	0.016852	0.897156	4.003979
Within Groups	45385.73	59	769.2496			
Total	45398.69	60				

The vita has been removed from the scanned document