

MX 7226927 8



The development and control of traffic jams caused by incidents in rectangular grid networks

A thesis submitted to Middlesex University
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

PENINA ROBERG-ORENSTEIN

School of Mathematics and Statistics

Middlesex University

September 1997

K2730062

Site HE	MIDDLESEX UNIVERSITY LIBRARY
Accession No.	7226927
Class No.	388.31 R06
Special Collection	

Acknowledgements

I would like to express my special thanks to the Director of Studies, Mr C R Abbess for his constant interest and enjoyable discussions throughout the research and for his incisive remarks concerning the material in the thesis.

Thanks are also due to my second supervisor, Professor C C Wright for his helpful suggestions and enthusiasm sustained throughout the project. The clear and painstaking comments made on the material in the thesis are also gratefully acknowledged.

I want to thank members of the Maths School at Middlesex University who encouraged me throughout the years as well as Psychology staff members for their thought-provoking lunchtime discussions at Enfield.

Finally, many thanks to my husband Binyamin for seeing me through the final stages and to my parents, extended family and friends for their perseverance.

Abstract

Urban traffic congestion is becoming a central issue in transport planning. If the present growth in car ownership and use continues, traffic jams are likely to increase in frequency and extent, particularly within the central areas of major cities. Whilst it is important to study the impact of congestion in the field, there is an urgent need for a fundamental understanding of the causes of congestion and the way in which it propagates. But, although a number of control schemes for controlling traffic congestion exist, no comprehensive rationale for an effective dispersal strategy has been developed.

This research is mainly concerned with the properties of incident-induced traffic jams on rectangular grid networks, and possible measures for preventing and controlling them. The research investigates the underlying structure of such jams using a combination of theoretical and simulation models developed for this purpose. Using these models, gridlock is identified as a crucial stage in the evolution of traffic jams.

However, most conventional traffic management measures aim to increase capacity and hence postpone the onset of gridlock and are unsuitable when gridlock has already set in. This thesis develops several alternative strategies for protecting networks from gridlock and dissipating traffic jams once they have formed. The treatment focuses on the installation of bans at specific network locations. The bans come in two forms: turn or ahead. Turn bans are imposed on selected links to break gridlock cycles at the nucleus of the traffic jam. By contrast, ahead bans are implemented around the traffic jam envelope to reduce input into critical sections of the road.

The control strategies are tested extensively using the simulation model and as a result, some general control principles have emerged. These are not intended to be immediately applicable to real networks since they incorporate some simplifying assumptions. However, they point to certain characteristics of traffic jam growth and dispersal which would not be accessible in any other way.

List of Terms

μ	Expected number of vehicles per source per minute
p	Proportion of turning vehicles
κ	Ahead saturation capacity: maximum number of ahead vehicles able to cross the upstream intersection during a single time slice
α	Proportion of stopline width allocated to ahead queues
σ	Proportion of a link devoted to storage of segregated queues
τ	Cycle time
s	Spillback factor: controls the degree of queue interaction
t_A	Time taken for a capacity restriction affecting the ahead discharge from a link to propagate to the upstream end of the link
t_{LR}	Corresponding time taken for a capacity restriction affecting the turning discharge from a link to propagate to the upstream end of the link
D_n	Fractal dimension of traffic jam in $n \times n$ network
t_g	Time till gridlock
B_i	Nucleus intervention: Block strategy
Q_j	Boundary intervention: queueing version
R_j	Boundary intervention: re-routing version
$B_i Q_j$	Combination: boundary and nucleus intervention
$B_i R_j$	Combination: boundary and nucleus intervention
\bar{w}	Average number of vehicles in the system in the steady state

Contents

Title Page	
Acknowledgements	i
Abstract	ii
List of Terms	iii
List of Figures	ix
List of Tables	xii
1 Introduction	1
1.1 Summary	1
1.2 Background	1
1.2.1 Types of congestion	3
1.2.2 Measuring urban congestion	7
1.2.3 Traffic jams	13
1.2.4 Controlling traffic congestion	17
1.3 The study approach	21
1.3.1 Objectives of the research	21
1.3.2 Research method	22
1.3.3 Contents of thesis	23

2	Congestion modelling - theory and simulation	25
2.1	Summary	25
2.2	Analytic representations of traffic flow	26
2.2.1	Introduction	26
2.2.2	Theories of traffic flow	27
2.2.3	Congestion growth models	36
2.2.4	The need for simulation	39
2.3	Simulation models for congestion studies	41
2.3.1	Introduction	41
2.3.2	A survey of existing traffic simulation models	41
2.3.3	Developing a congestion modelling environment	47
2.4	Conclusions	49
3	An exploratory tool for traffic jam modelling	52
3.1	Summary	52
3.2	The traffic simulation model	52
3.2.1	Introduction	52
3.2.2	Previous work	53
3.2.3	The modelling assumptions	55
3.2.4	The two-way grid network	62
3.3	Model control	67
3.3.1	Model selection	67
3.3.2	Model structure and operation	69
3.3.3	Model parameters	71
3.3.4	Graphical features	74
3.4	Summary statistics	76
3.4.1	The road network's approach to stability	76
3.4.2	Quantifying the severity of a traffic jam	77
3.5	Conclusions	80

4	The development and dispersal of traffic jams on rectangular grid networks	82
4.1	Summary	82
4.2	Introduction	82
4.3	Spatial characteristics of one-way traffic jams	83
4.3.1	Theoretical background	84
4.3.2	Growth pattern of simulated traffic jam	91
4.3.3	Starvation processes	97
4.3.4	Dispersability	108
4.3.5	Implications for control	112
4.4	Traffic jams on two-way grid networks	119
4.4.1	Growth pattern	119
4.4.2	Effect of right-turn bin size on jam growth and decay	125
4.5	Conclusions	129
5	Controlling traffic jams	131
5.1	Summary	131
5.2	Background	132
5.2.1	Objectives for control	132
5.2.2	Types of control	133
5.2.3	Recovery from severe congestion	136
5.3	Counter-measures for traffic jam control	139
5.3.1	Types of intervention	139
5.3.2	Notation and assumptions	142
5.4	Results	145
5.4.1	Performance of intervention strategies	146
5.4.2	The integrated solution	155
5.4.3	The complexity of congestion phenomena	160
5.4.4	Extending the control principles	167

5.5	Conclusions	172
6	Conclusions	176
6.1	Traffic jam development and dispersal	176
6.1.1	Fundamental concepts	177
6.1.2	Control principles	179
6.2	Possible extensions of the research	181
6.2.1	Theoretical studies	181
6.2.2	Practical developments	183
	Bibliography	186
	Published and Conference Papers	194
	Appendices	195
A		195
A.1	Introduction	195
A.1.1	Location and application of steady state	196
A.1.2	Technique for establishing steady state	199
A.1.3	Convergence of system to steady state using random seeds	203
A.2	Summary	203
B		207
B.1	Background	207
B.1.1	Delay - wait totals method	209
B.1.2	Delay - Positive excess queue method	211
B.1.3	Comparison between two methods	212
B.2	The effect of random seeds on the dispersal pattern of a traffic jam	213
B.3	The effect of random seeds on the level of delay	215
B.4	Summary	226

C	228
C.1 Background	228
C.1.1 Queue spillback times	229
C.1.2 Link propagation times - derivation	229
C.1.3 Minimising traffic jam growth	231
C.1.4 Inhibiting gridlock	232
C.2 Fractal aspects of traffic jams	234
C.2.1 Box-counting technique	234
C.2.2 Discussion	235
C.3 Summary	236
D	238
D.1 Introduction	238
D.1.1 Software objectives	238
D.1.2 Userguide structure	238
D.1.3 Software versions	239
D.2 The screen display	239
D.2.1 Traffic directions on the network	240
D.2.2 Road link layout	241
D.3 Movement of vehicles	241
D.3.1 Calculation of spaces	242
D.3.2 Calculating spillback lanes	243
D.3.3 Capacities and obstructions	245
D.4 Filenames in TSN	246
D.4.1 TSN.PAS	246
D.4.2 TSN.PAR	247
D.4.3 TSN.DAT	247
D.4.4 TSN.LIS	249
D.4.5 SCR**.SSC	250

D.5	Running the simulation	251
D.5.1	Basic instructions	251
D.5.2	Change Defaults ? No	252
D.5.3	Default Obstruction ? Yes	252
D.5.4	Change Defaults? Yes	252
D.5.5	Default Obstruction ? No	253
D.5.6	Turn Bans	254
D.6	Quitting, suspending and re-starting the simulation	255
D.7	Summary	255
E		257
E.1	Simulation of Demand	257
E.2	Vehicle progression	258
E.3	Traffic directions on the network	260
E.4	Turn Bans	263
E.4.1	The CORE Algorithm	265
E.4.2	The DIAMOND Algorithm	265
E.5	Summary	272

List of Figures

1.1	<i>The principle of self-generation of urban traffic</i>	4
2.1	<i>The principle of gridlock in cellular automaton theory</i>	35
3.1	<i>Queue storage configuration for a one-way link</i>	57
3.2	<i>Stages in the propagation of a queue along a one-way road link where the left turning discharge is obstructed</i>	61
3.3	<i>Possible configurations for traffic on a two-way link</i>	64
3.4	<i>Queue storage arrangement on a two-way link</i>	65
3.5	<i>Stages of queue propagation along a two-way link where the right turn- ing discharge is obstructed</i>	66
3.6	<i>Model structure: schematic diagram</i>	68
3.7	<i>Graphical representation of one-way 4×4 network</i>	75
3.8	<i>Network's approach to stability</i>	77
4.1	<i>Spatial configuration of Type I jam ($\alpha < 1 - p$)</i>	88
4.2	<i>Spatial configuration of Type II jam ($\alpha > 1 - p$)</i>	89
4.3	<i>Anti-clockwise and clockwise development of gridlock cycles in one-way network</i>	91
4.4	<i>Initial queue formation</i>	93
4.5	<i>Right-turning gridlock cycle at centre of traffic jam</i>	94
4.6	<i>Diamond envelope of one-way traffic jam</i>	95
4.7	<i>Queue propagation paths for a traffic jam in a one-way grid network .</i>	96

4.8	<i>A large simulated traffic jam on a 256×256 one-way grid network . . .</i>	98
4.9	<i>The changing shape of the boundary of simulated traffic jams with $s > 0$</i>	102
4.10	<i>Box-counting algorithm</i>	105
4.11	<i>Logarithmic plots: negative slope of line equals fractal dimension . . .</i>	107
4.12	<i>Dispersal mechanism I: traffic jam decays</i>	109
4.13	<i>Dispersal mechanism II: queue lengths oscillate</i>	110
4.14	<i>Dispersal Mechanism III: traffic jam fails to clear</i>	110
4.15	<i>Effect of segregated queue storage on jam size for various channelisa- tion regimes</i>	114
4.16	<i>Effect of channelisation on time till gridlock: $\sigma = 0.4$</i>	115
4.17	<i>Rate of growth of traffic jam for various s</i>	117
4.18	<i>Spatial characteristics of two-way traffic jam</i>	122
4.19	<i>Spatial characteristics: left-turning cycle</i>	123
4.20	<i>Spatial characteristics: right-turning cycle</i>	124
4.21	<i>Effect of right-turn bin size on rate of traffic jam decay</i>	128
5.1	<i>Application of the 'block' strategy</i>	141
5.2	<i>The performance of control strategies for increasing turning proportions</i>	150
5.3	<i>The performance of B_4Q_6 for various channelisation regimes</i>	153
5.4	<i>Queueing vs re-routing: a comparison in delay times</i>	156
5.5	<i>An integrated control scheme</i>	157
5.6	<i>Operative domain of the integrated control scheme with respect to de- mand, turning proportion and stopline widths (Colour indicates total delay measured in veh-mins)</i>	161
5.7	<i>Snapshot of 'static' traffic jam after treatment</i>	163
5.8	<i>The prolonged dispersal process of a 'static' traffic jam</i>	163
5.9	<i>Traffic jam migration</i>	168
5.10	<i>Traffic jam splits in two parts following the implementation of the in- tegrated control scheme</i>	169
5.11	<i>Decay process of resonating traffic jam</i>	170

5.12	<i>Friction points in grid networks</i>	171
A.1	<i>The gradual approach to stationarity using a random seed</i>	205
B.1	<i>Dispersal Mechanism I: Traffic jam clears</i>	216
B.2	<i>Dispersal Mechanism II: Traffic jam oscillates</i>	217
B.3	<i>Dispersal Mechanism III: Traffic jam grows</i>	218
B.4	<i>The performance of control strategy B_4C_6 for increasing turning proportions and random seeds</i>	223
B.5	<i>The performance of control strategy B_4Q_6 for fixed α, increasing turning proportions and random seeds</i>	225

List of Tables

1.1	<i>Factors contributing to rise in congestion 1961 - 1991</i>	3
2.1	<i>Features and limitations of existing traffic simulation models when applied to study traffic jams in urban networks</i>	49
3.1	<i>Computer simulation model parameters</i>	74
4.1	<i>Proportion of turn spillback in a traffic jam, a: Diamond, b: Octagon, c: Square</i>	101
4.2	<i>Fractal dimension of traffic jam: $0.00 < \epsilon \leq 0.03$</i>	108
4.3	<i>Extent of traffic jam (No. of blocked links at $\tau = \tau_{10}$) with respect to s</i>	117
4.4	<i>Time to gridlock (minutes) with respect to s</i>	118
5.1	<i>A summary of notation used to describe control strategies</i>	143
5.2	<i>Total delay(veh-min): Increasing Demands</i>	148
5.3	<i>Guidelines for clearing gridlock-based traffic jams</i>	173
A.1	<i>Convergence to steady state using different seeds</i>	204
A.2	<i>Time-Series data (Input - Output) for steady state convergence using different seeds</i>	206
B.1	<i>Total Delay incurred under increasing demands using strategy B_3 for various seeds</i>	220
B.2	<i>Total Delay incurred under increasing demands using strategy B_3R_6 for various seeds</i>	221

B.3	<i>Total Delay incurred under increasing turning proportions using strategy B_4R_6 for various seeds</i>	222
B.4	<i>Total Delay incurred for increasing turning proportions using strategy B_4Q_6 with $\alpha = 0.50$ for various seeds</i>	224
D.1	<i>Interpretation of TSN.PAR data</i>	247
E.1	<i>Calculation of space availability in links BT and BF</i>	259
E.2	<i>Actual number of vehicles that are released from link A</i>	260
E.3	<i>Eastbound traffic movements</i>	262
E.4	<i>Westbound traffic movements</i>	262
E.5	<i>Northbound traffic movements</i>	263
E.6	<i>Southbound traffic movements</i>	264
E.7	<i>Implementation of bans</i>	264
E.8	<i>Implementation of CORE Algorithm</i>	266
E.9	<i>East oriented diamond cordon situations</i>	266
E.10	<i>North oriented diamond cordon situations</i>	267
E.11	<i>Starting point for cordon</i>	267
E.12	<i>Matrices for cases 1-4 required for offsets and increments</i>	268
E.13	<i>Matrices for cases 5-8 required for offsets and increments</i>	269

Chapter 1

Introduction

1.1 Summary

This chapter introduces the main ideas associated with traffic congestion with particular reference to incident-induced urban traffic congestion and ways of modelling it. The background section discusses some of the causes of the increase in traffic congestion, distinguishes between different types of congestion and describes ways of measuring urban congestion. The section concludes with a survey of approaches to congestion control. The objectives of the research are then outlined together with a discussion of the research method. The chapter concludes with the layout of the thesis.

1.2 Background

Traffic congestion is a persistent problem in both ancient and modern communities. In urban centres throughout the world it imposes considerable problems on the local economy, it contributes to accidents and atmospheric pollution, and it triggers anxiety, frustration and stress among all classes of road-user.

The development of congestion is related to the way urban networks have evolved.

Many were not originally designed to deal with congestion, nor with its consequences. The problem has become acute because many urban networks are operating at saturation level. This means that any slight perturbation in the system leads to the rapid escalation of queues. The role of the traffic engineer is precarious because experimentation can lead to serious and often unpredictable outcomes. A better understanding of the mechanics of congestion and the way in which it propagates is urgently required, in order to develop remedial strategies for its successful control.

In simple terms, congestion occurs when too many vehicles try to use the limited amount of road space available. Congestion slows the passage of all vehicles, so that each additional vehicle imposes costs on all other vehicles by increasing congestion and slowing traffic flow.

Transport statistics published by the Department of Transport (1992) illustrate clearly how this problem has developed. As can be seen in Table 1.1¹, in 1961, the proportion of households in Britain having regular use of at least one car stood at just 31%, while, by 1991, this had increased to 67%. Meanwhile, the number of passenger kilometers travelled by motorcars and vans increased from 164 billion in 1961 to 590 billion in 1991 - a rise of almost 260%.

Throughout this period, investment in road transport infrastructure has steadily increased. Between 1961 and 1991, the length of roads increased by 15%. This, however, has been insufficient to meet the increased demand for travel and has resulted in worsening congestion, Mogridge (1990).

Heran and Tostain (1995) attribute the rise in car use and travel to a process of 'self-generation'. This can be summarised as follows. The availability of the car increases drivers' tolerance of growing distances between the places in which they lead their lives. This increased tolerance contributes to increased dispersion and makes public transport, bicycles and walking less attractive. As a result, it becomes more necessary

¹Source: Transport Statistics Great Britain 1992

	Year						
	1961	1965	1970	1975	1980	1985	1991
Proportion of households with regular use of one or more cars (%)	31	41	52	56	59	63	67
Passenger kms travelled by car and vans (billion)	164	236	299	333	389	443	590
Length of road (all classes) (thousand kms)	314.2	323.6	322.5	330.0	339.6	348.7	360.0

Table 1.1: *Factors contributing to rise in congestion 1961 - 1991*

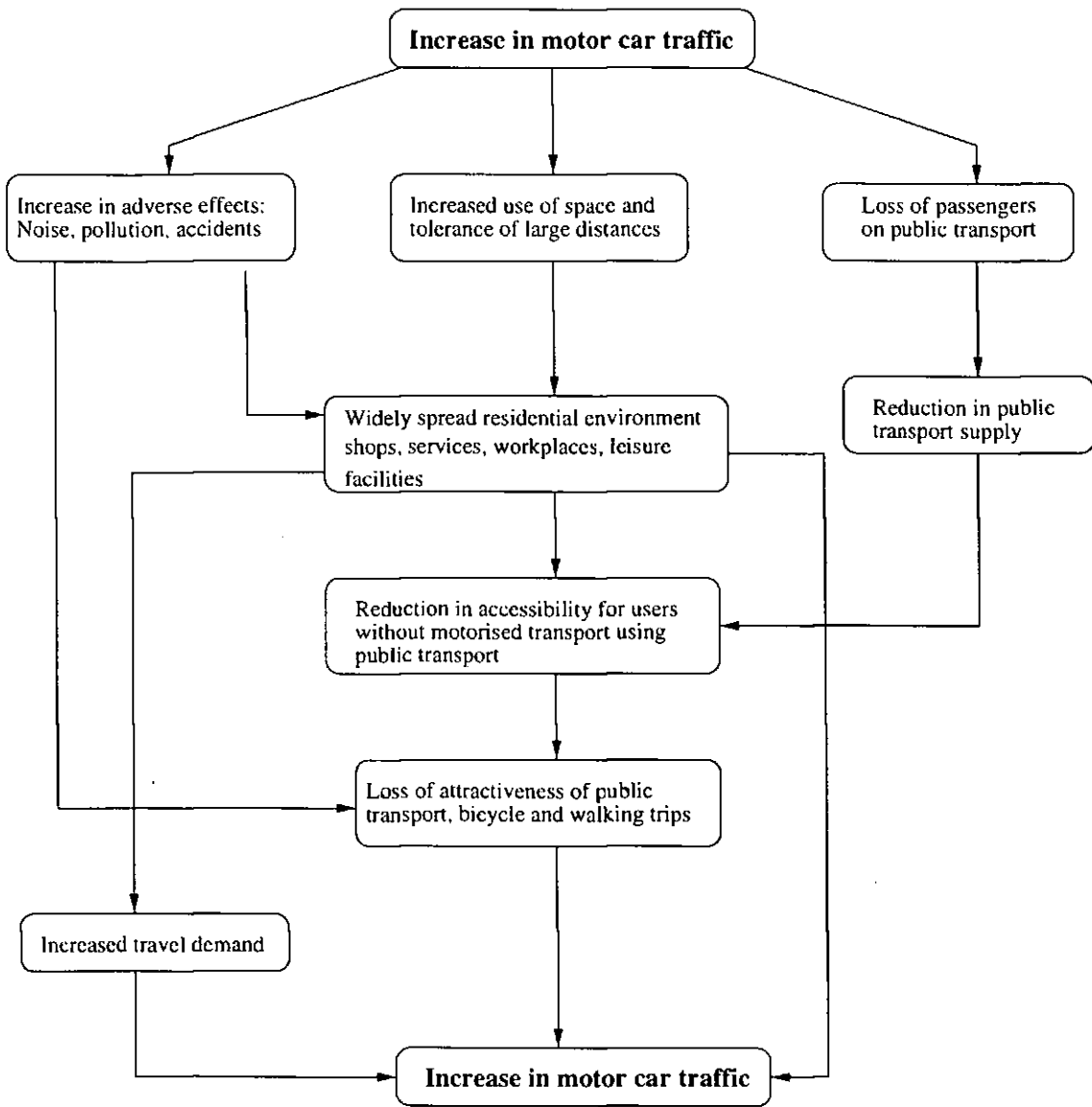
to travel by car. This chain of events represents a self-generating process of urban traffic which is shown in Figure 1.1.

The continuing expansion of road traffic when coupled with the lack of road space leads to increased levels of traffic congestion. It is unlikely that the situation will improve considerably unless a integrated strategy is formulated.

1.2.1 Types of congestion

The underlying cause of congestion is quite simple: when traffic demand exceeds capacity, either locally or over an area, then queues form which cause delays to vehicles - the greater the demand, the larger the queues, and hence the longer the journey times.

It is important to distinguish between two types of traffic congestion: recurrent congestion and incident-induced congestion, because the latter is unexpected and can cause the rapid build-up of queues on top of the usual daily queues which are a feature of recurrent congestion. In addition, with recurrent congestion, the system is still operational, and the queues are moving albeit slowly. By contrast, with incident-induced congestion, the system has irretrievably broken down at a particular point



Source: Heran, 1995

Figure 1.1: *The principle of self-generation of urban traffic*

and this effect is transmitted through the network. The rapid build-up of queues on top of the usual queues cannot be processed and eventually the system grinds to a halt.

A familiar example of recurrent congestion is the slow crawl of vehicles along busy shopping streets. Queues that form as a result of an accident are examples of incident-based congestion.

Incidents, because of their unpredictability, are likely to increase the risk of breakdown of the orderly movement of the traffic in the vicinity of the incident, and if they occur when there are already busy traffic conditions, may trigger levels of congestion which are out of proportion to the severity of the incidents themselves.

Traffic queues

One of the most acute forms of traffic congestion, however, is the area-wide traffic 'jam', in which vehicles can become embedded for sustained periods. The queueing mechanism is identified by Wright and Huddart (1989b) as fundamental for understanding the nature of such jams, and how they propagate. Queueing occurs when the demand placed on a road system, exceeds its capacity. The cause may be due to the following:

- external sources of "friction" such as crossing pedestrians and parked vehicles;
- internal sources of "friction" such as vehicles turning into and out of side streets and access ways, and slow vehicles such as heavy lorries, buses and cyclists;
- isolated pinch points where the physical layout of the road effectively reduces capacity to a value less than that for neighbouring locations: typically these occur at narrow sections of road;
- instability in the traffic stream as flow approaches capacity.

All these can be referred to as "bottlenecks", although the last is demand related, while the others are due to supply effects. On any road containing a bottleneck, traffic speeds tend to fall as the flow increases. The effect can be analysed as a queueing mechanism: passage through a bottleneck is pictured as a "service" which requires a given amount of time per vehicle. If the upstream flow varies randomly from moment to moment, there will be periods when more vehicles arrive than can be serviced, and these vehicles must form a queue. During other periods the demand will be less than the service rate, and the queue present will diminish.

Over a long period of time, the average queue length will be determined according to the ratio of flow to capacity. As this ratio approaches unity, the number of vehicles in the queue (and hence the average delay per vehicle) will increase without limit.

Bottlenecks may be difficult to isolate; there may be temporary obstructions which occur at different times and places on different days. These include crossing pedestrians, parked and loading vehicles, turning traffic and driver behaviour.

In these circumstances, traffic cannot be represented as a simple queueing mechanism with only a single server. It is usual to express traffic behaviour in terms of empirically determined relationships between average speed, flows and concentration.

Note that queueing theory assumes that the service characteristics are not affected by changes in demand. At high levels of demand, movement through a bottleneck may break down. The flow or service rate may actually fall below the level previously sustained through the bottleneck.

Two situations where these effects have been observed are described by Wright and Huddart (1989b). The first example occurs in saturated motorways where vehicles encounter a stopping shockwave, caused by sudden braking of one driver where vehicles travel close together. This reduces the traffic stream to a slow crawl. Vehicles that eventually clear the front of the queue trigger a 'starting' wave which also propagates from vehicle to vehicle within the traffic stream. Because there is a lag before

each vehicle accelerates, the flow discharging from the front of the queue will be less than the flow arriving at the rear. The capacity of the lane therefore, is effectively reduced.

The second case arises on narrow two-way urban streets. If one direction is obstructed by a parked vehicle, the unobstructed stream must occasionally give way so that the opposing stream can pass. Suppose a second vehicle parks nearby. There will now be two bottlenecks, each of a single lane width, with a pocket of clear two-way road in between them. Now, if vehicles from the two traffic streams enter these two bottlenecks simultaneously, the system will "lock-up" so that nobody can move.

In both examples, the capacity of the road is reduced and the average speed and flow fall to zero or nearly to zero. When this occurs, traffic enters into a 'congested-flow' regime. The transition from free-flow to congested-flow is unstable: it cannot be reversed without some form of external intervention. In the case of the saturated motorway lane, this would mean reducing the supply of vehicles, whereas in the case of the urban street, one or more vehicles would have to reverse out of one of the bottlenecks. Both speed and flow are limited by the physical characteristics of the bottlenecks.

1.2.2 Measuring urban congestion

Over the past few decades, much time and effort has been spent in developing methods and algorithms for measuring traffic congestion in the motorway environment. For example, Lindley (1987), defines congestion using the Highway Capacity Manual level of service concept. The limit between levels C and D (ie, average speeds of 54mph and a v/c ratio of 77) defines the conditions at which congestion starts to occur. But, this definition is not applicable to urban networks because of the interactions between intersections and road links.

Even so, many of the lessons learned from studying motorway congestion carry im-

plications for the urban equivalent, particularly in the context of measuring and automatically detecting incident-induced traffic congestion.

However, in this respect, the urban street environment is more challenging than the motorway counterpart because traffic flow discontinuities are introduced by traffic signals, traffic entering and leaving side streets, variations in signal timing and geometric characteristics. In particular, there are a number of clear distinctions between congestion that arises in motorways and that which occurs in urban streets, Leonard (1992). These can be summarised as follows:

1. Urban travel takes place on a road network which forces vehicles to interact at junctions. Because of the linking between junctions a queue at one junction can extend back to an upstream junction and block the free movement of traffic into that junction.
2. The network offers the driver a choice of route, enabling them to bypass areas of congestion with varying degrees of success.
3. Limited link lengths coupled with the types of junction control lead to a rapid escalation of congestion to upstream junctions and throughout the network.
4. Travel demand is varying and strongly peaked.

In addition, the complexity of street networks can also contribute to severe congestion directly, because it typically occurs when traffic patterns are conflicting and complicated.

Consequently, *urban* congestion is less straightforward than its motorway equivalent, particularly because free-flow conditions differ between cities, and junctions and links do not operate independently. The definition of urban congestion cannot be confined to variables such as speed but should include some reference to the key aspects of urban congestion, namely, queues blocking back to upstream junctions, or the main effect on drivers' decisions through increases in travel time.

Blocking back

Blocking-back (or queue interaction as it is sometimes known) occurs when a queue (or queues where there are multiple streams of traffic) of vehicles on a link extends back to the previous junction thereby 'blocking' free access to the link from the upstream links. The net effect is the reduction in the throughput capacity of the upstream links for as long as the blocking-back condition persists.

One can usefully discriminate between spillback that prevents vehicles turning into the *first* blocked link (primary spillback), and spillback that prevents *other movements* (secondary spillback). It follows that the spread of congestion is closely related to the way blocking-back occurs between affected links, and should therefore be taken account of when measuring the extent of urban traffic congestion.

But, most existing traffic models have various shortcomings in their ability to predict the behaviour of traffic at closely-spaced intersections, especially in accounting for the interaction of downstream queues and upstream capacities, see Johnson and Akcelik (1992) for a detailed review. To overcome these deficiencies, Roupail and Akcelik (1992) developed a queue interaction model for a single traffic stream. The results indicate that at intersections with limited queueing space, the presence of downstream queues effects the system performance both in terms of delay and capacities. Similar conclusions were drawn by Liu (1991) who examined the finite storage capacity of a link by developing a model to describe the dynamic congestion patterns for a single traffic stream passing through a signalised intersection.

However, a common characteristic of urban intersections is the existence of a number of movements at the upstream end of a link discharging into a common lane or lanes on a downstream link. The capacity of these various upstream movements may be limited by the storage capacity of the road linking the two intersections. The procedure described by Prosser and Dunne (1994) addresses the problem of estimating the capacities of these upstream movements. But the model relies on a

somewhat crude condition for determining whether blocking-back has occurred. The downstream queue is considered to interfere with upstream movements only when the storage capacity of the downstream link is full. The authors propose a possible refinement which implies a secondary storage capacity measure which could be used to model the complex behaviour of vehicles using the upstream intersection.

The blocking-back process is thus more complicated where multiple streams of traffic are involved. Effectively, the mechanism can be divided into two distinct phases. Initially, there is some interference between the traffic streams on a specific link. The severity of interference determines the degree to which upstream links are affected. The second stage of the blocking-back process occurs when blocking effects are transmitted to upstream links. The entire process then repeats itself.

The interference between traffic streams on a link arises as a result of the dependencies between them. When vehicles discharge into various exits from a road link their paths will interact because (a) vehicles making different turning movements may share the same traffic lanes, and (b) even if there is a separate turning lane or lanes for each exit, drivers do not necessarily position themselves in the correct lane at the entrance to the link under consideration and there will be a certain amount of weaving between the entrance and the exit during which the different turning movements may interfere with one another's progress.

Consequently, if a particular exit (say, the left-turn exit), is blocked, vehicles intending to turn left will form a queue that spills back along the link. If the lane discipline of drivers at the various points of entry to the link is not perfect, the queue may eventually spread across the other lanes and block all the traffic.

Very little is known about the mechanism of interaction between queues in practice, although work is currently being conducted in this area, see Allsop and Bell (1995). Preliminary results from the study, which used video cameras to examine how drivers at upstream junctions responded to blocking-back, showed that blocking-back affected

the capacity of entering lanes through reductions in the saturation flow. Although the authors are now conducting a more detailed statistical analysis to estimate these effects lane by lane, the video recordings and the data from the study offer scope for further behavioural as well as quantitative analysis.

For modelling purposes, we shall picture a road link as divided into two distinct zones: a downstream queue storage area where vehicles are organised into separate turning movements, and an upstream 'reservoir' where the turning movements are mixed. The interactions occur at the transition between these two zones. The degree of interaction can be altered to suit different traffic conditions. If drivers are well disciplined, then the degree of interaction will be marginal.

By making these assumptions, we hope to be able to study the blocking-back mechanism and its contribution to the spread of traffic jams. The extent to which the assumptions reflect reality will need to be analysed. In addition, we will have to assess the validity of the results if the assumptions are relaxed. The sensitivity of the model to its underlying assumptions will be discussed in Chapter 4.

It is useful to describe the development of congestion as a sequence of events, each one leading to the other. Having established that the onset of congestion coincides with the blocking effects experienced along a single link, we now examine the other stages which form part of this process.

Stages of congestion

There are a number of ways of describing the stages of urban traffic congestion. One method, Pignataro et al (1978), considers the realm of congested operations from the traffic signal viewpoint. Congested operations are divided into two sub-categories: saturated operations and over-saturated operations. Saturated operations is a term that describes that range of congestion where queues form at various junctions, but their adverse effects on the traffic in terms of delay are limited to the affected link. Over-saturated operations occurs when queues grow to a point where traffic operation

in upstream links are adversely affected. This disruption is described as blocking-back, which can produce primary and secondary adverse effects. Primary blocking affects traffic entering a blocked link whereas secondary blocking affects other traffic movements in upstream junctions.

Another way is to divide the stages in terms of the state of the evolving queues. For example, Longley (1968) and Quinn (1992) distinguish between primary and secondary congestion. *Primary congestion* is caused by the development of queues at controlled junctions. These queues spill back and interfere with flow through junctions upstreams. This is known as *secondary congestion*. Eventually, the queues encircle the original source of congestion, which becomes embedded in the network. This last phase is associated with 'gridlock' and represents a particularly serious case of degeneration which should be avoided, where possible. This is because once a gridlock cycle has formed, it becomes more difficult to clear the resulting congestion.

Both definitions share a common aspect in that they refer to the concept of blocking-back introduced earlier. However, there is another factor to consider when estimating the impact of urban congestion - increases in travel time. This issue is addressed by Leonard (1992) who provides a useful classification of congestion using a medical analogy.

Congestion is thought as the breakdown of the orderly movement of traffic. This definition forms the focus of a 'congestion-scale' which distinguishes between 'busy' traffic and conditions where there is so much traffic that the orderly movement of the traffic has started to break down. The most acute form of congestion is referred to as 'gridlock' and this occurs when many junctions are linked together into one large group blocking-back. Traffic is then virtually at a standstill over a large proportion of the network. This stage is described as an area-wide 'jam' in which vehicles can be trapped for long periods.

We are concerned with this last stage of congestion: the area-wide 'jam'. In particular,

we consider the sources of such jams, the way in which they propagate, the factors that influence their severity and their impact on the sustainability of modern urban centres.

1.2.3 Traffic jams

There are a number of ways in which a traffic jam may start. These include

1. A temporary obstruction may occur within a road link or at a junction. This causes a reduction in the capacity for movement for some (if not all) of vehicles using the link or junction which is below the corresponding level of demand.
2. The average demand for movement at some point may exceed the value that can be accommodated within the unobstructed capacity of the road at that point; in other words, traffic flow is constrained by a capacity bottleneck in the network itself.
3. An upward stochastic fluctuation in the rate at which drivers wish to access a particular sector of the network may lead to queue propagation over a wide area, even though the average demand is below capacity everywhere within the system.

We are concerned only with jams arising from a single source, that might be categorised under (1) or (2). Whichever of these applies, one would expect the subsequent process of jam development to be the same.

Starting from an initial obstruction located somewhere in the network, we visualise a traffic jam as propagating from link to link via a process of branching, each queue generating new branches at each junction. This propagation of queues through the network happens as a direct consequence of the crossing of paths that occurs at the junctions in the network.

Initially, the topology of the branches resembles that of a simple tree, but at some stage, a queue will tail back around four sides of a block to form a closed circle. This

represents the onset of gridlock. In practical terms, gridlock represents a significant stage in the evolution of a jam because once it has occurred, it is difficult to clear the jam even if the demand falls below the theoretical capacity of the obstructed network. The conflicts that arise between traffic streams competing for the limited road space are irresolvable without some form of external intervention.

The influence of incidents on traffic jam severity

One of the major causes of traffic jams is the disruption caused by incidents. An incident is any event which causes a reduction in capacity. Incidents occur in a variety of forms, such as road traffic accidents, vehicle breakdowns and roadworks. Initially, one needs to identify the cause of the incident and then implement some form of corrective action in order to handle the incident's immediate effects.

Technological advances in surveillance techniques have made incident identification easier than in the past. Possible techniques include the use of close circuit video cameras, infra-red technology, and cellular phone callers. The next stage is to handle or remove the incident or its direct impact. Where necessary, this implies removing a vehicle from the roadside and coordinating the actions of the responsible agencies (eg police, fire department, emergency medical services, etc).

A key feature of many incidents is that they are random, unplanned events and because of this, they can have an impact on congestion which is out of proportion to the original incident. The severity of the ensuing congestion can be exacerbated further by high levels of traffic, typically during morning and evening peak periods.

There are a number of factors which influence the severity of incident-induced congestion. These are the frequency of occurrence, the time of day in which the incident occurs, the duration of the incident, the level of background congestion and the location of the incident.

The duration of an incident, whether predictable (eg caused by roadworks), or un-

predictable (caused for example by an accident), is an important contributor to the severity of the ensuing congestion. Swift action is needed to remove an incident in order to minimise the amount of congestion incurred. Again, this result is confirmed in Holmes and Leonard (1992) and in Mongeot (1993), who show that the severity of congestion increased proportionately with the length of time before the incident was removed.

A study conducted by Holmes and Leonard (1992), found that during a six-month period in 1991, on average, about twenty significant traffic incidents were recorded daily for the London area. The severity of the resulting congestion depended on a number of factors, in particular the level of background traffic at the time of the incident and its duration.

The severity of congestion at a particular location is also linked to the characteristics of a network (eg link lengths, queue storage arrangements and junction control), the level of demand in the area and the type of control that can be implemented there, Van Vuren and Leonard (1994).

Huddart and Wright (1989) identify three features of link arrangement which can delay the onset of gridlock in oversaturated conditions. These are the allocation of separate turning lanes with high stopline saturation flows, adequate storage areas within the turning lanes and good lane discipline so that the interference between movements is reduced.

The level of demand and the demand pattern in an area will also influence severity. For example, an incident on a major or trunk road gives rise to more severe congestion because of the sheer volume of traffic carried by these roads. This was shown by Holmes and Leonard (1992) who demonstrated that the severity of incident-induced congestion increased with the background level of traffic, as might be expected. Thus, if an incident occurs when there are already busy traffic conditions, it may trigger levels of congestion which are out of proportion to the severity of the incident itself.

By contrast, a severe incident which occurs in light traffic conditions may not cause extensive damage because of the insignificant level of background congestion.

The time of occurrence of an incident coupled with the network characteristics in the incident vicinity are also important factors especially when determining the suitability of control strategies. For example, PM peak congestion in central areas does not allow for queue displacement strategies, as insufficient road space will be available for the displaced queues.

In addition, the location of the incident in the link itself affects passing opportunities, the possibility of removing the incident and the onset of secondary congestion, due to reduced storage capacity upstream of the incident.

The impact of traffic jams on urban sustainability

If the present growth in car ownership and use continues, traffic jams are likely to increase in frequency and extent, particularly within the central areas of major cities. It has been hypothesised that unless something is done, large traffic jams will effectively paralyse city centres for long periods of the day.

The potential for traffic jams to occur in London has been analysed by Mogridge (1988) who has evaluated traffic flows in the city over the past twenty years. London traffic has reached saturation point and it is only a matter of time before a relatively small accident sparks off an uncontrollable traffic jam.

Two events have been reported to support this claim, Wright and Huddart (1989b). An accident at Blackfriars Bridge in October 1988 was said to have brought 50,000 vehicles to a halt. On another occasion, the simultaneous closure of three Central London bridges resulted in severe congestion, although on that occasion the queues were observed to be moving steadily, if slowly. It has also been claimed in the media Rufford (1987), that, for example, an evening peak accident at the junction of Earls Court Road and Cromwell Road could cause extensive queuing eastwards to Marble

Arch and southwards to the River Thames within 40 minutes. There is also evidence of some road networks being particularly sensitive to gridlock phenomenon, for example the Victoria area in Paris, Seemama (1995). The complexity of urban networks often contribute directly to the formation of gridlock because traffic patterns using these streets are complicated and conflicting.

It is unclear at present whether the examples mentioned above are simply freak events, or whether they are symptoms of imminent collapse. With the continuing growth of travel demand over the past few years, severe area-wide traffic jams occurring in London and ultimately in other city centres, is a serious possibility.

1.2.4 Controlling traffic congestion

In many cases of traffic congestion, the problem can be traced to a single source. Initially, this may be a bottleneck on a major route, which generates a queue upstream. If the bottleneck cannot be eliminated, it may be possible to prevent the onset of secondary congestion, by regulating the input into this queue and hence avoid propagation to other streams of traffic. We shall refer to this queue as the *critical queue* since it is the first event in a chain which ultimately leads to the breakdown of orderly movement in the network.

Consequently, the first task in controlling area-wide congestion is to identify those road links or groups of links in a network, where critical queues form, and where queue control is most desirable. Possible measures for queue control include queue regulation, capacity improvements and demand restraint. Queue regulation can be achieved by regulating the traffic input into the tail of a critical queue in order to control the queue length. This type of control can also be implemented over a wide region of critical links. By restricting the flow of traffic entering a busy area it is possible to minimise the likelihood of critical queue formation elsewhere.

Objectives for congestion control

In the past, control measures aimed to minimise delay for all vehicles using the road network. However, with increasing concern for environmental issues the range of possible objectives for urban traffic control strategies should be reconsidered. A range of criteria for congestion control have been summarised by Huddart and Wright (1989), as follows:

- Minimise overall delay to vehicles
- Minimise overall delay to public transport services
- Minimise delays to emergency services
- Minimise delays to pedestrians
- Equitable distribution of delays between competing streams of traffic
- Maximise reliability, ie minimise unpredictable variations in journey time for vehicle users
- Maximise network capacity
- Minimise accident potential for all users
- Minimise environmental impact of vehicular traffic (noise, atmospheric pollution, and possibly visual intrusion)
- Energy efficiency

It is important to note that some objectives might conflict and a compromise may have to be sought in the selection of objectives.

However, some criteria may be met in tandem, for example minimising delay to vehicles would contribute to a reduction in fuel consumption, and thus to less atmospheric pollution and increase the network throughput.

Alternative approaches to congestion control

The effects of traffic congestion may be mitigated by implementing traffic control strategies. Van Vuren and Leonard (1994) divide these strategies into two categories depending on whether they are exercised on the supply side (road and intersection capacity) or the demand side (travellers). For the former, the available capacity of the network is utilised to minimise the impact of the incident. In the second type of strategy, information is supplied to drivers so they can avoid the area where the incident has occurred.

Most of the research on traffic control strategies has concentrated on the supply side of the equation. For example, Van Vuren and Leonard (1994) studied several strategies to dissipate congestion, and found that traffic signal optimisation was one of the most effective, regardless of whether drivers were re-routed or not. Signal optimisation downstream of the incident also proved to be beneficial in dispersing the traffic jam.

Other forms of 'supply' control measures include capacity improvements. For example, the elimination of bottlenecks can be a useful measure in parts of the network where demand regularly exceeds capacity. This can be achieved through junction improvements and road widening schemes which can reduce critical queue lengths to an acceptable level, eventually leading to an efficient use of the available network capacity. However, such measures are expensive, and not always feasible. Many bottlenecks are associated not so much with the physical layout of the network, but with pedestrian and commercial activities which spill into the carriageway. Measures to control parking, loading and pedestrian crossing movements are frequently applied, but they are inconvenient to users.

Re-configuring the traffic routing plan within a city can also increase its capacity. For example, a one-way system can re-distribute its conflicts and exploit unused stopline capacity on adjacent streets.

However, any traffic control strategy which relies on capacity expansion will tend to promote further traffic growth. Consequently, it may postpone the problem - and even make it worse - rather than providing a long term solution.

On the demand side, Van Vuren and Leonard (1994) found that the usefulness of a diversion strategy strongly depends on the level of demand on the rest of the network, and on the route selected. Their results show that re-routing can be harmful if the level of background congestion is already high, or if several incidents occur simultaneously. Regarding this strategy, the authors conclude that traffic should be diverted before the stationary queues are met and that travel time on the anticipated alternative routes should anticipate the travel times with the increased demand, and that under congested conditions, diversion should be to multiple routes.

Demand restraint also includes the use of metering or gating techniques. These can be "static" - fixed or pre-planned measures, or "dynamic" - measures which respond to variations in traffic flow over time.

However, static signal control plans are restrictive since they are mainly designed for fixed-time operation, in which the control algorithm changes at preset times to suit the anticipated traffic demand. But if the algorithms are applied too late, or if unexpected congestion occurs, queues build up quickly leading to catastrophic collapse. This suggests the need for dynamic or on-line control which responds to changes in the traffic situation.

Dynamic control measures may take on various forms. These include local-scale measures as well as area-wide measures. Local measures apply to a single intersection or pairs of intersections, whereas area-wide measures extend to cover larger sections of the urban road network.

Rathi (1991) develops a dynamic control scheme using metering strategies. These have successfully been applied to a central district in New York. However, care should be exercised with the modelling of the boundary areas, as the effect of metering may be

displacement of the congestion to other areas. This will be discussed in Chapter 5.

Another example of dynamic control is proposed by Shepherd (1994) who describes an automatic gating technique which may be employed in cases of both recurring and incident induced congestion. However, the strategies can only be used to reduce the effects of blocking-back thereby preventing the onset of gridlock. But once gridlock occurs, the 'auto-gating' principle fails to remedy the situation and the system breaks down irretrievably.

This is a hallmark of many traffic management measures, because whilst most can increase capacity and hence postpone the onset of severe congestion, they are not suitable once gridlock has set in. This is the realm of traffic congestion with which this thesis is particularly concerned.

1.3 The study approach

1.3.1 Objectives of the research

Traffic jams can be observed in many modern urban centres, yet the underlying mechanism is not well understood. Whilst it is important to study the impact of congestion in the field, there is an urgent need for a fundamental understanding of the causes of congestion and the way in which it propagates. But, although a number of control schemes have been developed to curb traffic congestion, no comprehensive rationale for an effective dispersal strategy has yet been developed.

This research is mainly concerned with area-wide traffic jams (as opposed to queues on isolated sections of road), and possible measures for preventing and controlling them. The research investigates the underlying structure of such traffic jams using a combination of theoretical and simulation models. The models are later applied to idealised, rectangular grid networks. This enables us to identify those features of traffic jams which would apply independently of the underlying network configuration.

The approach sacrifices a good deal of realism, and the results are not intended to be immediately applicable to real networks. However, they yield qualitative insights which may lead to a better understanding of the congestion problem in general terms. In particular, they point to certain characteristics of traffic jam growth and decay which would not be accessible in any other way, and that can in principle be tested.

Specifically, the primary objectives of this research initiative are:

1. to investigate traffic jam formation, growth and dissipation on rectangular grid networks,
2. to test the sensitivity of the model results to the assumptions on which the model is based,
3. to assess the potential of possible counter-measures to deter traffic jam growth, and
4. to suggest effective mechanisms for the controlled dispersion of gridlock-based traffic jams.

The realisation of these objectives will lead to a working theory which can explain the way traffic jams develop and disperse on rectangular grid networks.

1.3.2 Research method

The method of research combines theoretical analysis with simulation modelling. The study takes place in two stages. First, we consider several alternative approaches to congestion modelling. This leads us to the preliminary study carried out by Abhess and Wright (1991) on the spread of congestion on one-way rectangular grid networks. The results of this study are analysed further and some qualitative conclusions are drawn about the structure of traffic jams, and their sensitivity to model assumptions.

In the next stage, we develop a research tool which simulates the development and dispersal of traffic jams on rectangular grid networks in much less than real time,

allowing a wide range of parameter variations to be explored at a reasonable cost in terms of computer resources. The model incorporates features such as stochastic perturbations in the demand pattern and facilities for developing and testing remedial control strategies.

The simulation tool is used to validate the results of the analytic study, but in addition, it points to certain features of traffic jam growth and decay which cannot be predicted using the theoretical model. In particular, it acts as a device for the development and testing of remedial control strategies. The simulation model is thus used to formulate a working theory which explains the mechanism of traffic jam growth and dispersal in one-way and two-way rectangular grid networks.

1.3.3 Contents of thesis

The thesis is divided into a number of chapters. The first chapter describes a typology for congestion and ways of dealing with it. This is followed by the research aims and the research method.

Chapter 2 considers some alternative representations of traffic congestion and describes how these relate to the modelling methodology adopted later in the study.

The simulation tool which has been developed for one-way and two-way grid networks is introduced in Chapter 3. The main features of the tool are also discussed in this chapter.

Chapter 4 investigates the main characteristics of traffic jams on grid networks via the direct application of the simulation model. The results are subjected to detailed sensitivity analyses and some general principles about the behaviour of traffic jams in rectangular grid networks emerge.

This leads to Chapter 5 which considers how such traffic jams can be controlled. This is achieved by developing and experimenting with various forms of intervention.

The control mechanisms are applied to one-way traffic jams. Thereafter follows a discussion of how the mechanisms could be adapted to treat traffic jams in other types of network environment.

The conclusions of the thesis are brought in **Chapter 6** along with the implications of the results to future work.

The thesis is accompanied by five appendices whose contents can be summarised as follows. In **Appendix A**, we describe the mathematical background for locating the steady state of the simulation tool along with some experiments concerning the stochastic nature of the software. This leads to **Appendix B** which provides the algorithms which have been developed to measure the total delay in the road network. The algorithms are supported by further simulation analysis. **Appendix C** deals with the analytical modelling of traffic jams and derives some results concerning traffic jam minimisation and gridlock inhibition under certain circumstances. In **Appendix D**, we provide a USERGUIDE to accompany the simulation software described in **Chapter 3** of this thesis. The final appendix, **Appendix E**, summarises the main algorithms which have been implemented in the traffic model. These are provided in the form of truth tables.

Chapter 2

Congestion modelling - theory and simulation

2.1 Summary

Over the past few decades, traffic modelling has evolved along two paths: theory and simulation. Theoretical models concentrate on the mathematical analysis of a specific problem which is often stated in simplified terms. Computer simulations are employed when the dynamics of the problem are too complex for the mathematical approach to handle. The two approaches can thus be seen as complementary processes.

The aim of this chapter is to identify a dynamic model for the study of traffic congestion in an urban environment which can deal with the spatial characteristics of queues and the interaction mechanisms operating within them. We first describe various approaches to traffic congestion modelling. Some examples of analytic representations are provided, which apply particularly to the spatial modelling of traffic queues. The need for simulation is then identified. This is followed by a brief discussion of the usefulness and availability of conventional simulation models for representing congestion. The chapter concludes with some general requirements for modelling traffic jams in urban areas.

2.2 Analytic representations of traffic flow

2.2.1 Introduction

The representation of traffic flow in transportation networks is a complex task because of the numerous interactions which govern the system as a whole. During recent years several approaches have been applied to solve specific traffic problems. These include single-lane flow studied in terms of "follow-the-leader" theory, multiple-lane flow handled by stochastic or statistical theory, the application of queueing theory to highway crossing and merging problems, and general intersection and network problems.

Traffic flow theory has concentrated largely on one-way flow on isolated roads because in these situations, traffic flow is simplest. The mathematical analysis is manageable because everyone is travelling the same way and there is no interference from crossing traffic. When the problem of traffic in a city's network of streets is considered, mathematical analysis gives way to computer simulations, in order to handle the interaction between the large number of variables.

The modelling of road traffic was largely guided by the statement of the user equilibrium condition due to Wardrop (1952), which can be expressed as follows:

(For each origin-destination pair) the journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.

This statement led to the development of a new subject in traffic theory - that of equilibrium assignment methods for transportation networks. A comprehensive review on the subject may be found in the book by Sheffi (1985). The work deals with static assignment problems, in which the demand for travel, expressed in the origin-destination matrix, is fixed for the modelled period, and so are the network supply parameters, such as the signal settings. But in real traffic networks, traffic

behaviour is dynamic - the volume of traffic varies in response to network conditions and there are also regular fluctuations corresponding to peak and off-peak requests for travel. Dynamic traffic assignment plays an important role in modelling the formation and dissipation of congestion in oversaturated networks and hence in developing and evaluating traffic management strategies.

In this section, we consider several traffic models and discuss their suitability for representing traffic congestion in urban networks. The discussion highlights a need for the study of traffic jams in their own right. Two possible models for traffic jam growth are proposed. The section concludes with the relative advantages and disadvantages of the analytic approach in modelling traffic jams in urban networks.

2.2.2 Theories of traffic flow

Traffic models perform various functions depending on their individual aims and objectives. The models can be static or dynamic in nature. Static models assume that the demand for travel is time independent. In real traffic networks, travel demand changes in response to events. There are also regular, cyclic changes in demand, for example peak and off-peak conditions. Hence the introduction of a new generation of traffic models called dynamic traffic assignment models. These models have been developed to represent the changes or variations due to transient peaks in demand for travel or temporary changes in capacity. This makes them particularly suited to modelling the effects of congestion over a given time period.

Traffic models strive to find an acceptable balance between simplicity and realism. Highly detailed models have the advantage that they offer plausible descriptions of traffic flow, including the propagation of any congestion in the network. On the other hand they have the disadvantage of being both analytically and computationally demanding.

When formulating a model for traffic flow, it is convenient to consider the range

of traffic intensity for which the theory is intended. These ranges can be grouped into three qualitative categories: light, moderate and heavy traffic. Light traffic corresponds to the realm of traffic travelling at the speeds desired by the drivers with little, if any, interference between vehicles. Vehicles travelling in platoons or clusters are classified as moderate traffic. In heavy traffic, these platoons become very long and tend to run together.

A useful analogy is drawn by Edie (1974) between the states of traffic intensity and some states of physical matter. Light traffic is analogous to the gaseous state, where each molecule exists independently of neighbouring ones; moderate traffic corresponds to the vapour state, where some molecules have formed droplets and others remain free; and heavy traffic is analogous to the liquid state, where all molecules are in contact with each other. A fourth classification is added to this comparison: the jammed state. This corresponds to the solid state of matter.

We are not concerned with those theories which deal with light or moderate traffic conditions, but concentrate mainly on models which describe the transition between heavy and jammed traffic conditions. This is the realm of operations associated with severe congestion problems.

There are many papers that develop traffic flow theories for heavy traffic intensities. There are two basic approaches: the microscopic view and the macroscopic one. Microscopic traffic models involve a description of the behaviour of single vehicles. Macroscopic traffic models treat the traffic stream as a continuous fluid, and ignore the behaviour of individual vehicles. There is a third class called mezzoscopic models which are often referred to as intermediate models because they fall in between the microscopic and macroscopic categories. The mezzoscopic approach is concerned with the way an aspect of the traffic conditions varies in an environment with specific detail not required.

Here, we present some microscopic and macroscopic theories of traffic flow and discuss

their applicability to the representation of congestion in urban traffic networks.

Car-following theory

An early example of a microscopic model is the car-following one. This model provides a basic formulation for follow-the-leader theory which attempts to describe a single lane of fairly dense traffic in terms of the detailed manner in which vehicles follow one another in a traffic stream, with overtaking not allowed. The underlying assumption is that each vehicle in the line follows the one in front according to a stimulus-response relationship. The car-following approach is not intended for the study of large traffic networks but for understanding traffic behaviour in a single lane or link. A summary of car-following models, their history and application is provided by Edie (1974).

Hydrodynamic theory

The fluid approximation of traffic flow, often referred to as the hydrodynamic approach, was developed in the early 1950s by Lighthill and Whitham (1955) and independently by Richards (1956). This theory replaces individual vehicles with a continuous 'fluid' density distribution which, among other things, satisfies the law of conservation of vehicles. The model can be developed using techniques of partial differential equations to describe the spatial dynamics of traffic. Although the hydrodynamic approach is macroscopic - it ignores some of the detailed behaviour of vehicles in a traffic stream, nevertheless, it develops some interesting results which exhibit a great deal of realism.

The hydrodynamic approach to traffic flow is powerful due to its inherent simplicity, and, it underlies most of the research concerning the time-dependent traffic flow problem of determining how the flow, density and speed vary along a roadway with time.

The approach was adopted by Vaughan et al (1984) who posed this problem for a long arterial street. The analysis relies on the fact that no intersection becomes saturated.

In practice, traffic conditions vary, and saturation may occur. This is the subject of a separate set of papers, Vaughan and Hurdle (1992) and Hurdle (1992), which deal with the dynamic behaviour of traffic along a single arterial street. The model is macroscopic: it considers the effect of the origin-destination pattern on the resulting traffic dynamics. Moreover, the model is an explicit queueing model: intersections along the roadway have capacities and when the capacity of some key intersection is exceeded by the arrival flow, the large queue that results forms the dominant element controlling the subsequent operation of the system. Thus, the model addresses the build-up and subsequent decay of congestion patterns along a single urban arterial. Unfortunately, the solution to the differential equations resulting from the hydrodynamic model of traffic flow, is tedious even for a single link, let alone a series of links along an arterial street. The mathematical analysis is cumbersome and difficult to interpret and cannot easily be extended to cope with large traffic networks consisting of numerous intersecting urban arterial streets.

In a recent paper, Newell (1993) proposed a new method for evaluating the cumulative flow along a single link. The method predicts the variation of traffic flow at one end of the link from the behaviour of traffic at the other end, without evaluating the behaviour at intermediate points. Thus, the model ignores details such as fluctuating vehicle spacing and the speed in a flow but it provides a fully macroscopic representation of traffic flow which can be applied in principle over a wide area. A qualitative description of the model is given by Holmes (1994) who states that in Newell's model

” . . . motorway traffic is treated much like a stream of water being poured through a set of nested funnels. Each funnel collects the water flowing through the previous funnel, plus any added water (traffic joining a motorway), less any water that is siphoned off (traffic leaving a motorway). Each funnel's output is simply its net input or the funnel's maximum flow rate, whichever is less. Any excess water stays in the funnel until it has room to pass through.”

In this respect, the queue propagation model introduced later in this chapter, (Section 2.2.3), is quite similar. However, our model is microscopic and intended for urban as opposed to motorway networks. Consequently, we need to model the queuing behaviour on each link on a lane by lane basis. This is not the case for Newell's model which was developed for the motorway scene and hence did not require this level of detail.

But, Newell's model is unique in that it can be used to represent *systems* of motorway networks as opposed to isolated sections of motorway. Moreover, the mathematical analysis is manageable because travellers are all moving in the same direction and there is little interference from external sources. However, in urban networks, the situation is more complicated because there are a large number of interacting components for example, junctions, parking, pedestrians and pedestrian crossings, local and through traffic, and public transport. This model cannot therefore be used to solve the detailed problems associated with urban traffic congestion.

A similar model for predicting the behaviour of traffic in a single link is provided by Daganzo (1994). This representation, called a cell transmission model, is based on hydrodynamic theory for a highway with a single entrance and exit. The cell transmission model relies on evaluating flow at a finite number of carefully selected intermediate points including the entrance and exit. The model predicts traffic evolution over time and space including the building, propagation and dissipation of queues. Although it requires more computer memory than Newell's method, the procedure has been extended to complex networks such as those involving loops and diverging branches, Daganzo (1995). However, like Newell's method, the model is not designed to cope with the complexities associated with urban networks.

The work of Holden and Risebro (1995) is worth mentioning because they too have modelled networks of unidirectional roads by systems of hyperbolic equations. The work corresponds to a generalisation of the fluid-flow approach due to Lighthill. In contrast to other hydrodynamic models, the study departs from the conventional

transportation planning model of vehicles being bound for specific destinations. The authors propose a model which does not allow for the individual tracking of vehicles. Instead, the model relies upon a set of local turning laws which apply in highly congested conditions. In such circumstances, vehicles would rather make a small detour and drive a longer distance at a higher speed rather than risking a severe bumper to bumper traffic jam. This is the principle of 'least-resistance'. The model is macroscopic and takes a global point of view, assuming that drivers try to maximise flow locally at each intersection. However, although the model exhibits some familiar features of traffic flow, the mathematical investigation is difficult to evaluate and it is unclear how the results can be applied to a real world example.

Queueing theory

Queueing theory is often used to model aspects of traffic. Classical queueing theory involves 'customers', eg vehicles, arriving at 'service points', eg traffic intersections. The service facility requires a certain time to serve each customer and is capable of serving only a finite number of customers during a single time period. If customers arrive faster than the facility can serve them, then customers must wait in a queue. Typically, both the customer arrivals and service times are assumed to have given probability distributions.

A large part of the theory of queues deals with stochastic models but in some of the simple approximations it is convenient to disregard the discrete nature of customers and treat them as a continuous fluid. Thus the hydrodynamic approach is closely linked with the queueing theory analogy of traffic. A comprehensive review on the subject may be found in Newell (1971).

In the past, queueing theory provided detailed analysis of the behaviour of a single queue and of networks that in a sense could be decomposed into a product of single queues. More recently, investigations have focussed on how network components interact. In particular, queueing network models are of current relevance for analysing

congestion and delay in computer systems and communication networks including road traffic networks. These systems are often referred to as multiclass networks because they process more than one class of customer and may also have complex feedback structures. For example, in a computer communication network one may have voice, video and data being transmitted through each node in the network and a job may return to the same node several times for different stages of processing.

These systems can become heavily loaded and their performance cannot easily be analysed as they stand. One method for analysing their performance is to approximate them using particular diffusion processes involving Brownian motions, Williams (1996a). There is now a substantial theory for these diffusion processes which approximates multiclass networks under conditions of heavy traffic.

However, much of the work is concerned with the proof of limit theorems to justify the approximations, Williams (1996b), and although the results may bear some relevance to urban traffic networks, it is not easy to see how these would apply to our study.

Cellular automaton models

The simulation of complete cities using the hydrodynamic approach to traffic modelling requires powerful computing facilities because of the complexity of the problem. In the early 1990s, Bihain and Middleton (1992) considered how a simple and flexible model called a cellular automaton could be applied indirectly to describe the evolution of traffic on urban grid networks. The problem was not so much concerned with the dynamics of the traffic flow through an area, more with the state of the system as a whole. The approach was to reduce the traffic problem to its simplest form whilst maintaining the essential features. The aim was to discover any global properties that would emerge as a result of a set of simple rules.

The automaton approach can be summarised as follows. Cars move inside a town which consists of one-way perpendicular streets arranged in a square lattice. The vehicles may move in an east-west or north-south direction. Cars sit at intersections

and during each time slice, they can move to their nearest neighbours. Two cars cannot be at the same intersection simultaneously. In the deterministic version, Bigham and Middleton (1992), a fixed number of vehicles of each direction is introduced each time-slice. These vehicles progress according to the status of the traffic lights and the states of the neighbouring sites on the lattice. The stochastic model, Molera et al (1993), differs only in that upon entering the network, each car is allocated a preferred direction, controlled by a variable $w_i(r)$ which is the probability that car i located at node r , moves to the allowed neighbour in the east-west direction, and $1 - w_i(r)$ is the probability for a north-south move. Thus, the direction of movement is controlled by a random variable.

In both the stochastic and deterministic versions of the model, time is modelled as a sequence of short increments commencing at time $\tau = \tau_0$. Traffic lights permit east-west motion during increments of even time, ie when $\tau = \tau_2, \tau_4$ and so on, and north-south motion during the increments of odd time, for example when $\tau = \tau_1, \tau_3$.

Simulations of up to 512×512 lattice points reveal a sharp transition that separates a low density dynamical phase where all the vehicles move at maximum velocity without interruption and a high density, static phase where all vehicles are stuck in a static global traffic jam.

The model can be extended so that the lattice resembles a one way grid network. This can be achieved by introducing four possible directions of travel: north, south, east and west. The rules which govern the movement of cars within the lattice are similar to those in the basic model, except that there are four possible directions of travel. Unlike the basic model, the extended version can produce gridlock. Gridlock occurs when a northbound vehicle blocks a westbound vehicle, which in turn blocks a southbound vehicle which blocks an eastbound vehicle. This loop of locked vehicles can happen at any density of vehicles. This idea is shown in Figure 2.1.

The cellular automaton approach can be used to develop a comprehensive under-

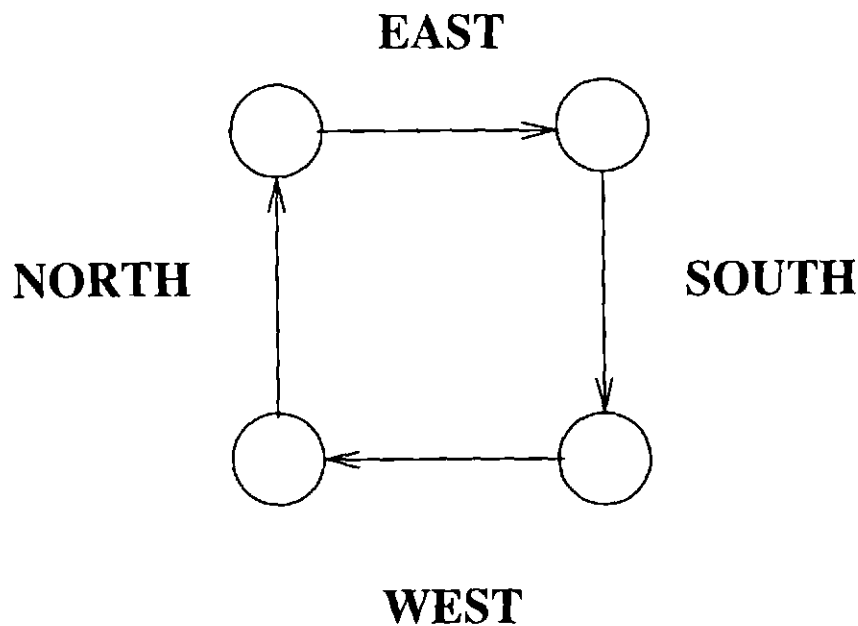


Figure 2.1: *The principle of gridlock in cellular automaton theory*

standing of the mechanism leading to the growth and evolution of traffic jams in grid networks. Although the application of the cellular automaton approach to real traffic problems is still at an early stage, there are a number of general conclusions which can be drawn. In particular, cellular automata seem suitable for studying large traffic networks because of their computational efficiency. The models incorporate some of the essential ingredients of traffic flow whilst maintaining an inherent simplicity. However, the lattice structure employed is typical only of networks in a minority of UK cities. The results are therefore currently limited in their applicability to solve real urban traffic problems.

While these models have not yet been adapted to realistic traffic problems, Wilson (1995) has developed a deterministic model for traffic on an urban square grid which exhibits some features associated with stochastic cellular automaton models. The model departs from the classical origin-destination approach and introduces a local turning law which describes drivers' avoidance of congestion. Thus, on arrival at a green light, each vehicle makes a turning decision based solely on the lengths of the

queues which it may join, and not on any inherent desire to travel in a particular direction. However, unlike the cellular automaton model where traffic is treated discretely, this model considers traffic as a continuum. The model is capable of describing the dynamic behaviour of queues on a network of roads. But, like the cellular automaton models, the study is concerned with the global properties of the system, and the way congestion waves may build and decay, rather than with the spatial characteristics of the queues and their interaction with the underlying road network structure.

2.2.3 Congestion growth models

Traditionally, investigators have analysed large interactive systems in the same way as they have small, orderly systems, mainly because the methods developed for simple systems have proved so successful. Thus, it was thought that the behaviour of a large system could be understood by studying its individual elements in detail. During the past few decades, it has become increasingly apparent that many complex systems do not yield to this approach.

This has led to a new generation of models which emphasise how uncoordinated local decision-making rules lead to system-emergent properties. But because complex systems contain many components and are governed by numerous interactions, analysts cannot possibly construct mathematical models that are both realistic and theoretically manageable. They therefore have to resort to simple, idealistic models that capture the essential features of the real system. If the simple models are robust, they might be able to extrapolate the findings to real situations.

The application of cellular automata to model the motion of traffic is an example of how local actions (turning laws) may give rise to self-organising structures (traffic jams) with emergent properties. Traffic jams grow in a way that might be expected to resemble the growth of two-dimensional aggregates of particles. Starting from an initial obstruction, traffic queues propagate from link to link via a process of

branching, each queue generating one or more branches at each junction in turn. This process generates a cluster whose structure resembles a fractal.

The model of diffusion-limited-aggregation, Witten and Sander (1981), often referred to as DLA, has been invoked to explain the apparently fractal nature of many physical phenomena, for example urban morphology, Batty and Longley (1994), and the evolution of drainage networks, Masek and Turcott (1993). Here we show how the principles of DLA can be applied to describe the mechanism of traffic jam growth in an urban environment.

DLA - A model for traffic jam growth

DLA is a process which generates a single cluster of particles over a network of sites. The model's name suggests a pattern in which particles cling to a self-evolving structure. The *aggregation* of particles can be described as a *diffusive* process since the structure spreads out over the lattice. The diffusion mechanism is governed by local laws which *limit* the development of the cluster over the network.

The mechanism of DLA can be summarised as follows. A seed particle is fixed somewhere near the centre of the network. During each time-slice particles are launched from the perimeter of the bounded region. When a launch occurs the particles begin random walks on the network. During each step of a particle's random walk, the particle can assume one of three states: fixed, transient or destroyed. Fixed particles cease to move because they have encountered the developing structure. Transient particles continue to move until either leaving the system in which case they are destroyed, or until they become fixed. Similar to cellular automata, at each time-slice, a particle can move one step in one of a number of directions. The number of directions depends on the underlying lattice structure.

The structure of DLA clusters is dendritic, with tentacles extending from the seed particle in a tree-like fashion. This is because when a particle sticks to another, the probability of more particles sticking in the neighbourhood is much increased.

Ribbons of particles begin to form around the centre of the cluster, making it ever more likely that new particles will stick to the tips of existing dendrites. These effectively screen the fissures between the emerging tentacles from receiving further particles, Batty et al (1989). Consequently, the rate of growth is faster at the tips of the cluster than in the centre.

The DLA mechanism can be modified to reflect the process of traffic jam growth, as described later in Chapter 4. However, while the model reveals insights regarding the overall structure of traffic jams, it cannot be used to understand the mechanism of queue propagation in detail. This is the purpose of a second, more practical model, which is described next.

A queue propagation model

Conventional traffic assignment models focus on the movement of vehicles (or streams of vehicles) through the road network. But a traffic jam is essentially a process of queue propagation in the opposite direction to the flow, and in some respects the mathematical tools required to analyse this process are simpler - and require less data - than those required for conventional traffic modelling.

The model proposed by Wright and Roberg (1997) is a theoretical one involving a number of simplifying assumptions. The authors have used the model to investigate the way in which traffic jams form and evolve on a macroscopic scale when a bottleneck is inserted at some point in the system. The network is an idealised, uniform, one-way rectangular grid system. The model assumes steady-state demand conditions: localised dynamic effects such as stochastic variations in demand and cyclical flow variations associated with traffic signal phases are deliberately ignored. The study predicts the shape of the jam boundary by deriving theoretical expressions relating to the jam's growth over time. These are used in turn to study the effects of varying traffic management and road layout parameters, and qualitative conclusions are drawn about physical countermeasures such as the allocation of queuing space between

ahead queues and turning queues.

The model sacrifices a great deal of realism, and the results are not intended to be immediately applicable to real networks. However, they yield qualitative insights which may lead to a better understanding of the congestion problem in general terms. In particular, they predict certain features of jam behaviour that would not be accessible in any other way, and that can in principle be tested. Later, we shall describe the underlying principles of the model in more detail and show how the model provides a framework for the study of networks of queues in a simplified urban environment, with particular reference to the interaction between streams of traffic in the queues and their dynamic behaviour.

2.2.4 The need for simulation

Traffic congestion studies such as the ones proposed so far have concentrated mainly on techniques of mathematical analysis. In spite of the simplifying assumptions made by these models, they often lead to general qualitative results concerning network behaviour. But the analytic method is limited because the results often involve complex mathematical expressions which are difficult to evaluate and the equations do not always hold if the boundary conditions are modified. This imposes a restriction on the generality of the results and may question their validity.

This often leads the problem-solver to the realm of computer simulation which can be attractive where the mathematical modelling of a system is impossible or where the analytic method has no simple solution. Simulation is particularly suited to modelling time-varying, complex, and stochastic processes such as traffic flow and can often be used to validate the results of a theoretical study. In addition, simulation provides a high level of detail and accuracy and eliminates the need for costly and sometimes infeasible experiments. Simulation also allows for the testing of alternative systems under identical conditions and enables a 'systems approach' when measuring the effects of interacting variables.

But simulation is a laborious, iterative and experimental problem-solving technique, which may prove inefficient and expensive in terms of manpower and computer resources. Simulation models may yield suboptimal solutions and do not always allow one to make general predictions in the way that analytic models do. Thus the virtue of simulation is its very disadvantage - one may be able to test specific complex regimes but in the case of a large parameter space it becomes impractical to determine the effects of varying all parameters individually. Consequently, simulation and analysis are best regarded as complementary disciplines.

The simulation technique is often supplemented with observational studies. However, it is quite difficult to study congestion problems via direct observation. Van Vuren and Leonard (1994) state a number of reasons for this. First, data-collection exercises are expensive particularly when considered over large areas. Second, an approach based on data-collection alone does not enable the researcher to test the sensitivity of the resulting congestion with respect to incident duration, severity and driver behaviour assumptions. Third, the generalisation of the results of one particular congestion study to similar or different ones might be difficult. Finally it is not easy to investigate the effects of remedial strategies in real-life observation studies. This is particularly true of studies of traffic jams because it is quite difficult to *find* a traffic jam at the moment of birth and to observe its evolution in its entirety.

Consequently, simulation is the only practicable solution for studying traffic jams. But although there have been many previous attempts in developing simulation models for traffic congestion studies, few can be applied to study the behaviour of traffic jams. This is the topic of the next section.

2.3 Simulation models for congestion studies

2.3.1 Introduction

In developing a simulation model, there are two ways to represent the changing states of the variables. The 'fixed-time increment' method reviews the state of each variable at fixed increments of time whereas with the 'event-based' technique, the states of all the variables are updated as a result of the scheduling of an event. Thus, only events are represented explicitly in the model, and the time between events is treated as inactive.

Fixed time increment methods are suitable for the simulation of continuous systems (which vary smoothly over time), and systems where there are large numbers of state variables. In the case of the latter, it is more efficient in terms of computer resources to update the states of all the variables at intervals of fixed time rather than each time an event is triggered. On the other hand, discrete simulation models often favour the event-based approach and are sometimes known as discrete event simulation models.

Simulation programs can be written in a variety of computer languages. There are also a number of languages which have been developed specifically for simulation purposes, for example GPSS, SIMULA and SIMSCRIPT. A discussion of the main characteristics of these languages is provided in Neelamkavil (1987).

2.3.2 A survey of existing traffic simulation models

A number of traffic simulation models have been developed over the last few decades. These usually describe the behaviour of traffic by tracing various system states as a function of time and then collecting and analysing summary statistics. Some models have been applied specifically to study the problem of traffic congestion and include facilities for describing incidents and for representing the dynamic behaviour of queues in congested conditions.

When studying a new problem, it is worthwhile considering whether existing simulation models can be used. We are particularly interested in those models which are suited to the study of urban congestion. There are two important components to examine in such models (a) the method of assignment and (b) the queueing representation.

Traffic assignment is the process of finding routes through a road network and loading demand trips onto them. Traffic assignment models are generally used to forecast flow patterns through a traffic network given a particular demand for travel and supply of road space in the street network. They are usually static in nature, ie, the demand for travel, expressed in the origin-destination matrix, is fixed for the modelled period, and so are the network supply parameters, such as the signal settings. This is not the case with dynamic traffic assignment models where the system adjusts itself continually according to the flow patterns and conditions on the network. This is important when modelling incident-induced traffic congestion which is often unpredictable. Traffic systems need to be able to respond to such unexpected events and provide motorists with accurate information.

The behaviour of queues in congested conditions can be described using either a vertical or horizontal queueing model. Vertical queueing fails to take the spatial extent and blocking-back characteristics of a queue into account. Instead, the vertical representation relies on the relationship between the flow exiting a junction and its capacity. Thus, for an individual junction, the physical length of the queue and its relationship to space, is unimportant. All that is required is to minimise the queue length. However, for a network of queues, this assumption creates problems since the minimisation formulae assume that vehicles arrive at the stopline, but if queues have formed, vehicles will be at the end of the queues and not at the stopline. This means that the queues will be underestimated. In addition, they do not show the precise location of vehicles in the network and there are instances where the length of a theoretical queue may exceed the storage capacity of a link, Paksarsawan (1994).

By contrast, the horizontal queueing mechanism gives a precise indication of the vehicles maintained in the queues. Thus, vehicles pass the stopline if the saturation capacity of the junction exceeds the flow and there is space ahead in the downstream link. If either of these conditions are not satisfied, then some of the vehicles may be retained on the upstream link. Thus, the number of vehicles leaving a junction is related to the physical constraints of the neighbouring links.

In this section, we consider the suitability of several commercially available traffic simulation tools with particular emphasis on the assignment and queueing representations which have been employed. These generally do not model gridlock.

CONTRAM

CONTRAM (Leonard (1989)) is a traffic assignment model developed by the Transport Research Laboratory to assist in the design of urban traffic management schemes. For a network having specified traffic flows between specified origins and destinations that vary over time in a known fashion, it predicts the routes that vehicles will use and calculates the traffic flows, queues, journey times and journey costs through the network.

The model uses an 'incremental' form of vehicle loading which assigns packets of vehicles to their minimum journey time routes in the network through a number of iterations. However, if vehicles are assigned to the quickest route, then the implication is that drivers have prior knowledge of the network. This is consistent with the idea that drivers who make the same journey each day will find the quickest route by repeated experience. On any one day, conditions may be different from the average, in ways which are largely unpredictable. Incidents such as accidents, breakdowns or roadworks cause a major form of planned unpredictability, but it is unrealistic to assume that drivers do not make some attempt to avoid the resulting congestion. A version of CONTRAM called CONTRAM-I has been developed by Southampton University (1992) to explore the effects of travellers' self-diversion, ie where drivers

change their original route in response to seeing unusually large queues ahead. CONTRAM has been enhanced further to model the effects of route guidance. A summary of the CONTRAM suite of models is provided in White et al (1994).

Although CONTRAM is a dynamic model and can thus describe the build-up and decay of congestion in a network over a given time period, the queueing representation is vertical. CONTRAM also imposes a constraint on network size and is limited by hardware specifications. (We have only managed to construct a 10×10 one-way grid network.) In addition, the graphical component is not commercially available in the standard version. This restricts the spatial analysis of traffic congestion patterns which change in response to the conditions in the network.

An example of a recent congestion analysis study which used the CONTRAM suite of programmes is described by Van Vuren and Leonard (1994). The study considers the spread of urban congestion in two networks: Kingston and Reading, at different times of the day. The study describes the impact of incidents on urban congestion as a function of incident and network characteristics. Some remedial control strategies have also been developed and tested using model simulations. The authors conclude that the severity of congestion is strongly determined by general network characteristics, such as average link lengths and junction control. The location of the incident is identified as an important factor in the spread of congestion. Finally the work reinforces the need for modelling in the assessment of the spread of congestion caused by incidents and highlights that an accurate queueing model should be employed to represent the build-up and decay of queues. The authors recommend that any graphical facility provided by the model should allow for a comprehensive analysis of traffic patterns through the network.

SATURN

SATURN is a simulation model for the analysis of traffic management schemes, Van Vliet (1982). It is a static traffic assignment model which simulates delays at intersec-

tions. The underlying traffic pattern is represented by a fixed O-D matrix. The model incorporates two phases: the calculation of delays and the assignment of routes.

The main limitation of SATURN as applied in congested situations is the queueing model. SATURN assumes queues at junctions to be vertical queues instead of horizontal queues which appear in real life. This is a simplification which is acceptable in less congested conditions, but when modelling oversaturated conditions, an accurate representation of queueing behaviour and blocking effects is essential.

The static nature of SATURN makes this tool unsuitable for modelling the effects of traffic congestion. Moreover, the graphical facilities incorporated into SATURN cannot easily be adapted to describe the build-up and decay of congestion over time.

Nevertheless, SATURN has been applied, albeit within certain limits, in many cities, mainly in developed countries with similar conditions to those in England. More recently, SATURN has been used to model traffic conditions on the Bangkok network, May et al (1993). In particular, the traffic behaviour and the formation of queues at junctions were analysed. Although in general the model worked well, it underestimated the capacity of junctions. This led to a further study which was conducted by Paksarsawan (1994), who describes the development of queueing simulation procedures in more detail using a different micro-simulation model. The TRAF-NETSIM package was preferred to the SATURN model, because the former could be adapted to model traffic queues horizontally.

TRAF-NETSIM

TRAF-NETSIM, Rathi and Santiago (1990) was developed for the Federal Highway Administration (FHWA), USA, more than twenty years ago and has been continually enhanced. The model adopts a fixed time discrete event simulation approach which describes the dynamics of traffic operations in urban street networks. TRAF-NETSIM incorporates a car following and lane changing model. Vehicles enter the network via sources and progress subject to traffic control, pedestrian activity, transit

operations and various other events which influence traffic behaviour. The effects of blockages and the blocking back of queues can be simulated in a detailed manner. The model employs a horizontal queueing representation. This makes the model capable of modelling the effects of spillback in urban networks.

TRAF-NETSIM has been successfully applied to model spillback conditions particular to the Bangkok network. The model was used to investigate appropriate control strategies for oversaturated conditions, see Paksarsawan (1994) for further details.

However, the program is cumbersome and difficult to adapt as it requires a very detailed description of the road network and traffic demand pattern, and hence carries a substantial 'overhead' in terms of setting up costs and run times. This and the lack of visual facilities make the package unsuited to the spatial analysis of congestion patterns in large, isotropic networks.

TRAFFICQ

TRAFFICQ is a dynamic stochastic simulation model of traffic and pedestrian activity in road networks, Logie (1979). It was designed to investigate in detail the effect of introducing a new traffic scheme in a complex urban situation where extensive queueing and congestion of one junction on another is prevalent. TRAFFICQ is not an assignment program: the O-D matrix and the vehicle routes are fixed by the user. It was originally intended to be used for examining part of a network in detail and used in conjunction with a less detailed model; it employs a horizontal queueing model and provides graphical output of the congestion patterns. However, TRAFFICQ is a highly detailed model and as a result, the maximum size of the network is severely limited.

While TRAFFICQ incorporates some fundamental assumptions regarding the behaviour of queues at intersections, nevertheless, the size constraint coupled with the level of detail provided by the model, makes TRAFFICQ more suitable for studies conducted over a section of road rather than for area-wide surveys of congestion

problems.

NEMIS

Another traffic model which has been applied to the study of congestion is the NEMIS model, Mauro (1991), developed by Mizar Automazione SpA. NEMIS simulates individual vehicles according to a car-following law, lane changing rules, junction regulations and traffic light status. It is a traffic assignment model which includes an additional micro-simulation module. This feature distinguishes the model from most conventional traffic assignment models because it takes into account the availability of storage space downstream. Consequently, NEMIS can represent the blocking-back of upstream junctions and hence the disruption caused to the traffic.

The program uses a shortest path algorithm to assign vehicles to their quickest routes in the initial stage of the simulation, but thereafter these routes remain fixed. The NEMIS model presents results in animated visual form which makes it particularly applicable to the study of congestion patterns in urban networks.

The model has been applied to represent the central Turin network and has been used as a tool to develop various auto-gating strategies, Shepherd (1994). In addition, the author describes experiments which have been conducted on 4×4 one-way and two-way grid networks, and discuss the implications of gridlock phenomenon to the performance of the gating strategies. However, NEMIS was not compatible with our requirements because of the size limitation imposed by the model and also was not available at the time of the study.

2.3.3 Developing a congestion modelling environment

So far we have discussed different aspects of traffic simulation models and their capability for modelling congestion. However, it was decided not to use any of these existing simulation models because no single model could be applied directly to study traffic jams in urban networks. Nevertheless, the models exhibited certain features

which are important in the traffic jam modelling context. Thus, in developing a congestion modelling environment for the study of traffic jams one would require:

- a microscopic approach which provides an appropriate level of detail for representing the movement of vehicles through the network,
- a horizontal, lane-by-lane queueing model which accurately describes the interactions operating both within and between the queues in the network
- a dynamic representation of the traffic system, ie the model should be capable of responding to variable traffic conditions
- a graphical facility which displays the evolution and decay of congestion patterns and how they vary in response to network conditions over a considerable time period
- a 'control' component which would enable the testing, evaluation and implementation of general control strategies
- a flexible, modular program structure which is easy to modify and enhance for a specific network environment
- a large study area with few restrictions imposed on network size or speed of operation

Table 2.1 shows the relationship between these criteria and the traffic simulation models mentioned in the previous section. The summary suggests the need for a single framework which would combine these features to provide an environment for describing urban traffic jam growth and decay.

However, in the past, the majority of congestion analysis studies have been context-specific and do not necessarily generalise. Our aim is to build a research tool for the study of traffic jams which would apply independently of the network structure and

	CONTRAM	SATURN	T-NETSIM	TRAFFICQ	NEMIS
Approach	mezzo	macro	micro	micro	micro
Queueing	vertical	vertical	horizontal	horizontal	horizontal
Assignment	dynamic	static	none	none	static
Graphics	non-standard	not online	none	not online	online
Structure	difficult	adaptable	difficult	detailed	adaptable
Size limit	yes	yes	yes	yes	yes

Table 2.1: *Features and limitations of existing traffic simulation models when applied to study traffic jams in urban networks*

localised factors. This would enable a detailed analysis of the spatial dynamics of traffic queues and the interactions operating between them.

Although the research tool does have its disadvantages, for example it can only be used in an idealised setting and some of the underlying assumptions concerning vehicular movement are somewhat unsophisticated, nevertheless, the model does capture the essential features of traffic jam growth and dispersal, and providing the model is sufficiently robust, it may be possible to extrapolate the findings to less idealised situations. Ultimately, this would lead to a better understanding of the concept of blocking-back, and its role in the formation and dissipation of traffic jams in urban networks.

2.4 Conclusions

This chapter has considered two approaches to urban traffic congestion modelling: theory and simulation. A selection of examples underlining these approaches has led to a discussion concerning the application of analytic and simulation modelling techniques to developing a representation for the spatial analysis of traffic jams. The study found that in this context, simulation and analysis were best regarded as complementary disciplines, but that simulation was particularly suited to the modelling

of time-varying, complex and stochastic processes such as traffic flow.

As a result, a selection of the leading simulation models were considered in order to determine the extent to which they could be utilised to support this study. The models which were evaluated were CONTRAM, SATURN, TRAF-NETSIM, TRAFFICQ and NEMIS. It was thought that these could be applied to the complexities associated with urban networks, because of the level of detail they provided. Thus, for example, the models can be used to predict route choice and journey times on a road network, for any demand pattern and network configuration. The results of the evaluation are summarised in Table 2.1.

However, the survey found that no single model was entirely compatible with the requirements for modelling the development and dispersal of area-wide traffic jams. Both CONTRAM and SATURN employed a vertical queueing representation, and consequently were not capable of modelling gridlock conditions. Their graphical capabilities were also deficient and the dynamic effects of control could not easily be monitored. On the other hand, TRAFFICQ, TRAF-NETSIM and NEMIS used horizontal queueing conventions, and even provided graphical facilities. The graphical representation was crucial for the spatial analysis of dispersing traffic jams. However, the models required detailed descriptions of the road network and traffic demand pattern. This slowed the speed of operation and imposed restrictions on the size of network allowed.

Based on these considerations, we have developed a new, dedicated research tool to explore the mechanisms of traffic jam growth and decay, over a wide portion of the network in less than real time. Whilst the model does have some limitations, nevertheless it can be used to investigate the fundamental characteristics of queue development and decay in idealised road environments. The model incorporates those features which are essential for the study of traffic jam formation and dissipation. These include an accurate representation of queues, an online, visual description of dynamic congestion patterns, and facilities for developing and testing the effects of

remedial control measures. In addition, the model is capable of displaying congestion patterns at high speed over a considerable area.

In the next chapter, we describe the underlying assumptions of the tool and outline its main features and drawbacks. We then show how it can be used to investigate the development and dispersal of area-wide traffic jams in rectangular grid networks.

Chapter 3

An exploratory tool for traffic jam modelling

3.1 Summary

In the previous chapter we showed how traffic congestion could be modelled using a combination of analytic and simulation techniques. We outlined the features of a selection of theoretical and simulation traffic models, and considered their applicability to traffic jam modelling. Although appropriate in a number of respects, no single model could be identified as entirely suitable for this task. This chapter describes a model developed at Middlesex University, specifically to explore the characteristics of traffic jams in rectangular grid networks.

3.2 The traffic simulation model

3.2.1 Introduction

The characteristics of vehicle movement in a traffic jam differ from their behaviour under normal conditions. Under normal conditions, the length of a traffic queue may fluctuate, according to the time pattern of discharge from the head of the queue and

the arrival pattern at the tail of the queue. Consequently, the variations in arrival and departure rates are important, and one cannot predict the behaviour of a queue without a detailed knowledge of the dynamics of the upstream and downstream flow.

But in severely congested conditions, a queue grows without limit and the effects of stochastic variations in arrival and departure rates become less significant as the traffic jam grows in size. Although it will still be affected by momentary changes in the arrival and departure rates, in the longer term, the rate of growth is largely determined by the average arrival and departure rates. As a result, the way in which a queue propagates is not very sensitive to the precise behaviour of individual vehicles as they move from one queue to another. A simple model is sufficient to demonstrate the key features of jam formation without the need for a detailed representation of (a) platoon dispersal or (b) junction layout or (c) stochastic fluctuations.

In this section we propose a model which simulates the way in which traffic jams evolve on a macroscopic scale when a bottleneck is inserted at some point in the system. The networks are idealised, uniform, rectangular grid systems. The model builds on earlier work by Abbess, Wright and Huddart (1991) who develop a preliminary model for traffic jam growth, and show how it could be extended to investigate the fundamental features of traffic jams on rectangular grid networks.

3.2.2 Previous work

In a recent study, Abbess, Wright and Huddart (1991) developed an analytical model for traffic jam growth arising from a single bottleneck in an idealised network. In the model, the pattern of queue propagation depended on the numbers of vehicles wishing to turn and move straight ahead at each junction, and on the storage areas and stopline capacities associated with these movements, but not on the routes adopted by individual vehicles through the network. This enabled traffic jams to be modelled independently of the origin-destination matrix.

The theoretical analysis predicted that, under idealised conditions, there were essentially two possible spatial configurations for a traffic jam on the type of network considered. In addition, the rate of jam growth was affected among other things by the severity of the original blockage, the overall level of traffic demand, the proportion of link storage area devoted to the segregated queues, and the balance of storage area between ahead and turning queues.

The authors supplemented the theoretical analysis with a preliminary computer simulation model. Unlike the analytic version, the simulation model reproduced the interaction which occurs when queues tail back (often at different speeds) from more than one source.

The computer model was also used to examine the influence of network parameters on the rate of traffic jam growth, and identified the extent of queue segregation and block length as important variables. It was postulated that for any given set of parameter values, the rate of delay increased approximately as the square of elapsed time. Subsequently, a number of practical measures for inhibiting area-wide congestion were reviewed, Wright and Huddart (1989a,b).

The models could be used to predict the evolution of traffic jams on one-way grid networks. However, the models were restricted and could only be applied under steady flow conditions. In addition, the models could not represent the process of dispersal nor could they be adapted to two-way grid networks. This suggests the need for a more comprehensive tool which could simulate the formation, growth and dissipation on one-way and two-way grid networks. The tool should also provide facilities for developing and testing remedial control strategies, so that the effects of different queue control measures might be examined.

Aims of the research tool

Since the formulation of the original model, this author has undertaken its development and introduced several new features. The objectives of the model are now

1. to develop previous modelling work in more detail incorporating stochastic variations in the demand levels,
2. to simulate the formation, growth and dissipation of traffic jams on one-way and two-way grid networks,
3. to identify the main parameters which govern the formation, rate of growth and dispersal characteristics of traffic jams, on the above networks.
4. to test the sensitivity of the model results to the assumptions on which it is based,
5. to develop and test different queue control strategies for protecting networks from gridlock and dissipating traffic jams once they have occurred.

The remainder of this chapter describes in more detail the main features of the simulation tool and how it can be applied to study the characteristics of traffic jams in one-way and two-way rectangular grid networks. The section begins with the modelling assumptions for the different types of networks, and then sets out the parameters and notation adopted within the model. A summary of the output statistics available in the models, and how these can be used to quantify system characteristics and the severity of congestion, brings this chapter to its conclusion.

3.2.3 The modelling assumptions

The model networks are restricted to uniform one-way and two-way rectangular grid systems. The same general assumptions apply to both network types, but there are a number of differences in detail. We outline the assumptions for the one-way grid system first, and then deal with the differences in a separate section. The underlying algorithms for the assumptions which follow are provided in Appendix E.

Grid network

The one-way system consists of two sets of parallel one-way streets intersecting at right angles, with the direction of traffic alternating between neighbouring roads, and the roads at equal spacing. There are no intermediate origins or destinations within any of the links, and only the severed links on the boundary act as traffic sources and sinks. Vehicles travel from any source towards any sink. The routes which vehicles follow are not pre-determined. The mechanism by which the vehicles progress is described in the section on *vehicular movement*. The algorithm describing the way vehicles decide where to go is provided in Appendix E.

The queueing behaviour at intersections in the network preserves the 'yellow-box' convention whereby intersections remain clear and are not affected by cross-blocking. Vehicles approaching intersections 'look ahead' to avoid stopping within the intersection itself, and spillback does not block crossing traffic at the upstream junction. Instead, vehicles wait on their respective approaches until there is space to proceed. (In practice, drivers are not always so well disciplined, but we have not pursued this aspect here.)

Link storage configuration

The queueing arrangements are relatively simple. Each link contains just two segregated queue storage areas terminating at the downstream stopline: one for the 'ahead' movement, and one for the turning movement. Each segregated storage area is of constant width; it may consist of one or more lanes, but there are no shared lanes. Upstream of these segregated storage areas, there is an upstream 'reservoir' that feeds into them both. It occupies the full width of the road. One can visualise the reservoir area as being channelised (ie, it may possess lane markings), but within this area, vehicles are allocated positions regardless of their intended direction of movement at the downstream junction. The reservoir *may* be of zero length, allowing simulation with 'perfect' lane discipline everywhere.

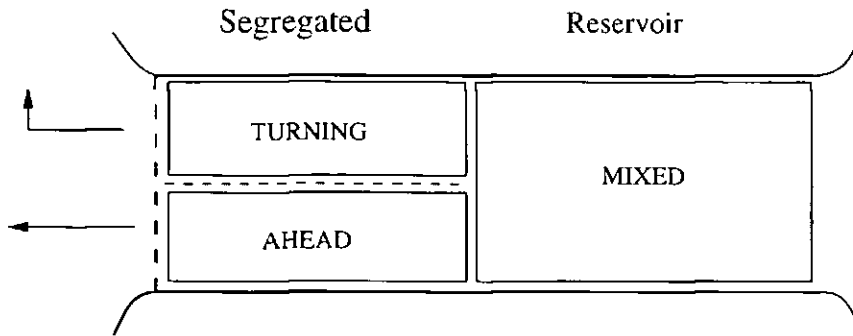


Figure 3.1: *Queue storage configuration for a one-way link*

The arrangement of the segregated and reservoir portions on a one-way link is shown in Figure 3.1. The mechanism of queue spillback within these storage areas will be described later.

Demand profile

In the preliminary model of Abbess, Wright and Huddart (1991), traffic flow was deterministic, ie, stochastic variations were ignored and demand was spatially uniform everywhere within the system. But in real networks, this is not the case. A new feature of the simulation tool is the Poisson model which has been incorporated to allow for stochastic fluctuations in the level of demand. The Poisson model is used to generate vehicles from each of the sources (See Appendix E for details). The mean of the demand at each access point is described by a constant, μ , in units of vehicle arrivals per cycle. It has the same value for all the sources to create a steady flow of vehicles over the long term throughout the system.

The demand pattern remains the same throughout the modelling period. In principle, the demand profile could be modified to cope with transient peaks in the level of demand, but we wanted to eliminate the factors which might complicate the mechanism of traffic jam growth and decay. Initially therefore, a constant demand profile was selected.

Vehicular movement

The mechanism of vehicle movement can be summarised as follows. Vehicles are partially segregated into traffic lanes and progress subject to space being available downstream, capacity constraints at junctions, and interactions between queueing vehicles.

Vehicles travelling along any given link may choose from two possible movements at the downstream junction - 'ahead' and 'turning'. The turning movements alternate between left and right turns at successive junctions along any given road.

Conversely, at any given junction, vehicles may arrive via either of two approaches. These have equal priority; by this we mean that the opportunities for discharge are shared between the approaches in proportion to the demands.

The model addresses the movements of individual vehicles. However, the routes that they follow are not predetermined, and it does not 'track' them through the system. Suppose that X vehicles are discharged from any particular reservoir during a given time slice. They are split into two groups: the number of vehicles allocated to the 'turning' group is the nearest whole number to pX where $p \in \{0 \dots 1\}$ while the remainder are allocated to the ahead group. The groups are progressed respectively to the downstream ahead and turning queues, regardless of origin. Implicitly therefore, the O-D matrix is predetermined with smooth flows between each origin and destination pair, although it is not possible to refer to the matrix directly. The pattern of queue development is thus determined by the overall turning movements as opposed to the specific routes followed by individual vehicles. More details on vehicular movement can be found in Appendix D and E, at the end of this thesis.

This represents a simplification in the conventional modelling of road traffic flow. A similar stance is taken in Wilson (1995) and also Holden and Risebro (1995), whose work was described in the previous chapter.

Spillback

We picture the mechanism of spillback on a one-way link as follows. At the upstream end of the link there will be a junction with one or more approaches that feed traffic into the link. At the downstream end there will be a junction with one or more exits through which the traffic may leave the link. If a particular exit becomes blocked, (say the left turning exit), vehicles intending to turn left will form a queue that spills back along the link. If the lane discipline of drivers at the various points of entry to the link is not perfect, the queue may eventually spread across the other lanes and block all the traffic (Figure 3.2).

The extent to which the queue spreads and blocks the movement of all traffic will depend on the network conditions, and also on the behaviour of drivers in the network. Skillful drivers often pass through gaps that arise in queues of traffic, and the assumption of imperfect lane discipline may not always hold. This has led to a refined spillback definition which is described as follows.

In the model, the road link is divided into two distinct zones: a downstream queue storage area where vehicles are organised into separate turning movements, and an upstream 'reservoir' area where the turning movements are mixed. The interactions occur at the transition between these zones. The degree of interaction may be altered via an additional parameter, denoted by s , which assumes values in the range $0 \dots 1$.

With $s = 0$, it is assumed that when a segregated queue spills back to the reservoir area, then no further movement is allowed from the reservoir area into either of the segregated regions.

By contrast, when $s = 1$, a queue in one segregated area does not affect the discharge of vehicles from the reservoir area into the alternative segregated area. This implies perfect lane discipline along the whole link: on entry vehicles immediately select the correct exit lane. Values of s between 0 and 1 can be selected to model the effects of partial interaction.

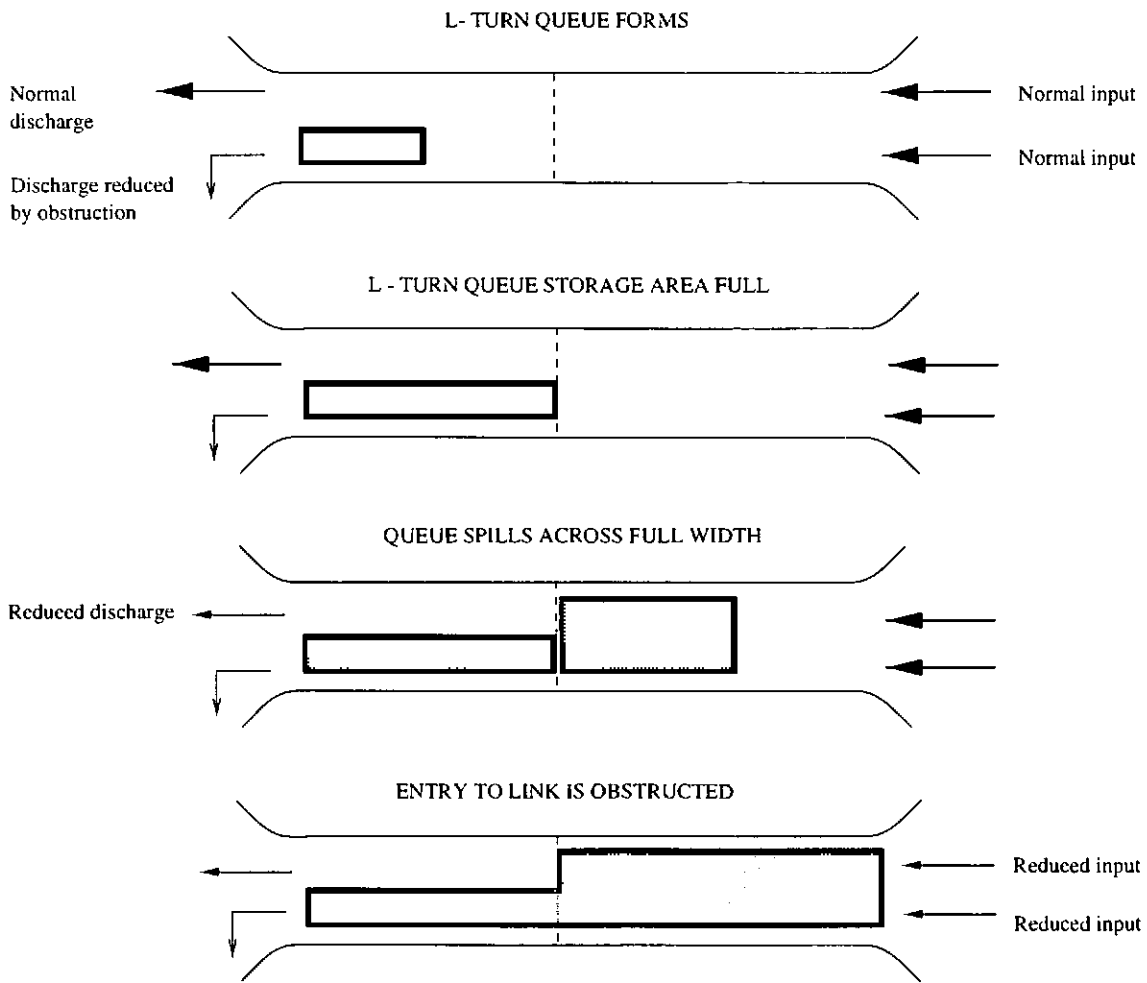
This modified spillback definition is a new feature of the simulation tool and it provides improved flexibility particularly when investigating the sensitivity of the model to its underlying assumptions. This will be described later in Chapter 4.

Queue propagation

Starting from an initial obstruction located somewhere near the centre of the study area, we visualise a traffic jam as propagating from link to link via a process of branching, each queue generating new branches at each junction in turn. Initially, the topology of the branches resembles that of a simple tree, but at some stage, a queue will tail back around four sides of a block to form a closed circle. In practical terms, this represents a significant stage in the evolution of a jam because once it has occurred, it is difficult to clear the jam even if the demand falls below the theoretical capacity of the obstructed network.

Intuitively, one might expect the rate of queue propagation to diminish with increasing distance from the original obstruction, because vehicles whose routes are destined to pass through the bottleneck will form a diminishing proportion of the total traffic. The jam will 'die out' at points remote from the source. However, within the idealised framework of the model, this is not so. Under the steady-state demand conditions assumed here, jam propagation is a 'domino effect'. Once the system has failed at any point, the constriction is transmitted indefinitely from link to link until the whole system has broken down. Essentially, queue propagation is controlled by interference between turning movements, and is independent of the routing pattern as such.

This property of queue propagation is fundamental to the model, and it has two significant implications. First, it allows the behaviour of the system to be analysed without reference to the pattern of origin-destination movements - a factor that simplifies modelling and reduces computer simulation times. Second, for geometrically regular networks having static traffic demands, the pattern of queue propagation will also be in some sense geometrically regular, repeating itself indefinitely as the jam



Source: Wright C C and Roberg P, 1995

Figure 3.2: Stages in the propagation of a queue along a one-way road link where the left turning discharge is obstructed

grows in size.

In the theoretical model, each queue is treated separately and the effects of queues arriving at the same junction from different sources are deliberately ignored. This is not the case with the simulation model which can handle the effects of multiple queues converging on the same junction, often at different speeds.

This may affect the overall pattern of queue development since the growing queues may themselves form blockages of upstream junctions turning previously congested links into ones where there are unusual free-flow conditions. This is known as 'starvation'. Starvation begins to occur when an obstruction is placed on the network, causing a reduction in the flow downstream. This effect is propagated from one link to the next and may ultimately lead to empty spaces that develop simultaneously with the traffic jam itself. For example, where queues develop around an area encompassing a number of city-blocks, certain links may become totally 'starved' of flow due to the blockages caused by the developing queues. Starvation also affects the boundary of the traffic jam and causes a peripheral 'halo' which encompasses the jam. This will be discussed in more detail in the next chapter.

3.2.4 The two-way grid network

The author has also adapted the simulation tool for two-way grid networks. The same general assumptions apply, although there are some differences in detail. These are outlined next.

The environment

Like the one-way version, the two-way grid network consists of a series of parallel streets intersecting at right angles. However, there are a number of differences. In particular, we have introduced an additional type of road link called a right-turn bin. These can only accommodate a small number of vehicles and act as storage space for opposed right-turning vehicles.

Opposed right-turning traffic may wait in the dedicated right-turn bins until gaps arise in the opposing stream of traffic. When this occurs, (typically at the end of the cycle), opposed right-turning vehicles proceed to the selected destination link. During the cycle, a right-turn bin may become full of waiting vehicles. If this happens, no more right-turning vehicles requiring this opposed movement, are released into the storage area. This may cause queueing on the link feeding into this area. Thus, the right-turn bin acts as an additional constraint on the throughput of right turns.

There are no intermediate origins or destinations within any of the links, and only the severed links on the boundary act as traffic sources and sinks. Vehicles travel from source to sink. The assumption of no cross-blocking still holds. There are no changes in the way vehicle arrivals are generated.

Vehicle movement

The direction of travel adopted in the two-way model conforms to the left-hand conventions accepted in the UK, where right-turning vehicles are forced to cross a traffic stream moving in the opposing direction.

In the one-way grid, vehicles travelling along a link may select from two alternatives: ahead and turning (either left or right, but not both). But in the two-way network, vehicles travelling along any given link may choose from three possible movements at the downstream junction - ahead, left or right.

Conversely, for any given link, vehicles may arrive via any of one of three approaches at the upstream junction. These have equal priority; by this we mean that the opportunities for discharge are shared between the approaches in proportions to the demands. Later, we will see how the number of available movements affects the characteristics of the resultant traffic jam.

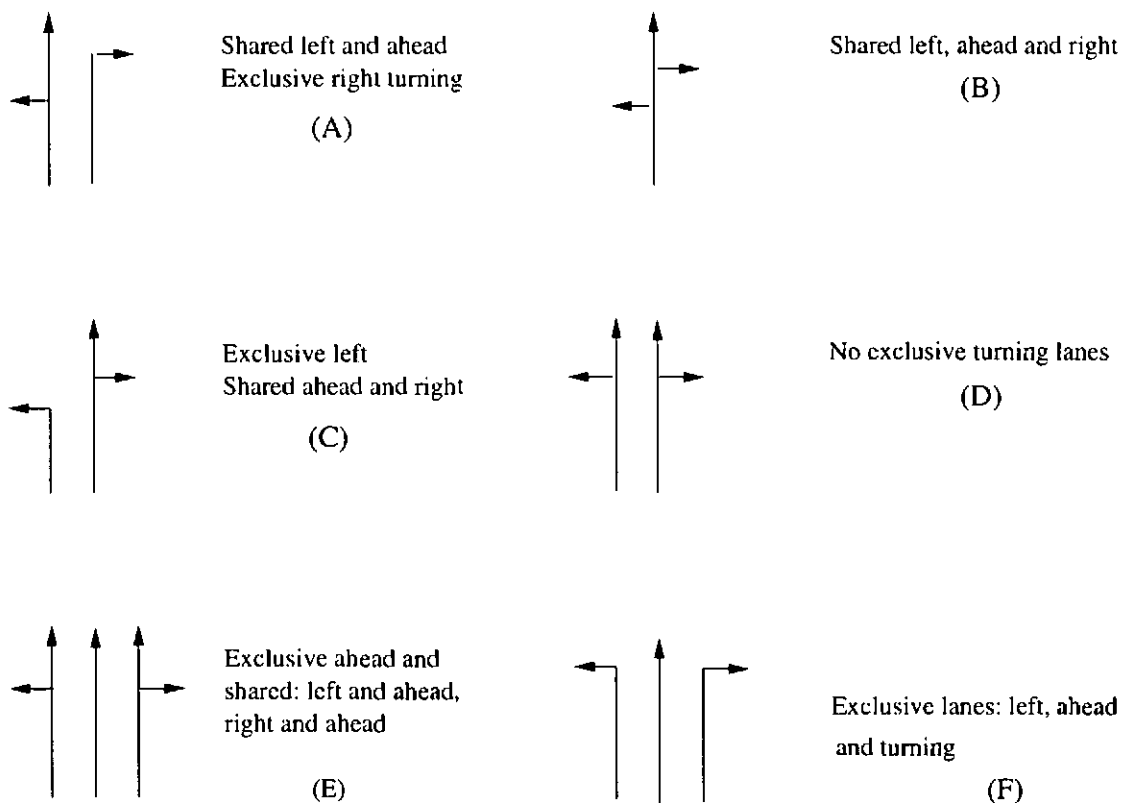


Figure 3.3: *Possible configurations for traffic on a two-way link*

Storage configuration

The storage arrangement of a two-way link is similar to the one developed for the one-way link. However, the availability of three traffic movements: ahead, left and right, complicates the queue storage arrangement, because of the variety of ways in which the segregated regions can be configured. Some of the possible arrangements are shown in Figure 3.3. The examples show separate lanes devoted to each of the available movements, shared lanes allocated to a selection of movements (for example, left and ahead share the same segregated region), as well as a combination of these options. We have chosen the arrangement shown in Figure 3.4. Conceptually, this arrangement reflects the layout shown in Figure 3.3F, in which the link has a number of segregated lanes (for ahead, left turning and right turning vehicles) as well as a reservoir area where turning movements are mixed.

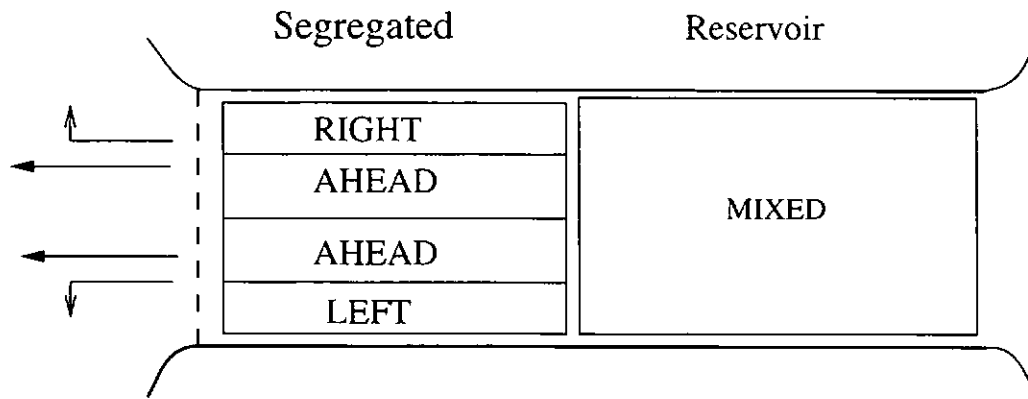


Figure 3.4: *Queue storage arrangement on a two-way link*

Spillback on a two-way link

Suppose an exit becomes blocked, at the right exit, say, then vehicles intending to turn right will form a queue that will spread along the offside lane. Eventually, the queue will extend into the reservoir area where vehicles spread across the whole width of the link, blocking all the traffic in the reservoir area. The use of the upstream reservoir enables us to model the fact that drivers' do not necessarily chose the correct exit lane on entry to a link but defer this decision until later on downstream. Of course, we can simulate perfect lane discipline by allowing the reservoir to be very short.

Figure 3.5 shows the mechanism of spillback graphically. A similar pattern of events arises if the ahead or left-turning segregated regions block back to the reservoir area. Once the reservoir area fills with waiting vehicles the link becomes blocked and traffic may no longer enter this link. This causes queueing on the links upstream. The pattern then repeats itself, until a system of queues has evolved around the original source of obstruction.

Discussion

The two-way model proposed in this section is a preliminary one. It is intended as an exploratory device for identifying the main features of two-way traffic jams, and does not seek to represent all the complexities associated with movement on two-way

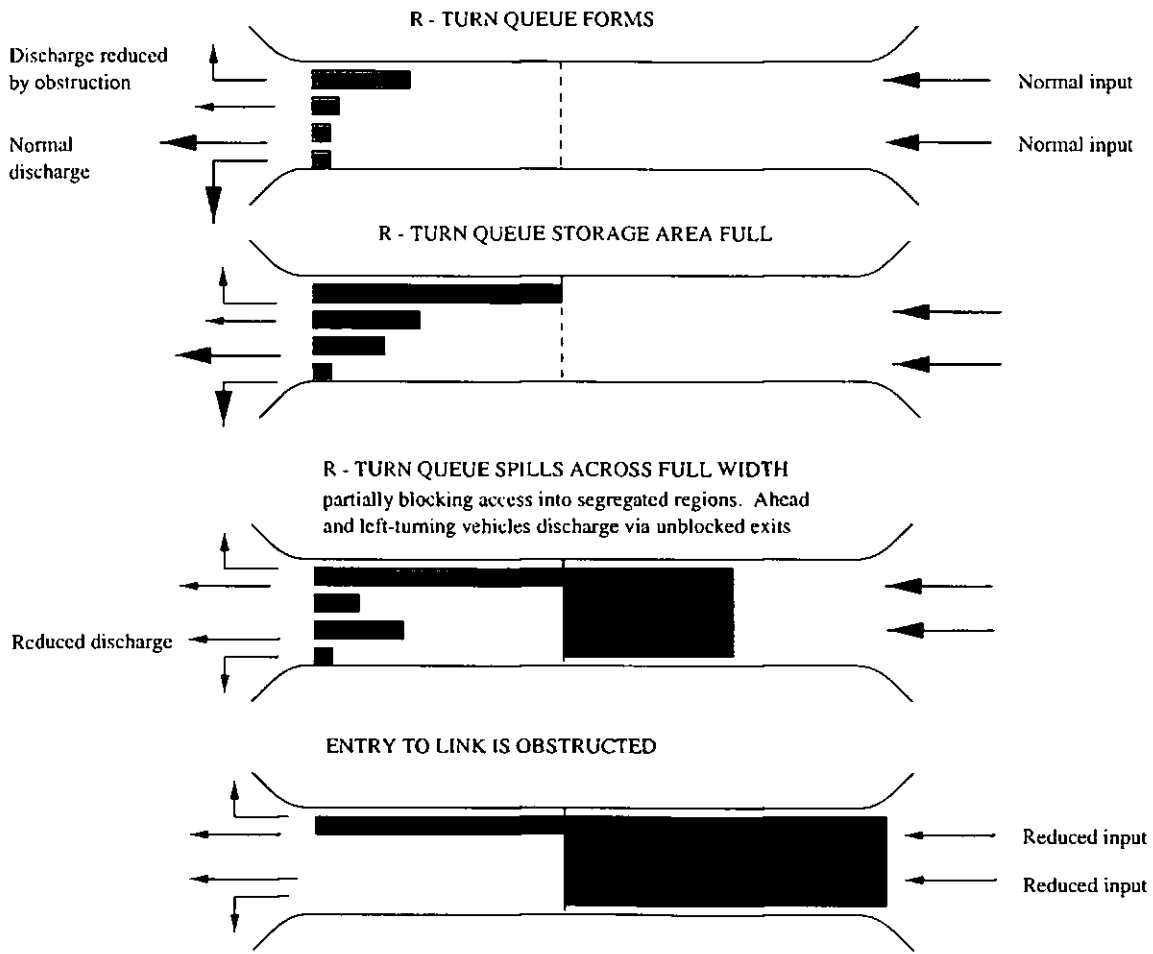


Figure 3.5: Stages of queue propagation along a two-way link where the right turning discharge is obstructed

networks. A number of simplifying assumptions have been made, and the model thus only represents the 'best-case' scenario. There may be additional complications, which have not been modelled, which may cause the system to lock-up much faster than predicted. The implications for traffic engineers might therefore be more severe than expected. However, the model will give a useful indication of the types of situation which are most likely to cause disruption, and point to strategies of how these can be avoided.

Having outlined the underlying assumptions of the traffic simulation tool, we now describe the mechanisms which govern its operation.

3.3 Model control

3.3.1 Model selection

There are two approaches for developing a traffic simulation model such as the one required for this study. These are (a) the construction of a purpose-built model using a suitable programming language or (b) the adaptation of off-the-shelf traffic simulation packages to investigate traffic jam growth and decay.

There are several programming languages which are available for small-scale simulation studies. These languages can be classified into two groups: general purpose and simulation languages. General purpose programming languages include Pascal, Fortran, C, Basic and Ada. Simscript, GPSS, CSL and Simula are examples of popular simulation languages. The translation of a model using simulation languages is preferred over general purpose languages because simulation languages are simple and often suited to the practice of simulation. On the other hand, programming in general purpose languages is comparatively more difficult and more intensive, particularly where there is a need for the replication of in-built routines which are standard in most simulation languages. However, the resulting program is often more efficient in terms of speed of execution and flexibility in program control.

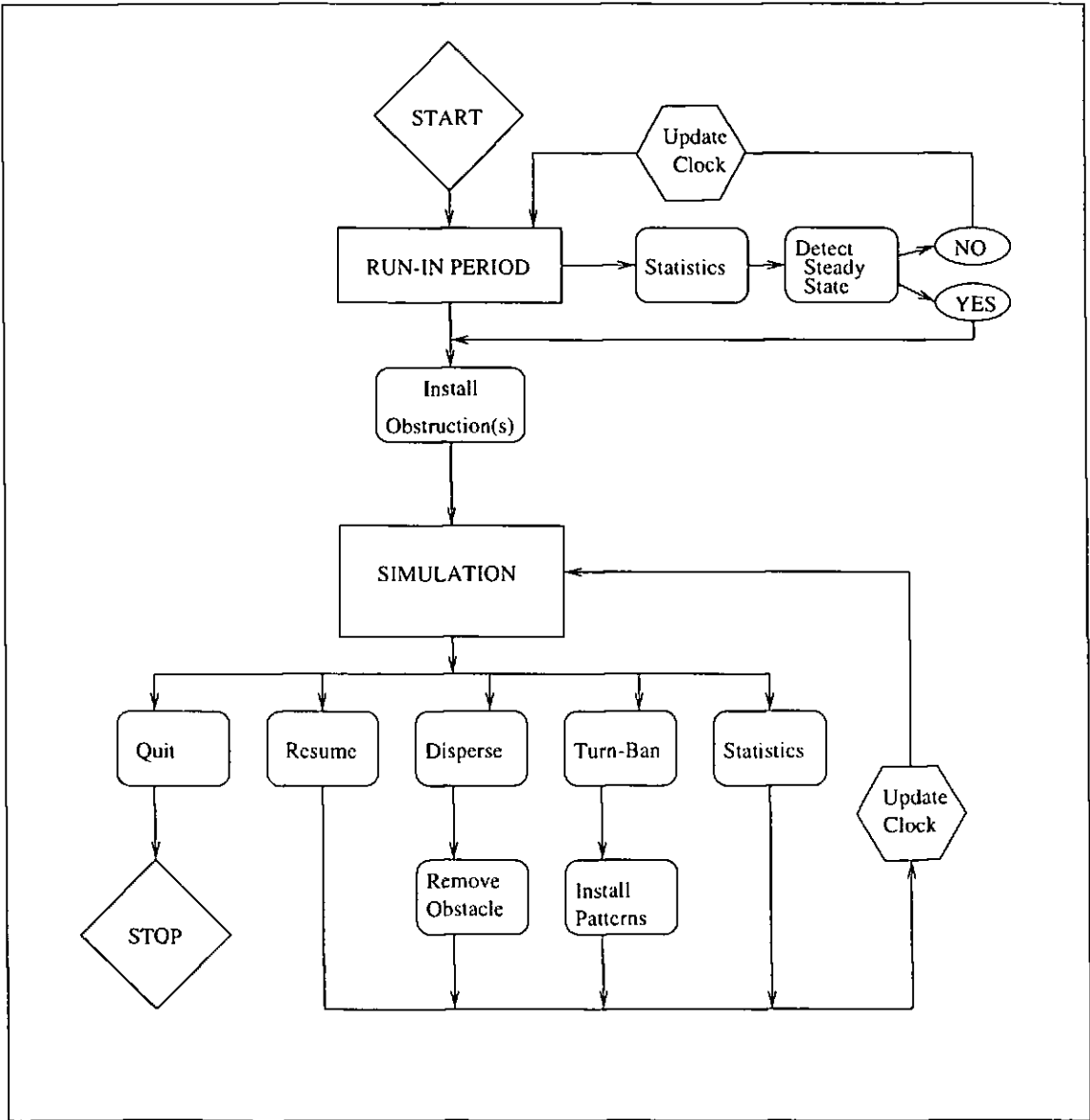


Figure 3.6: Model structure: schematic diagram

The adaptation of off-the-shelf traffic simulation packages is attractive because it enables researchers to focus on the problem at hand, rather than on details to which there are already standard solutions available. However, the models are not always intended for the specific problem and can hence be complex and difficult to implement. There may also be limitations on the size and configuration of the programming environment. Finally, when modifying an existing piece of software, there may be redundant code which can cause slower execution speed. This reduces the software's efficiency as well as its accuracy.

The traffic simulation model developed for this study was created on a *Digital Alpha 3000* workstation using the *PASCAL* programming language. This implementation was selected over the aforementioned options for a number of reasons. Firstly, it provided a platform for a simple yet structured program which would be efficient both in terms of its speed of execution and its flexibility of control. There were also no constraints on the size of network allowed. Secondly, the automatic statistics and input/output procedures generated by standard simulation models and languages were not entirely applicable to the study of traffic jam attributes. Thirdly, the program's modularity make it readable and suitable for constant development and enhancement. Finally, the language generality, programming power, ease of use and availability were amongst the criteria used for choosing *PASCAL* over other general purpose programming languages.

3.3.2 Model structure and operation

The modelling of congestion growth and its subsequent decay has been achieved by dividing the program into three main stages. The stages include a stabilisation period, a jam evolution period and an optional dispersion stage. In developing the simulation tool for the workstation environment, the author has introduced several new features which have substantially enhanced the model's operation and flexibility.

Figure 3.6 describes the operation of the simulation tool via a flow diagram. At

the start of the simulation, vehicles progress freely through the system. During this time, the program gathers statistics which are used to determine the steady state of the road network. This is known as the stabilisation period. It is terminated as soon as this condition has been established.

The growth of the traffic jam may then be initiated by installing an obstruction on the network. Although multiple traffic incidents may be introduced anywhere in the network, we have concentrated on investigating the effects of a single incident which occurs towards the centre of the grid. The severity of the obstruction can be adjusted by experimenting with the percentage discharge parameter, which assumes values between zero (for a total obstruction) and one hundred percent (no obstruction). Partial obstructions are represented as values between these upper and lower bounds.

The software can also be used to investigate the way in which traffic jams disperse by removing the blockage and by introducing counter-measures. These options can be activated via the DISPERSE and TURN-BAN menus shown in shown in Figure 3.6. The graphical display enables the user to evaluate the success of intervention over a considerable time period.

Simulation cycle

In the model, time is divided into a number of equal time-slices. At the end of each time-slice, all the queues are updated according to the inputs and outputs of the particular link. To eliminate any lumpiness that may result from this technique, the duration of a time-slice is typically short. This imposes substantial overheads in terms of program execution, since the sizes of all the network queues are continually being reviewed. Nevertheless, short time-slices impart a smoothing effect throughout the system since the queue lengths do not change substantially from one time-slice to the next, and any variation is evenly spread across the network.

A typical model cycle involves all the east-west roads being processed in turn, starting with the link at the downstream end, and moving progressively upstream from link to

link. Then, all the north-south roads are processed in similar fashion. Initially, this cycle is carried out a number of times to allow the system to stabilise. An obstruction is then introduced on a link near the centre of the grid. This activates the growth of the traffic jam. The cycle is then repeated as required. The growth of the traffic jam with time is presented in an animated visual form.

During each time slice, the outputs from each storage area are taken as the inputs to the appropriate storage areas immediately downstream for the next slice. The simulation model therefore takes account of the effect of one queue on its neighbours.

The program is intended to deal with large networks at high speed, to enable a screen display of the salient features of traffic jam growth in less than real time. The program can handle a jam extending over a one-way grid of 19×19 blocks (ie, 760 links and 400 junctions) in about two minutes. A range of network parameter variations can be explored at a reasonable cost in terms of computer resources.

3.3.3 Model parameters

The existing software includes facilities for manipulating the time interval between successive updates of the vehicle queues, the traffic demands, the proportions of turning vehicles, the saturation capacities, the stopline widths, and the queue storage arrangements. Each of these parameters may be altered individually, but their combined values contribute towards the overall traffic conditions. This means that certain combinations of parameters might give rise to impractical situations. For example, if the saturation capacities are less than the initial demands, then queues will form immediately throughout the network irrespective of the other parameter values. The impact of parameter values on the system characteristics should therefore be evaluated both individually as well as within a global context.

Table 3.1 summarises the main parameters of the model, including the symbol of reference, and the domain for which the symbol is defined. The range of each indi-

vidual parameter will depend on the values of the other parameters. The individual contribution of each of these parameters is discussed next.

The traffic demand can be altered by selecting higher values for congested conditions, and lower values for light, free-flow conditions. Clearly, the value of μ will be related to the saturation capacity parameter κ , with respect to the resultant type of flow conditions. However, given a fixed value of κ , (eg $\kappa = 50$ veh per min), light conditions can be generated by setting $\mu = 10$ veh per min and congested conditions can be provided by setting $\mu = 30$.

The propensity for a vehicle to turn is controlled by fixed probability values which are the same for each intersection and are used to determine the proportions of turning vehicles at each intersection. Thus, if p represents the proportion of turning vehicles, where $p \in \{0 \dots 1\}$, then $1 - p$ is the proportion of vehicles moving ahead. For the two-way grid network, the proportion p , of turning vehicles, may be split further between left turning and right turning vehicles. In other words, if $q \in \{0 \dots 1\}$, then pq is the proportion of right turning vehicles and the remainder, $(1 - q)p$, is the proportion of left turning vehicles. The proportions are subject to numerical rounding to ensure that vehicular flow is always integer. This improves the software's speed of computation and reduces run-time overheads.

The saturation capacity is defined as the maximum number of vehicles able to cross to the upstream intersection during a single time-slice. The program implicitly assumes that the turning saturation capacity, per unit width of stopline, is lower than the corresponding ahead one. This is because it takes longer to negotiate a turn than it does to move in the ahead direction. The turning saturation capacity per unit width of stopline is therefore represented as a percentage of the ahead saturation capacity which can be specified by the user. The ahead saturation capacity is referred to as κ .

In developing the simulation tool, the author has introduced a number of additional parameters which enable extensive testing of the model to the underlying assumption

of blocking-back. These parameters include the stopline width allocated between the segregated ahead and turning queues and the degree of interaction between segregated and reservoir queues.

The stopline widths can be altered to suit individual requirements. The allocation of stopline width between ahead and turning vehicles is governed by a single parameter, which in principle, can be set in the range $\{0 \dots 1\}$. The proportion of stopline width devoted to the ahead queues is denoted by α , with the remainder of the width $(1 - \alpha)$ allocated to the turning queues. For example, $\alpha = 0.66$ denotes the situation of two lanes for ahead traffic and one for turning vehicles. For practical use, the set of possible values for the α parameter has been restricted to the set $\alpha \in \{0.33, 0.50, 0.66, 0.75, 0.80\}$, as these form a representative set of road network configurations.

The queue storage configuration can be manipulated by varying the relative queue storage capacities within the segregated and reservoir areas on the link. This is controlled by a parameter, denoted by σ , which represents the proportion of the area of a link devoted to storage of segregated queues (the proportion devoted to the upstream reservoir will then be $1 - \sigma$), with $0 < \sigma < 1$.

The degree of interaction between the queues that may form along different sections of the link can also be altered. Thus, the blocking effects experienced on a one-way link can be monitored by selecting a value between zero and one. This value, denoted by s , represents the degree of interaction between ahead and turning queues.

In addition to the network control parameters, the model includes facilities for defining the duration of a single time slice, denoted by τ . The time slice can be fixed to the nearest second, although typically, a time slice lasts for one minute. The duration of the simulation period is maintained as a number of time slices, which can be altered depending on the individual requirements of the user. Note that under free-flow conditions and subject to no obstructions, a vehicle can travel across the network in

Parameter name	Reference symbol	Unit of measure	Domain
Demand	μ	discrete	positive integers
Turning proportions	p, q	continuous	$0 \dots 1$
Saturation capacity	κ	discrete	positive integers
Stopline widths	α	discrete	$\frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}$
Segregated proportion	σ	continuous	$0 \dots 1$
Spillback factor	s	continuous	$0 \dots 1$
Percentage discharge		continuous	$0 \dots 100$
Cycle time	τ	continuous	Minutes/seconds

Table 3.1: *Computer simulation model parameters*

approximately ten time-slices.

3.3.4 Graphical features

The simulation model incorporates a number of criteria which make it particularly suited to area-wide traffic congestion modelling. The model is interactive and it is intended to deal with large networks at high speed. The development of an area-wide traffic jam can be observed in much less than real time. External counter-measures can be introduced anywhere in the network. The graphical display is online, in other words, the operator can view the effects of intervention over a considerable time period. This means that the success of control strategies can be assessed in the short and long term.

The display of congestion patterns over a wide area is important for developing and determining the effectiveness of control strategies, particularly when the elimination of the original queues merely leads to their displacement elsewhere in the network. The online graphical display plays an integral part in the spatial analysis of traffic jams and so we next describe how to interpret the visual results provided by the model.

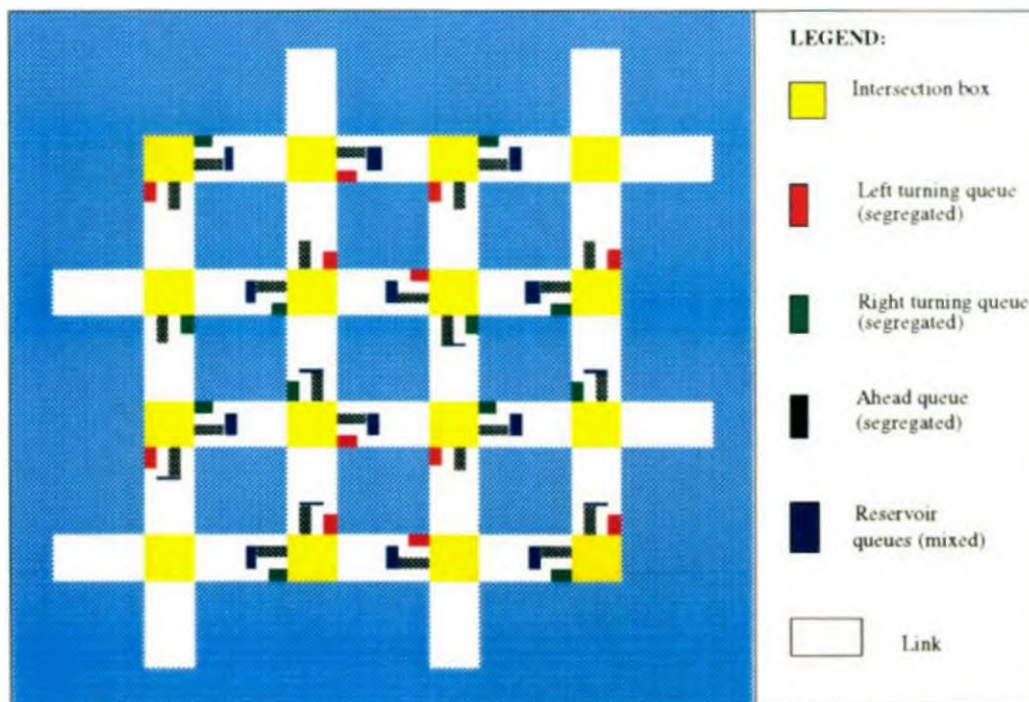


Figure 3.7: *Graphical representation of one-way 4 × 4 network*

The screen display

The graphical display consists of a series of white roads which are drawn against a light-blue background. The roads run horizontally and vertically across the grid. The intersections are shaded in yellow. This representation recalls the yellow boxes painted at the junctions of busy roads.

The storage arrangement of a link is divided between a segregated zone, where individual turning movements are separate, and a reservoir area, where the turning movements are mixed. The reservoir region occupies the whole width of the road and is shaded in blue. The segregated region is divided between the different movements. Left turning queues are coloured in red, right turning queues in green and ahead queues are shaded in black. Figure 3.7 shows the layout of a 4 × 4 one way road network. A similar scheme has been adopted for the two-way version. The extended roads represent the sources, whereas the severed ones represent the sinks. Vehicles travel from source to sink in the east west and north south directions.

3.4 Summary statistics

In addition to the graphical representation of congestion, the simulation model provides summary statistics which are used to quantify system characteristics and the severity of the traffic jam. At the end of each time slice, the program calculates the total number of vehicles present on the network, and the total number of blocked links within the traffic jam, as well as the total number of arrivals and departures of vehicles to the system. We shall now describe how these statistics can be used to measure specific attributes of the system, such as its steady state and the total delay incurred by vehicles that have become trapped by congestion.

3.4.1 The road network's approach to stability

At the start of the simulation, vehicles progress freely through the road network. Initially, there will be few vehicles in the system, but as time progresses, the number of vehicles will increase. Eventually, there will be little variation from one time slice to the next. In other words, the total input will be balanced by the total output. When this occurs, the system is said to have achieved its steady state.

In the original model, traffic flow was deterministic and, providing the difference between system input and output remained constant, it was possible to establish the system's steady state. The introduction of stochastic variation in the simulation tool, implied that this condition might not necessarily hold, ie the difference in system input and output might fluctuate between successive time-slices. This problem has been solved by employing time-series methods to determine the stabilisation of vehicle flows. The technique used is described in Appendix A.

Once the stationarity condition has been established, the stabilisation period is terminated, and among other, an estimate is made of the mean and standard deviation of the number of waiting vehicles. (These statistics are used when calculating the total delay in the system.) The gradual approach of the one-way road network to

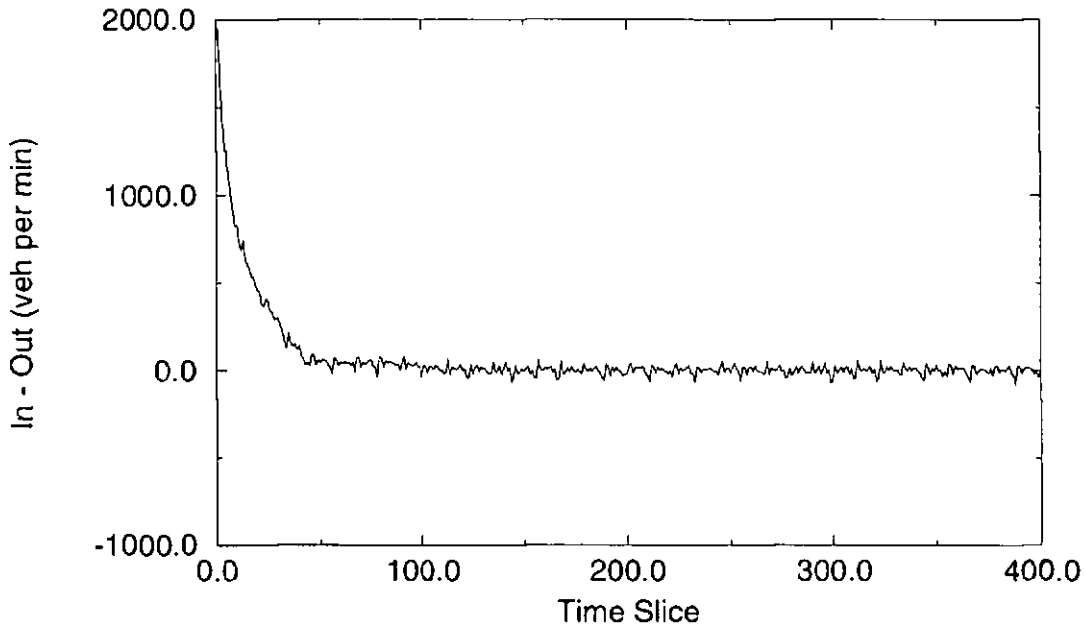


Figure 3.8: *Network's approach to stability*

its steady state is shown in Figure 3.8. *Further examples showing the convergence of the network to its steady state using different random seeds are shown in Appendix A.* Similar patterns have been observed for the two-way network.

3.4.2 Quantifying the severity of a traffic jam

One needs to be able to quantify the severity of a traffic jam, so that comparisons can be made between different traffic management measures, for example. There are several possibilities, which generally fall into one of two categories: measures that describe conditions at specific intersections, and measures that describe conditions in the system as a whole.

The total delay experienced by road users is an example of a useful indicator of the magnitude of congestion either at a particular intersection or over an entire system. It is one of a selection of measures proposed by Pignataro et al (1978) for the above categories.

The nature of traffic jams is that they are usually spread over a wide area, ie, congestion is not confined to isolated intersections but is propagated over a wide area. Consequently, we will measure the total delay due to a traffic jam as a function of the system's performance rather than as a function of individual intersections.

Another way of measuring the severity of a traffic jam is to count the number of blocked links enclosed within its boundary. The jam boundary is defined as an imaginary polygon joining the outer extremities of the queues at any particular moment in time. This is a useful measure when dealing with traffic jams that are spread out over a large area of network.

Estimating the total delay

In classical queueing theory, a random variable of interest, is the delay experienced by an individual. But in the traffic situation, (a) at certain times no departures are possible (eg during red periods), (b) vehicles do not necessarily depart in order of arrival (where there are multiple streams of traffic), and (c) the service points are often linked. Consequently, the random variable of interest in the traffic situation is the *total delay* due to congestion. It is also important because it gives a direct indication of the financial cost due to congestion. In any case, the mean individual delay could be obtained by dividing the total delay for a cycle by the number of vehicles arriving during the cycle.

In the simulation tool, vehicles are not individually tracked and as a result we needed to develop a measure which would estimate the delay incurred by the traffic jam. This could then be used to quantify its severity. This has been achieved using the following technique.

If we consider a deterministic system in which the average number of vehicles on a link when the system is in the steady-state is denoted by u . Due to congestion, there are v vehicles on the link. Hence, the extra delay experienced by vehicles which is

due to congestion is given by

$$\sum_{t=t_0}^{t=t_n} \sum_{links} (v - u) \Delta t$$

where Δt denotes the duration of a single time-slice.

In other words, the total delay in the system is the time waited by the excess vehicles in the system. This idea can also be applied to a stochastic system. Two methods have been developed for calculating the total delay incurred as a result of a traffic jam. The first method compares the total number of vehicles in the system in the steady state with the corresponding number in the congested state. The estimated delay denoted by Δ_i^* , represents the total delay in the system incurred as a result of the traffic jam, assuming that the original obstruction has been removed. This method will be referred to as the **wait totals** method. The method is described in full in **Appendix B**.

However, this method is deficient in two ways.

1. There is a time lag between the placing of the obstruction and the observation of an increase in delay. This is caused by the fact that excess levels of vehicles in obstructed queues are compensated by lower levels in queues downstream.
2. Having delay dependent on the total number of vehicles in the system may cause an *underestimation* of the actual delay particularly when strategies rely on rapid expulsion of vehicles from the system. In addition, random fluctuations in the demand levels may generate some undersaturated links which would also reduce the delay experienced in the network.

A second method has been devised to cope with the aforementioned problems. This method uses a direct approach to the excess queueing of vehicles at each link in the system. The average size of a queue which has formed on a link when the road network is in the steady state is recorded prior to the simulation phase. A comparison is then made between this value and the size of the queue during the congested state. The

difference between these two values represents the excess queueing due to congestion and as a result forms a measure of delay. The total delay, Δ_i^* , represents the total delay incurred in the system as a result of the traffic jam on the network. This method is described in detail in **Appendix B** and will be referred to as the **positive excess queue method**.

This method addresses the problems associated with the first method but in so doing, it may overestimate the actual delay experienced by drivers. This is because random fluctuations in the demand levels may cause some links to become temporarily oversaturated. This excess will be considered as delay caused by congestion. The implications of this are that delay may, in some cases, be over estimated using this technique.

Hence, whilst the first method may underestimate the total delay, the second may overestimate it. The actual delay incurred by the presence of a traffic jam is estimated as the mean of these bounds. We shall use this statistic to compare the relative efficiency of certain control strategies in various congested situations.

3.5 Conclusions

This chapter outlined the fundamental concepts associated with the traffic simulation model. The model was developed for one-way and two-way rectangular grid networks. It included various facilities which could be manipulated to suit different traffic conditions. There were also a number of run-time options which could be activated during the simulation period, such as the installation and removal of obstructions and the implementation of external counter measures.

An important feature of the models was the way in which queues were handled. On each one-way link, there were assumed to be two segregated approach queues, one for the 'ahead' movement and one for the 'turning' movement at the downstream junction. Upstream from the segregated queueing area was a common 'reservoir' area

occupying the combined width of, and feeding into, the segregated queues. A similar arrangement applied for the two-way link. Whenever a segregated queue overflowed its segregated storage space it was assumed to spill back into the reservoir area and block vehicles from proceeding into either of the segregated queues. This interference between queues is a central feature in area-wide traffic congestion, particularly to the form of congestion known as 'gridlock'.

The severity of congestion was quantified by estimating the total delay incurred as a result of a traffic jam or by measuring the number of blocked links enclosed within its boundary. This was thought to be a useful measure when dealing with traffic jams that were spread over a substantial part of the network.

The next chapter uses the simulation model to formulate a working theory which explains the processes of traffic jam development and dispersal in rectangular grid networks.

Chapter 4

The development and dispersal of traffic jams on rectangular grid networks

4.1 Summary

This chapter uses the traffic simulation model, introduced in the previous chapter, to investigate the fundamental attributes of idealised traffic jams, and considers how they vary in response to changes in network conditions and model assumptions.

The characteristics of traffic jams are explored by examining the underlying processes of growth, dispersal and starvation. Some general conclusions are drawn about the spatial structure of traffic jams in grid networks. This leads to an assessment of the relationship between traffic jam structure and its dispersability.

4.2 Introduction

Traffic jams are becoming more frequent everywhere, but not all take the same form. Some are relatively simple: they consist of isolated queues associated with bottlenecks

on long saturated roads. Others have a more complex spatial structure, for example area-wide jams in town centres. Typically, area-wide jams involve the propagation of queues from one road link to another. Subsequent interaction of the queues leads to a phenomenon commonly referred to as 'gridlock'.

Within certain limitations, the development of an area-wide jam can be analysed as a spatial phenomenon. The queues propagate in a predictable way to form a tree-like structure having regular geometrical features. Some useful indicators about the behaviour of jams and ways of inhibiting their growth can be gained by investigating this process with the help of models developed specifically for this purpose.

In this chapter, we describe the formation, growth and dispersal of a traffic jam on one-way and two-way rectangular grid networks, and investigate the likely effectiveness of various traffic management strategies in deterring traffic jams, or dissipating them once they have formed.

While the overall pattern of jam development will be the same for one-way and two-way grid networks, there are some important spatial differences. Consequently, we shall examine the structure of traffic jams in one-way grid networks first, and then explain how similar principles apply in two-way grid networks.

The networks are idealised uniform grid networks with an inelastic pattern of traffic demand. Dynamic effects of the type generated by traffic signals are deliberately omitted, allowing one to observe more clearly the underlying spatial structure, unobscured by local variations in network geometry and traffic flow.

4.3 Spatial characteristics of one-way traffic jams

It is useful to consider the study of traffic jams in two phases, the growth phase and the dispersal phase. The growth phase can be initiated by installing a blockage towards the centre of the network. The dispersal phase can be activated by removing

the obstruction or by introducing external counter-measures.

The stages of development and dispersal are affected by a secondary process: starvation. In general, starvation arises as a result of vehicles unable to access certain regions in the network due to the traffic jam's development. Starvation affects the spatial characteristics of the traffic jam, as well as the shape of its overall boundary. The dispersal process of the traffic jam is also affected by the extent of starvation.

There are few papers which deal specifically with the processes of jam development and dispersal. One exception is the report by Abbess and Wright (1991) which describes two models for traffic jam growth, the first analytic and the second, a computer simulation model. Both demonstrate the features underlying the process of traffic jam growth and point to possible traffic management measures which aim to minimise the spread of jams and inhibit gridlock phenomena. The salient points of the study are summarised in Wright and Roberg (1997), together with further investigation concerning the asymptotic shapes of one-way traffic jams.

In this section, we briefly outline the main concepts of this theory emphasising the predictions concerning the spatial characteristics of one-way traffic jams. The analysis is then extended by considering the sensitivity of the results to the assumptions underlying the model. This leads to additional features of traffic jam development and dispersal in one-way grid systems. There follows a discussion of the relationship between traffic jam structure and its dispersability. The implications for traffic jam control are considered separately.

4.3.1 Theoretical background

When an obstruction is set on a link which is located towards the centre of the network, it may reduce the *ahead* discharge or it may reduce the *turning* discharge by a factor c , say. If, say, the turning discharge is reduced, a queue will build up until it fills the the turning storage area. It is then assumed to spread across the entire width

of the road, blocking the upstream reservoir. Within this area, ahead and turning vehicles will be mixed together in proportion to their respective demands. It follows that the throughput of all vehicles will now be reduced by a factor c . The queues in the reservoir will in turn build up and eventually spill back to the upstream entry. This sequence of queue spillback and blocking was explained in the previous chapter and shown diagrammatically in Figure 3.2.

If the turning discharge is reduced, the turning area will fill up first, and then the upstream reservoir, in which case, the overall link propagation time t_{LR} will be the sum of individual spillback times for the turning storage area and the upstream reservoir.

However, if the ahead discharge is reduced, then the overall link propagation time, t_A , will be the sum of the individual spillback times for the ahead storage area and the upstream reservoir.

Note that within the theoretical framework of this model, traffic demand is spatially uniform everywhere within the system, ie, in the absence of any obstruction, the 'ahead' flows at all the stoplines are the same, and the turning flows at all the stoplines are the same irrespective of whether they are turning left or right (but not necessarily equal to the ahead flows).

It is possible to describe the individual link propagation times t_A and t_{LR} without having to make reference to the widths or lengths of the segregated storage spaces and upstream reservoirs. The effects of varying the queue storage arrangement can be represented entirely in terms of the stopline width allocation parameter α and the segregated proportion parameter σ . This greatly simplifies the problem so that we can write, for example,

$$t_A - t_{LR} = \frac{N}{q(1-c)} \frac{\sigma(\alpha - 1 + p)}{p(1-p)} \quad (1)$$

See Appendix C for the derivation of this result.

Spatial structure of idealised traffic jam

Once the link has filled up with queueing traffic, its capacity to accept vehicles from either of the two upstream links feeding into it will be affected in the same way: it will be reduced to a value c times the demand. Effectively, the initial obstruction is propagated to the two upstream links, which will now behave in a similar way.

Hence, on the model network, a traffic jam will spread out from a blockage, repeating the same sequence of events on all the links until the system is full of queueing vehicles. The queues grow upstream along well-defined paths, which can be identified by searching for the quickest propagation 'routes'.

A blockage which occurs at the downstream end of a particular link, may affect the 'ahead' vehicles, or the 'turning' vehicles, or both. After the first link becomes full, the subsequent pattern of queue propagation will be the same for all three cases: the only element which varies is the time taken for the first link to overflow.

However, the growth paths are affected by the relative values of the link propagation times t_A and t_{LR} . It is possible to distinguish between three possible configurations corresponding to the following three conditions:

1. $t_A < t_{LR}$ (Type I)
2. $t_A > t_{LR}$ (Type II)
3. $t_A = t_{LR}$ (Type III)

(Note that for the one-way model we have assumed $t_L = t_R$ since traffic demand is spatially uniform everywhere. Consequently, there is no need to consider the cases $t_L < t_R$ or $t_L > t_R$ individually. But, for two-way traffic jams one needs to consider the effect of unequal turning proportions on the spatial structure of two-way traffic jams. This is dealt with in Section 4.4 later in this chapter).

From eq(1) these conditions are respectively equivalent to

1. $\alpha < 1 - p$
2. $\alpha > 1 - p$
3. $\alpha = 1 - p$

Condition (3) represents the case where the ahead queue tails back at the same rate as the turning queue. This corresponds to the situation where the stopline widths are allocated in exactly the same proportion as the traffic demands. In terms of the simulation parameters, this condition can be represented as $\alpha = 1 - p$. In practice, this would not often occur and it is hence only of theoretical interest.

Conditions (1) and (2) on the other hand, represent unequal rates of propagation. The first condition implies that the ahead queue propagates faster than the corresponding turning queue. This would occur when $\alpha < 1 - p$ since in this case, relative to the demand, less stopline width is devoted to the ahead queue causing it to tailback more rapidly than the turning one.

By contrast, the second condition ($t_A > t_{LR}$), describes the reverse argument: relative to the demand, less stopline width is allocated to the turning queue which causes it to spillback faster than the corresponding ahead one. This corresponds to the situation $\alpha > 1 - p$. These conditions affected the shape of the traffic jam as follows:

When $t_A < t_{LR}$, queue growth takes place most quickly in a straight line along one axis of the grid network. Other secondary queues branch out in straight lines at right angles from this primary queue, and yet more queues branch out from the secondary queues, and so on. The resulting propagation paths are shown in Figure 4.1. The pattern is dominated by four primary paths (shown as thick black lines), that radiate from the source in four compass directions, dividing the system into four quadrants. The boundary of a Type I traffic jam is predicted to be diamond-shaped, centred on the initial obstruction.

The shape of a Type III traffic jam is closely related to that of a Type I - the only

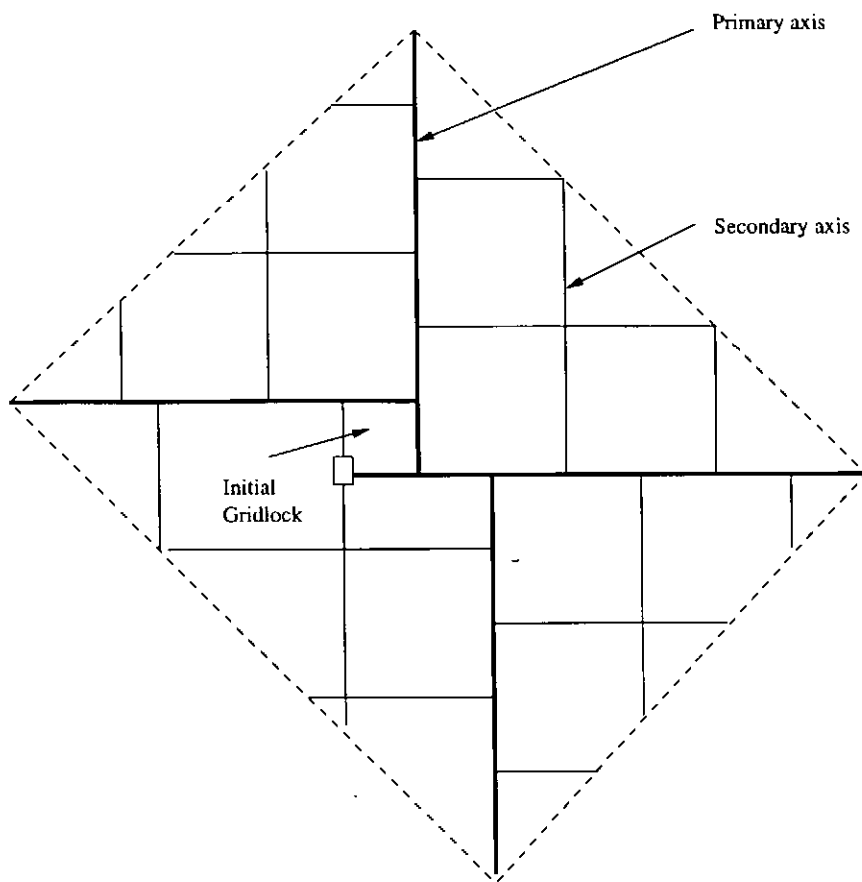


Figure 4.1: *Spatial configuration of Type I jam ($\alpha < 1 - p$)*

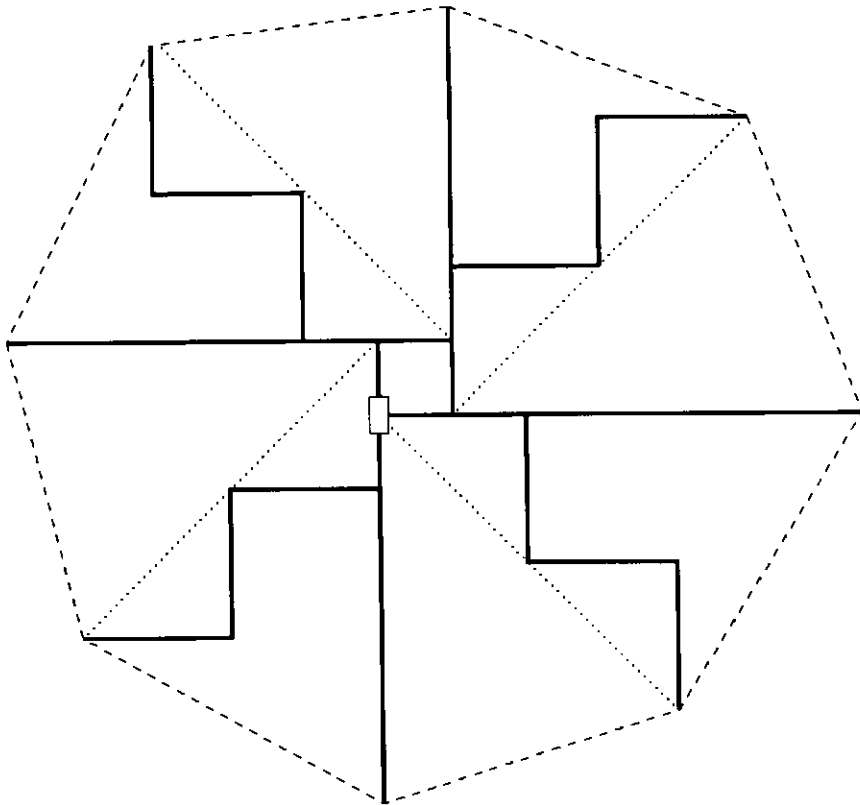


Figure 4.2: *Spatial configuration of Type II jam ($\alpha > 1 - p$)*

difference being that the rate of jam expansion for a Type III traffic jam is the same in all four directions producing a perfectly regular diamond shape. This is not necessarily the case for Type I traffic jams whose diamond shape can be elongated in any of the four directions depending on the relationship $t_A < t_{LR}$.

However, when $t_A > t_{LR}$, the pattern is different. The queues develop as small clusters of tightly packed 'curls' and the growth paths are relatively tortuous. In this case, (Type II), the overall pattern can be divided into eight sectors, as shown in Figure 4.2, and in the limit, the boundary of the emerging jam is predicted to be octagonal. It is not a regular octagon, although it possesses four axes of symmetry. As t_A increases still further, so that the ratio t_A/t_{LR} approaches infinity, the octagon approaches the shape of a square. These results were found by tracing out the queue propagation paths through the network.

While the analysis is useful in predicting the asymptotic shapes of traffic jams in one-way grid networks, it does not take into account randomness in the vehicle arrival times, nor the 'starvation' phenomenon mentioned earlier. Furthermore, it cannot be applied to study the process of dispersal, since the theoretical model treats each queue independently and cannot model the effects of queues interacting with each other. To see how these factors might affect jam structure, we must turn to a simulation model.

Using the simulation tool introduced in the previous chapter, we point to the main aspects of traffic jam development, and comment on the role of starvation in this process. We then describe the traffic jam dissipation mechanism, which cannot be explained using the simple theory, and discuss the relationship between traffic jam structure, and its inherent dispersability.

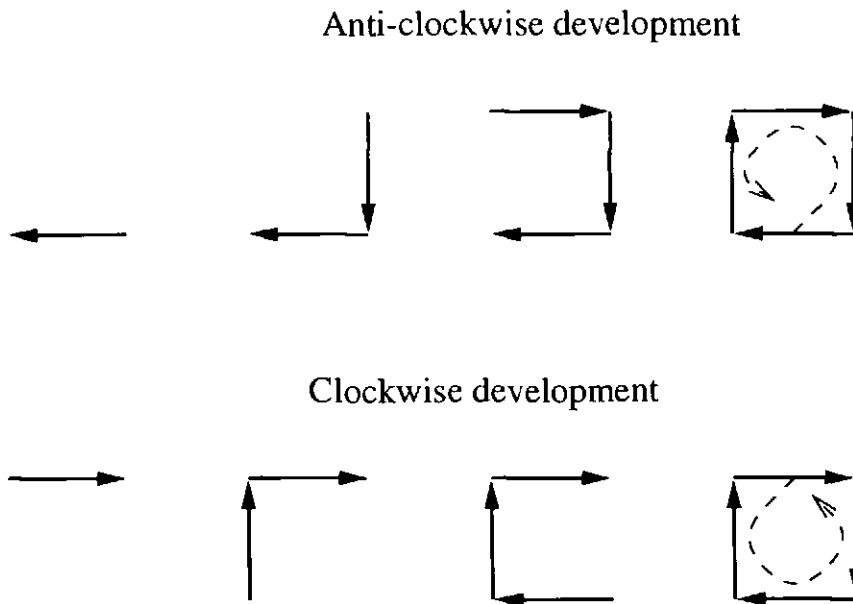


Figure 4.3: *Anti-clockwise and clockwise development of gridlock cycles in one-way network*

4.3.2 Growth pattern of simulated traffic jam

Stages in development

In the idealised model, traffic jams may be initiated as a result of obstructions in the network. As a result, there will be traffic queues on one or more links, but the queues are effectively *independent*. Then, queues will begin to tail back, causing interference with the flow through the junction. This is the *interaction* stage. The final stage in this process occurs when queues encircle the original source of congestion, creating a loop of locked vehicles. This stage is known as *gridlock* and was introduced in Chapter 2 with reference to the theory of cellular automata. A similar phenomenon arises in telecommunication networks, Glass and Ni (1994), where, for example, a packet of information may wait for a network resource to be released by a packet that is, in turn, waiting for the first packet to release some resource. This produces the circular wait condition known as gridlock.

On a one-way network, gridlock cycles can form in one of two ways. These are

demonstrated in Figure 4.3. The top gridlock cycle is created as a result of vehicles unable to turn left, whereas the bottom one forms as a result of vehicles unable to turn right. The overall pattern of queue development is the same for 'left-turning' and 'right-turning' gridlock cycles. The only difference is the direction of development: a left-turning cycle triggers anti-clockwise jam evolution, whereas a right-turning one causes clockwise development.

Similar patterns of gridlock formation have been simulated by Shepherd (1994), who used the NEMIS model adapted for a one-way grid. Despite major differences between the two simulation models, the conclusions are similar. For example, Shepherd caused the gridlock by specifying internal destinations around a central square, whereas we have triggered gridlock by reducing the capacity of a single link. Both studies found that gridlock was more likely to occur as the percentage of turns within the 'central square' increased. Similarly, it was observed that once gridlock had formed, it was impossible to clear even if the demand was dropped to zero. This will be discussed later, with reference to the blocking-back assumptions of the simulation tool.

The investigation of one-way, area-wide traffic jams has been conducted by the author using a 20×20 grid network. The growth of a traffic jam was initiated as soon as the network had achieved its steady state. A regular pattern of queue development was observed as a result of installing an obstruction. The process is shown in Figures 4.4, 4.5 and 4.6. These figures show the development of a simulated traffic jam around a 'right-turning' gridlock cycle. The network conditions are typical of a Type I traffic jam. In Figure 4.4, a queue has formed as a result of an obstruction, leading to initial interference with upstream junctions. Eventually, this leads to the formation of a gridlock cycle, shown in Figure 4.5. The traffic jam then expands rapidly across the entire network, as seen in Figure 4.6. The traffic jam is bounded by a characteristic diamond shape.

Referring to Figure 4.4, one can observe that as the first link fills with queueing traffic, its capacity to accept vehicles from either of the two upstream links will be

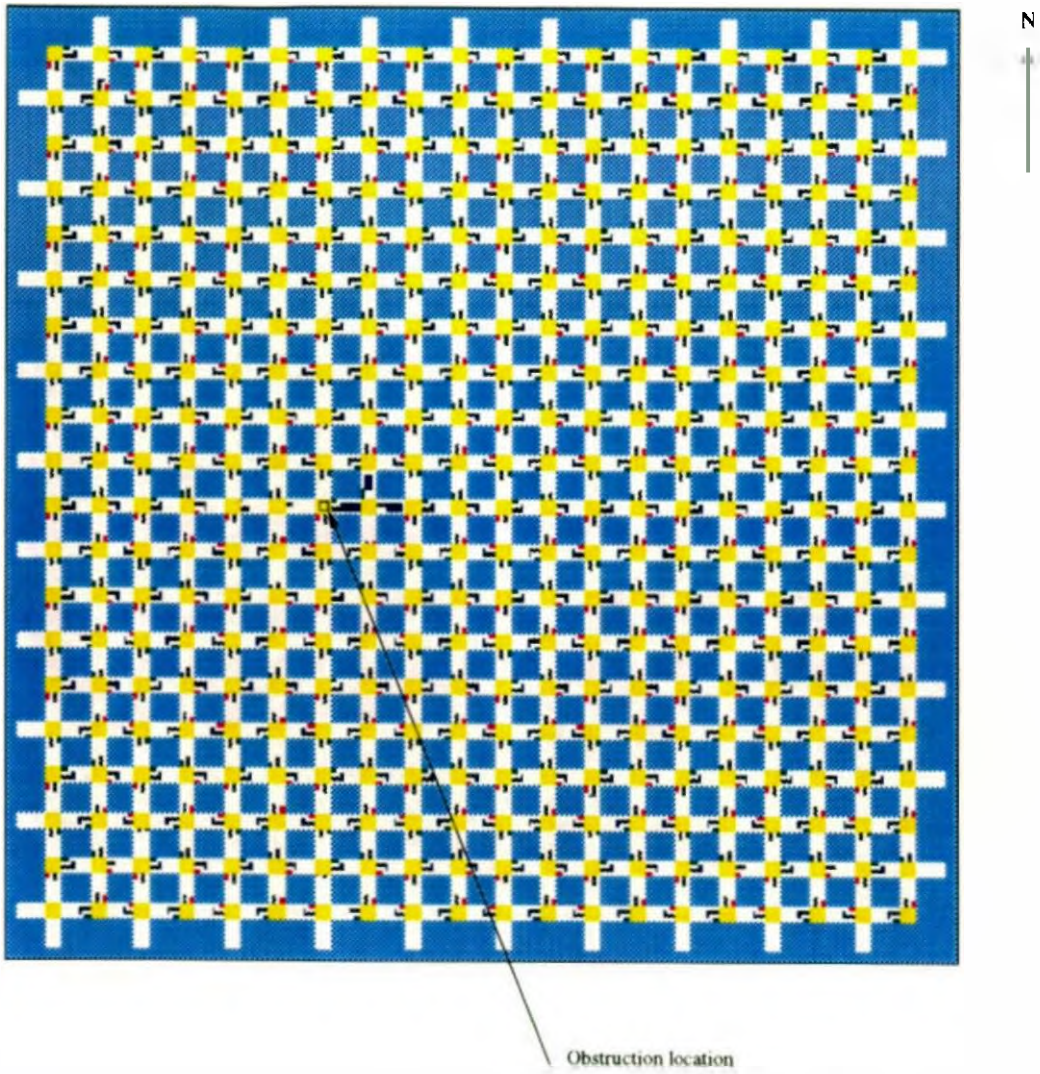


Figure 4.4: *Initial queue formation*

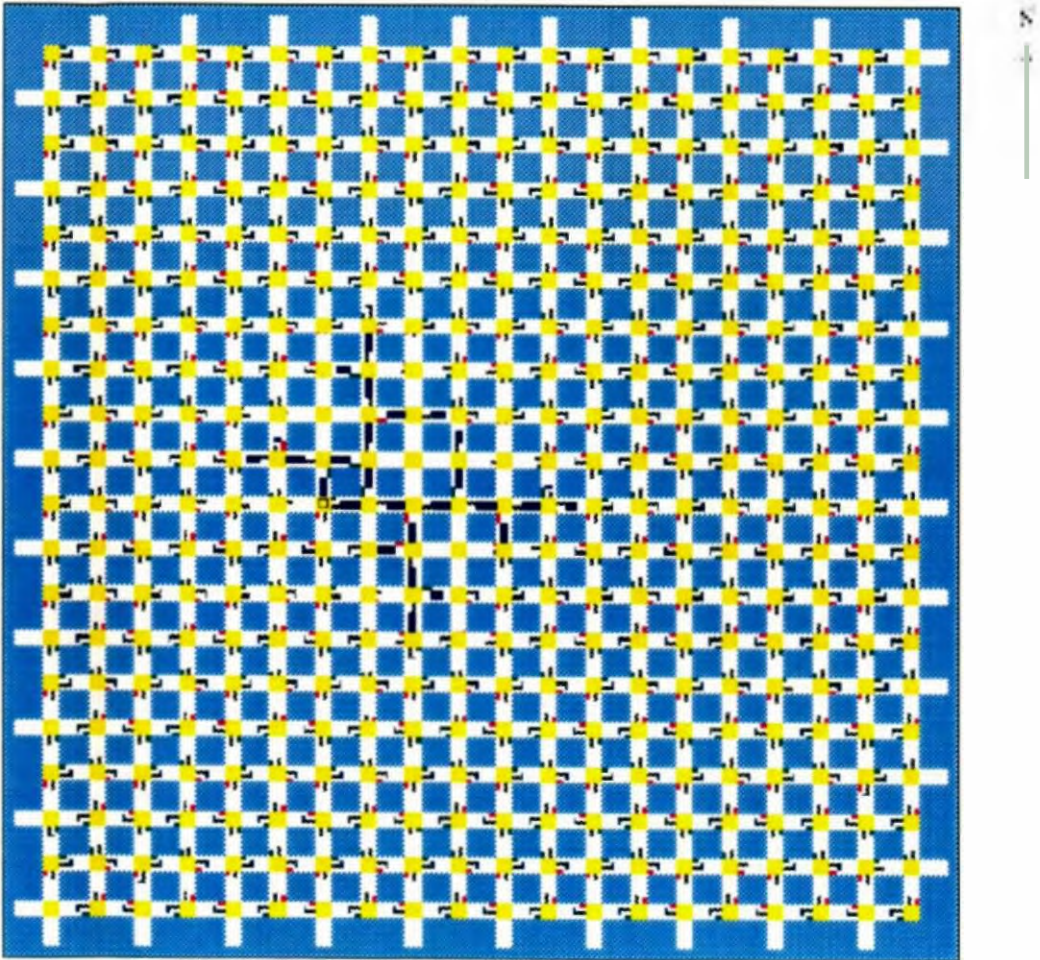


Figure 4.5: *Right-turning gridlock cycle at centre of traffic jam*

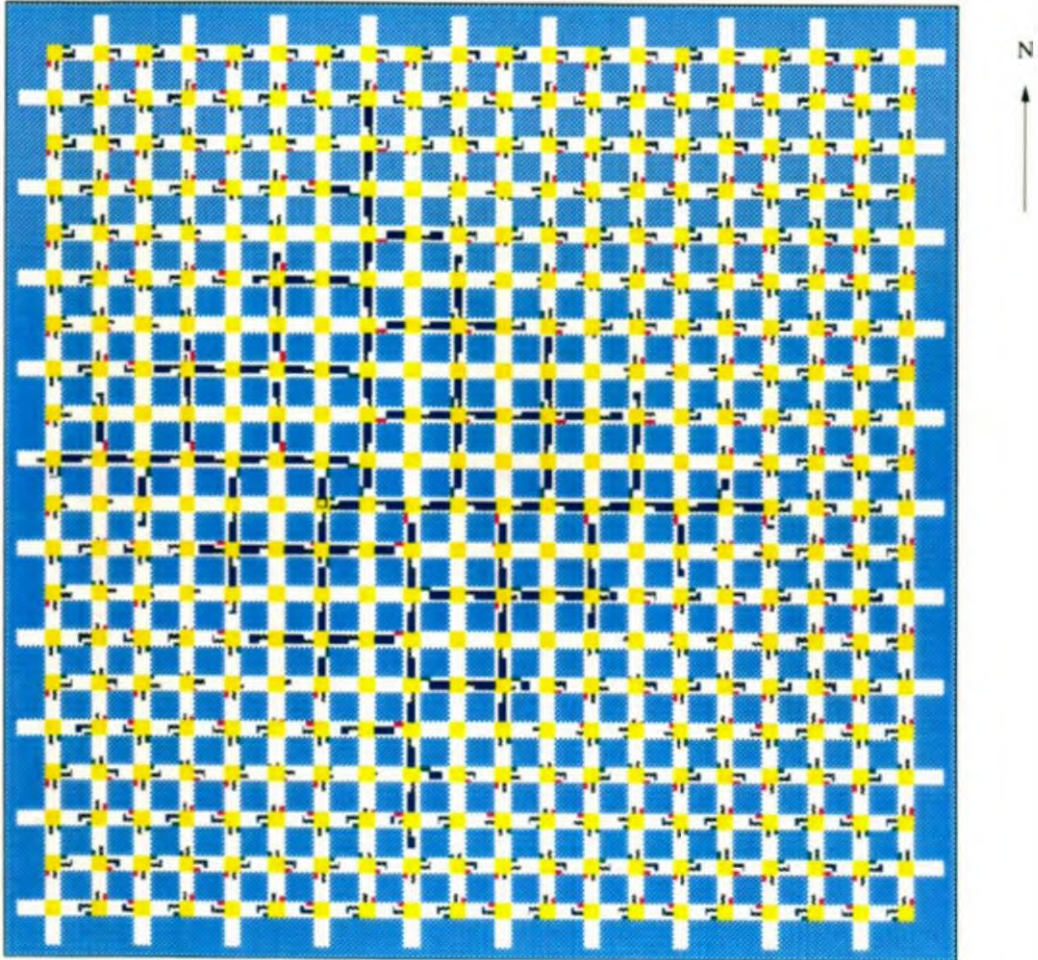


Figure 4.6: *Diamond envelope of out-way traffic jam*

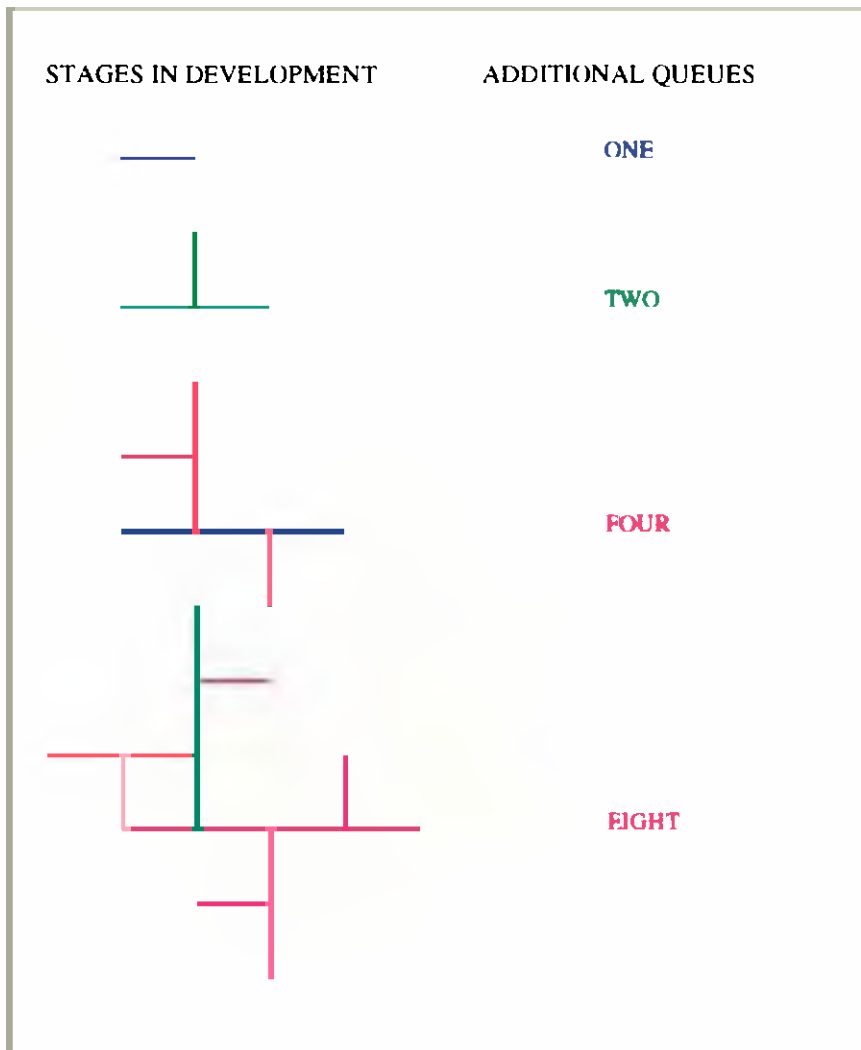


Figure 4.7: Queue propagation paths for a traffic jam in a one-way grid network

reduced. This causes a source of 'friction' at the upstream junction, as two traffic streams (ahead and turning) converge on the obstructed link and compete for the vacant road space there. Eventually, the two upstream links will fill to capacity, causing the transmission of blocking effects to a further four junctions. The pattern then repeats itself in a self-similar way, with each new queue contributing two additional branches to the evolving structure. This idea is shown diagrammatically in Figure 4.7.

4.3.3 Starvation processes

A secondary part of the development process of traffic jams is the phenomenon of 'starvation' - where vehicles are unable to reach parts of the traffic jam due to its own development. Starvation may affect the traffic jam's structure in a number of ways. We refer to the starvation which occurs in the internal regions of the traffic jam as *queue* starvation, and to that which affects the shape of the traffic jam envelope as *boundary* starvation. With queue starvation certain links become 'starved' of vehicles which are held back in upstream queues. Boundary starvation affects the shape of the jam boundary. Instead of following a regular structure, such as a polygonal shape, gaps appear along the periphery of the traffic jam to create a fractal structure. The relationship between queue and boundary starvation is explained in a later section.

Referring to Figure 4.6, one can observe that the boundary of the traffic jam is roughly diamond shaped, as predicted by the theoretical model. Starvation phenomena along the traffic jam boundary are apparent although not immediately obvious. However, for large traffic jams, their effects become more noticeable. Figure 4.8 shows a simulated traffic jam about the size of London. The picture was achieved by running the simulation model over a 256×256 one-way grid network with parameters typical of a Type I traffic jam. Each cross represents a blocked junction. Due to the process of starvation, the jam boundary is not entirely as predicted by the theory. It still exhibits the characteristic diamond shape but the jam is depleted at points



Figure 4.8: *A large simulated traffic jam on a 256×256 one-way grid network*

around the boundary, rather like a four-petalled flower. This will be discussed later in the chapter.

Queue starvation arises as a result of a reduction in the level of traffic feeding into particular links, because some of the traffic is held up elsewhere in the traffic jam. Thus, a proportion of the links inside the boundary of the traffic jam shown in Figure 4.6, become 'starved' of vehicles, and do not accumulate queues at all. Since the density of vehicles on these links is less than normal traffic, these queues have been termed *anti-queues*.

'Anti-queues' have also been observed by Scemama (1995), who describes these effects as unusually free-flow conditions which develop at previously congested sections. These conditions propagate in parallel with the queues. Since it is difficult to quantify the contribution made by 'anti-queues', the author suggests a spatial approach in analysing the extent and severity of any affected regions. This might be achieved using a graphical facility which would help operators visualise and hence analyse the effects of queue and anti-queue propagation.

Starvation has a third effect that is less immediately obvious than the two already mentioned: because vehicles are not able to leave the area enclosed by the jam boundary at the rate they would otherwise have done, outbound flows are depleted for some distance outside. The 'halo' of reduced traffic flows around the boundary is just perceptible in some of the computer visualisations.

The 'anti-queues' shown in Figure 4.6 propagate according to a regular geometric pattern which develops in parallel with the traffic jam itself. Consequently, it is possible to view the processes of growth and queue starvation as complementary ones. We consider the nature of this interaction next.

Sensitivity of growth pattern to underlying assumptions

The processes of queue and 'anti-queue' propagation are closely linked with the blocking-back mechanism. The assumptions underlying this mechanism can be summarised as follows:

In the model, the road link is divided into two distinct zones: a downstream queue storage area where vehicles are organised into separate turning movements, and an upstream 'reservoir' area where the turning movements are mixed. The interactions occur at the transition between these zones. The degree of interaction may be altered via an additional parameter, denoted by s , which assumes values in the range $0 \dots 1$. With $s = 0$, the drivers' discipline is assumed to be imperfect. Consequently, when a segregated queue spills back to the reservoir area, then no further movement is allowed from the reservoir area into either of the segregated regions.

By contrast, when $s = 1$, a queue in one segregated area does not affect the discharge of vehicles from the reservoir area into the alternative segregated area. This implies perfect lane discipline along the whole link: on entry vehicles immediately select the correct exit lane. Values of s between 0 and 1 can be selected to model the effects of partial interaction.

The effect of the degree of interaction between the queues on the pattern of jam growth has been analysed by conducting an experiment in which, all the model parameters remained fixed except for s (the degree of interaction) and p (the proportion of turning vehicles). The fixed parameters were $\alpha = 0.50$ (equal stopline widths allocated between ahead and turning queues), $\mu = 17$ (demand per entry point), $\kappa = 100$ (saturation capacity), the segregated portion of link held 20 vehicles and the corresponding reservoir area 40 vehicles.

The results are summarised in Table 4.1, which shows how the percentage of spillback caused by turning queues changes in response to variations in network conditions and blocking assumptions. For example, when $s = 0$ and $p = 0.2$, the percentage of blocked links triggered by the spillback of turning queues was thirty percent. In other words, the majority of spillbacks in the network (seventy percent) were caused by ahead spillback events. Thus, $\alpha = 0.50 < 1 - p$, so the traffic jam is predicted to be Type I in plan. This confirms the results of the theoretical analysis where, when $\alpha < 1 - p$, the result was a Type I traffic jam, whereas when $\alpha > 1 - p$, a Type II traffic jam formed. Similar conclusions may be drawn for the remaining entries of Table 4.1.

Note that when the lane discipline of drivers is imperfect ($s = 0$), an additional restriction is introduced into the model. This limits the proportion of spillback events triggered by turning queues, regardless of the overall turning proportion. This is apparent in the first row of Table 4.1, where even, for $p = 0.8$, where one would expect a Type II traffic jam ($\alpha > 1 - p$), the resultant traffic jam follows a Type I structure. This is due to the constraint $s = 0$ which ensures that the proportion of turn spillback does not exceed the corresponding proportion of ahead spillback. Thus, although the proportion of turning vehicles is much greater than the corresponding proportion of ahead vehicles, the proportion of turn spillback is limited by the additional constraint $s = 0$, and the ahead spillback movement is dominant. Invariably, $t_A < t_{LR}$ and the resultant traffic jam will be a Type I.

s	Turning				
	p=0.2	p=0.4	p=0.5	p=0.6	p=0.8
0.0	30 % (a)	36 % (a)	40 % (a)	42 % (a)	46 % (a)
0.4	28 % (a)	41 % (a)	40 % (a)	51 % (b)	72 % (c)
0.5	29 % (a)	38 % (a)	55 % (b)	54 % (b)	72 % (c)
1.0	31 % (a)	41 % (a)	55 % (b)	53 % (b)	73 % (c)

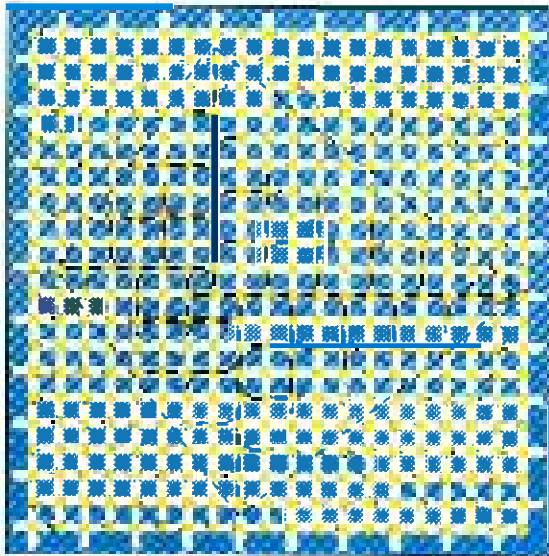
Table 4.1: *Proportion of turn spillback in a traffic jam, a: Diamond, b: Octagon, c: Square*

However, as the value of s increases beyond zero, additional discharge is allowed through the unblocked exit, enabling the proportion of turn spillbacks to increase proportionately with p . When the proportion of turn spillbacks exceeds the corresponding proportion of ahead spillbacks, the traffic jam assumes a Type II structure, whose boundary resembles an octagon. As this ratio increases still further, the octagonal shape changes to a 'square'.

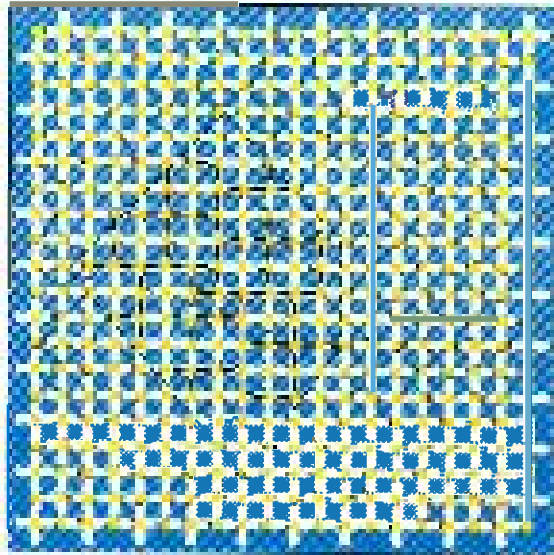
Unlike the theoretical analysis which predicted a single shape for each jam type, simulation has highlighted that there are various types of diamond and octagon shapes which characterise Type I and Type II traffic jams. A Type I traffic jam simulated with $s = 1$ and $p = 0.3$ can be seen in Figure 4.9A. The traffic jam is roughly diamond shaped. As the value of p increases, the diamond envelope changes and begins to resemble a hybrid structure of a triangle on one side, and a trapezoidal shape on the other, (see Figure 4.9B). This represents the threshold between Type I and Type II traffic jams.

As the proportion of turning increases beyond $p = 0.5$, then, providing $s > 0$, the queues begin to develop as small clusters of tightly packed 'curls' which become more tortuous as the value of p approaches unity. This characterises the development of Type II traffic jams, whose boundary broadly resembles an octagon. The octagon shape takes on various forms, an example of which is shown in Figure 4.9C.

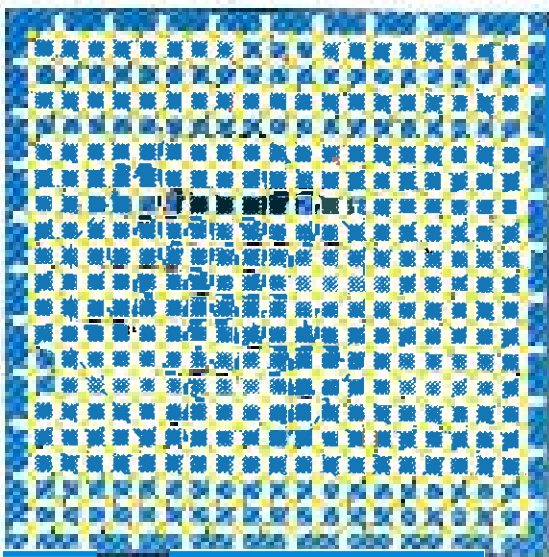
A: Diamond ($p = 0.3$)



B: Hybrid ($p = 0.5$)



C: Octagon ($p = 0.6$)



D: 'Square' ($p = 0.7$)

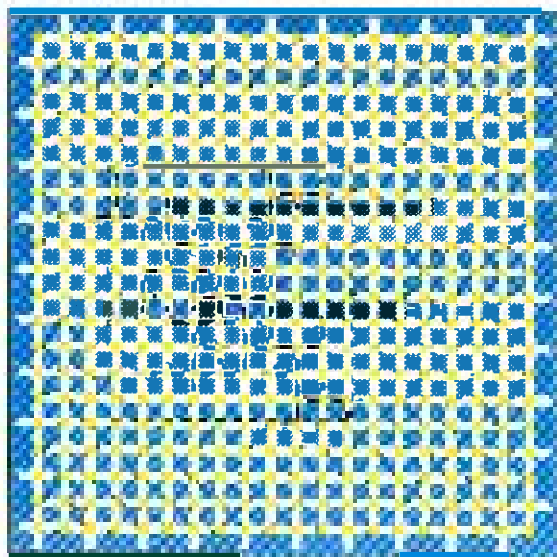


Figure 4.9: *The changing shape of the boundary of simulated traffic jams with $s > 0$*

When the proportion of turning vehicles approaches unity, the jam boundary is distorted into the shape shown in Figure 4.9D. If the tips of this structure are joined, then the resulting shape broadly resembles a square, although boundary starvation effects distort this shape quite substantially. Furthermore, the effects of 'anti-queues' have been eliminated completely, and the traffic jam evolves in tight gridlock 'cycles'.

This feature is typical of Type II traffic jams. These are dominated by turn spillback effects which propagate over the congested area. Increases in p contribute to a diminishing proportion of 'anti-queues', which lead to worsening boundary starvation effects. With Type I traffic jams the situation is reversed. Traffic jam development is dominated by ahead spillback effects, which cause extensive queue and 'anti-queue' propagation. Boundary starvation effects, on the other hand, are less noticeable.

We have already seen how Type I traffic jams incorporate a high proportion of 'anti-queues', which is not the case with Type II traffic jams. This has implications for dispersal. Later, we will see how a control strategy developed for Type I structures, might not be suited for clearing Type II structures. It would therefore be desirable if the extent of starvation in all its forms, could be quantified, so that traffic engineers could decide on the appropriate type of control. An exploratory technique for measuring the extent of starvation is proposed in the next section.

Fractal structure of traffic jams

Earlier in the thesis (Section 2.2.3) we considered the relationship between Diffusion Limited Aggregation (DLA), originally described by Witten and Sander (1981), and the process of queue propagation. A DLA is created by aggregating particles to the current structure which therefore evolves until the process is stopped or the structure reaches a particular spatial limit. The process of DLA generates fractal structures using a stochastic algorithm.

A DLA aggregate can be initiated by planting a seed particle at the centre of a lattice of sites. The growth process emanates from the seed, and over time, a cluster of

particles emerges. Particles are released from the edge of a pre-defined region, and are conserved. A particle can assume one of three possible states: fixed, transient or destroyed. Fixed particles cling to the evolving structure and cease to move. Transient particles continue to traverse the system until they become fixed or destroyed - in which case they are lost to the system. It is also possible to modify the rules which govern the growth mechanism of the aggregate. This enables a large variety of types of DLA to be generated.

The development of a traffic jam over a rectangular grid network is analagous to the DLA mechanism. Vehicles arrive at the network sources and leave via the sinks, whereupon they are destroyed. Like DLA, vehicles aggregate at an obstruction to traffic flow placed at a busy junction. However, there are also some important differences between the processes of DLA and traffic jam growth. First, the movement of vehicles in urban networks is not random walk - drivers leave specific origins to reach their planned destinations. Second, the growth process of a DLA aggregate is irreversible.

Like DLA, the resultant traffic jam structure possesses regular geometric features which are common to fractal objects. The growth pattern is replicated across several scales, although the 'scale' is limited by the block-size. For example, as shown in Figure 4.10, traffic jam development focusses on the central gridlock cycle (shown in black) which surrounds a city block. But secondary gridlock cycles may be observed as the traffic jam expands. These extend over a number of city blocks and are shown in blue (two blocks in each direction) and green (four blocks in each direction). In this context, the traffic jam is 'self-similar' in structure.

A secondary feature of the fractal structure is the gaps that appear both along and inside the outer boundary of the jam region, as shown in Figures 4.6 and 4.8. These irregularities lend a fractal characteristic to the developing traffic jam.

The small traffic jam image shown in Figure 4.6 covers a range of twenty blocks.

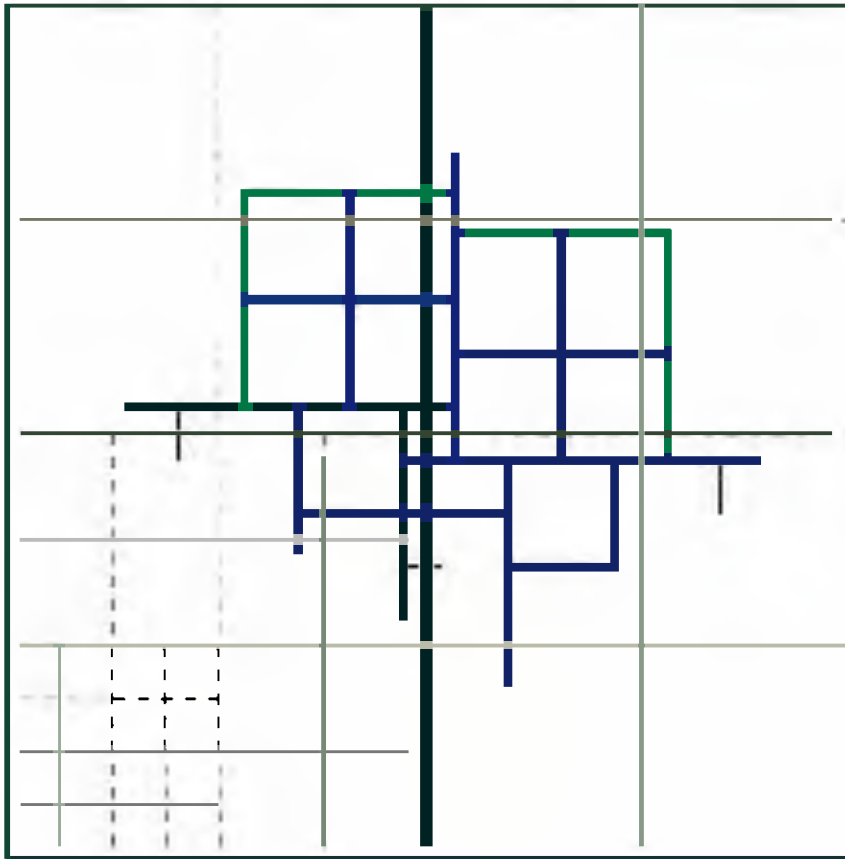


Figure 4.10: *Box-counting algorithm*

The heavily coloured regions indicate traffic congestion - gaps are clearly seen between the main jammed roads. *These gaps are known as 'anti-queues'*. Traffic flow is seen at other junctions, the clear gaps around them indicate that traffic is not being held up for any significant amount of time. The larger scale model, Figure 4.8, shows only congested areas and has a structure resembling a fern growth.

It is possible to measure the extent of the gaps in the jam structure by estimating the fractal dimension of the traffic jam. In general, the fractal dimension of any structure is obtained by plotting the logarithm of a particular measure against the length of the measuring scale. If this plot is linear, with negative gradient D , then $1 - D$ is an estimate of the structure's dimension. Mandelbrot (1983).

To estimate the fractal dimension of a traffic jam, we define a square region of side L

which can be covered by closed square boxes of side length 2^i . This is superimposed on the grid network as shown in Figure 4.10 to eliminate any possible ambiguities which may arise in the counting procedure.

Let $N(r)$ denote the number of boxes of side length $r = 2^i$ which intersect the traffic jam image. A plot of $\ln(N(r))$ vs $\ln(1/r)$ is then obtained. Providing the resultant graph is approximately linear with gradient D , then D can be interpreted as the fractal dimension of the image since

$$\ln(N(r)) = D * \ln(1/r) + \ln(P)$$

where P is a constant. Hence

$$N(r) * r^D = P$$

The full algorithm for this 'box-counting' procedure is provided in Appendix C. A similar technique has also been successfully applied by Morse et al (1985) to measure the fractal dimension of vegetation.

Figure 4.11. shows the relationship between the number of intersections and the corresponding scale of measurement. Each of the graphs appears as a negatively sloped line. This demonstrates the fractal nature of the image.

The fractal dimensions, denoted by D_n , for the traffic jams grown over grid networks of order $n \times n$, are shown in Table 4.2. They have been estimated by measuring the respective line slopes shown in Figure 4.11. The results are similar to those obtained by Meakin (1983), for a conventional DLA process. Note that the fractal dimension, D_n , may assume values in the range $1 \dots 2$. Mandelbrot (1983). Thus, if $D_n \rightarrow 1$, the traffic jam will resemble a linear structure, but as D_n approaches 2, then the image will tend to fill the 2-D space with few gaps in the structure. In other words, D_n measures the degree to which the traffic jam image fills a two-dimensional space.

The increasing values of D_n suggest that the traffic jam assumes a more dense structure as the size of the network increases. This effect can be observed by comparing

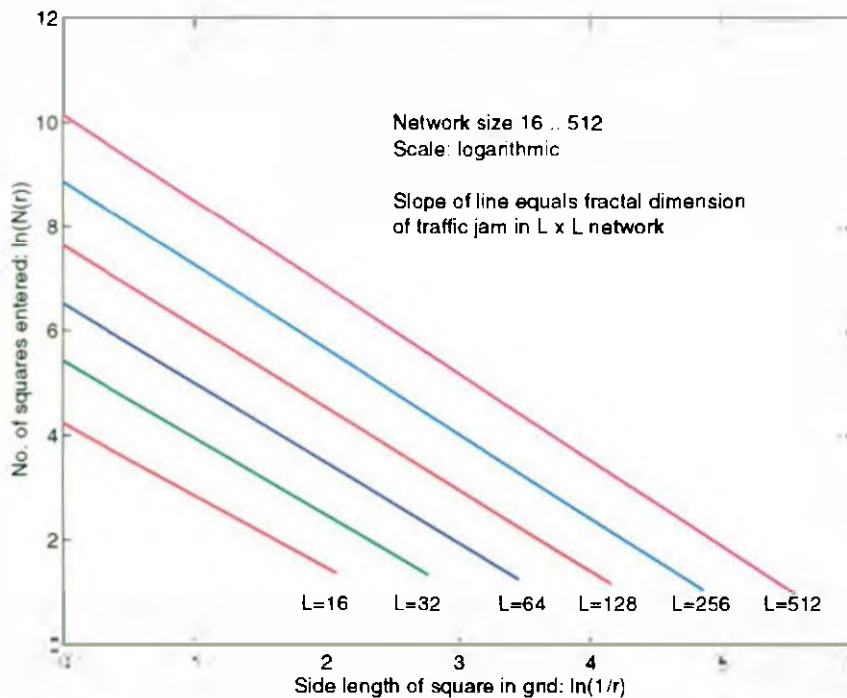


Figure 4.11: *Logarithmic plots: negative slope of line equals fractal dimension*

the shapes of the traffic jams shown in Figure 4.8 and Figure 4.6 respectively. In 4.6 the boundary of the traffic jam can be viewed as approximately step-like, and to a degree almost linear, whereas with Figure 4.8 the intricate shape of the boundary is reflected in a fractal dimension closer to 2 than 1.

It would be desirable to extend this technique to measure the fractal dimension of Type II traffic jams so that comparisons could be made between the two types. But this aspect was not pursued here because although the method provides a way of describing the structure of a traffic jam, it is a theoretical measure which in its current form, cannot be applied in a practical sense. The objectives of this research were of a practical nature and further development of the analysis could not be justified under the constraints of the research programme.

Having considered the structure of traffic jams in detail, we now turn to the problem of their dispersal.

Network Size: n	Fractal Dimension: D_n
16	$1.37 \pm \epsilon$
32	$1.48 \pm \epsilon$
64	$1.52 \pm \epsilon$
128	$1.55 \pm \epsilon$
256	$1.61 \pm \epsilon$
512	$1.65 \pm \epsilon$

Table 4.2: *Fractal dimension of traffic jam: $0.00 < \epsilon \leq 0.03$*

4.3.4 Dispersability

Stages in the dispersal process

The dispersal process of a traffic jam can be initiated by removing the original cause of the jam, or by introducing external counter-measures. The rate of dispersal will be affected, amongst others, by the network conditions, the location and severity of the original incident, and the time taken till control policies are implemented.

It is convenient to classify the transition from the growth phase to the dispersal phase into three categories. Where the traffic jam has been active for a short time, the removal of the original incident terminates the growth process, and the queues disperse at varying rates. The transition from the development state to the dispersal state is immediate, and the traffic jam is eliminated within a short time period.

The second category arises when the removal of the obstruction leads to an oscillatory phase in which queue lengths fluctuate throughout the congested region. Traffic jam growth accelerates and decelerates at regular intervals. Eventually, the queues begin either to dissolve or grow. The boundary between growth and decay is not clearly defined, and dispersal *without external forms of intervention* cannot always be reliably predicted.

In the third group, the elimination of the incident does not induce the traffic jam

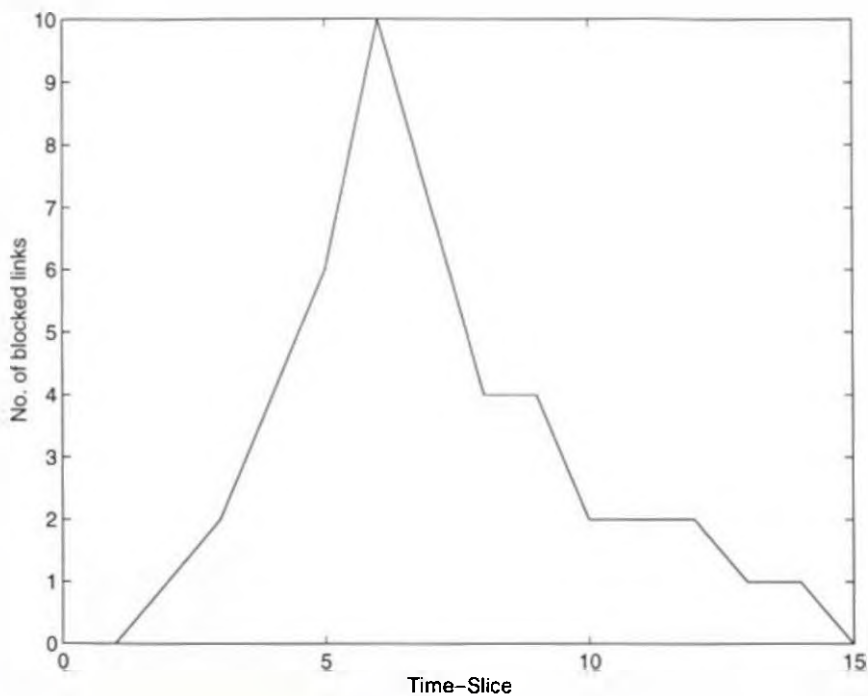


Figure 4.12: *Dispersal mechanism 1: traffic jam decays*

to disperse and the traffic jam continues to expand. In some cases, the removal of the incident leads to a temporary reduction in jam size, which is followed almost immediately by renewed growth. In this case, after a given time interval, the traffic jam fails to clear, and more extensive intervention is required in order to induce the dispersal state.

Figures 4.12, 4.13 and 4.14 demonstrate these three categories graphically. The graphs represent an example of a series of simulation experiments in which a traffic jam was grown for a fixed number of time-slices. This was followed by the removal of the obstruction. The timing of the removal varied and appeared to have a critical impact on the dispersal outcome.

Figure 4.12 exhibits the acceleration of growth caused by the obstruction. The elimination of the incident in the sixth cycle of the simulation triggers the dispersal process, and the traffic jam clears within a short period of time. The graph in Figure 4.13 exhibits the oscillatory phase which might arise in dispersing traffic jams. The

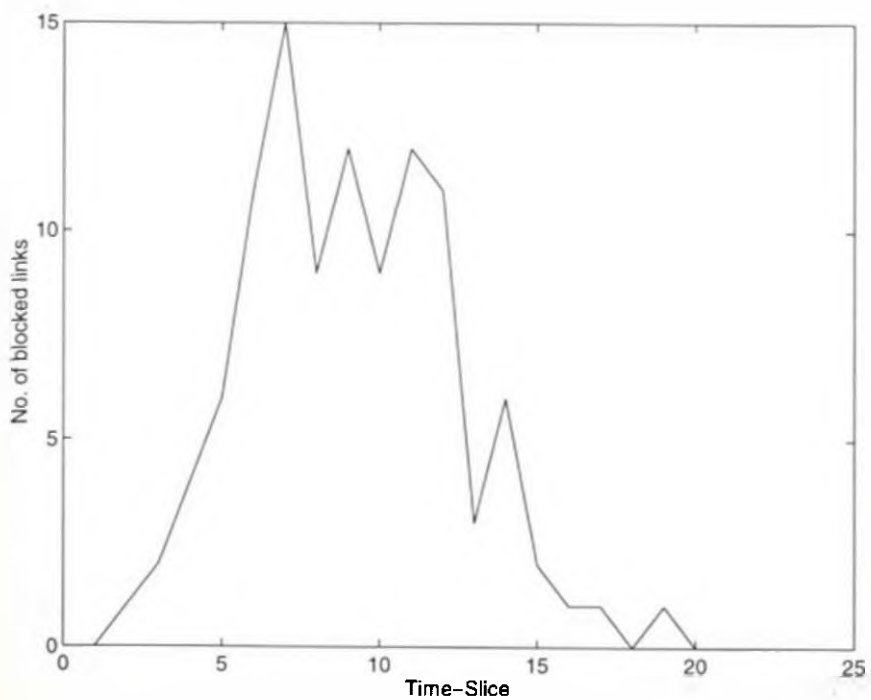


Figure 4.13: *Dispersal mechanism II: queue lengths oscillate*

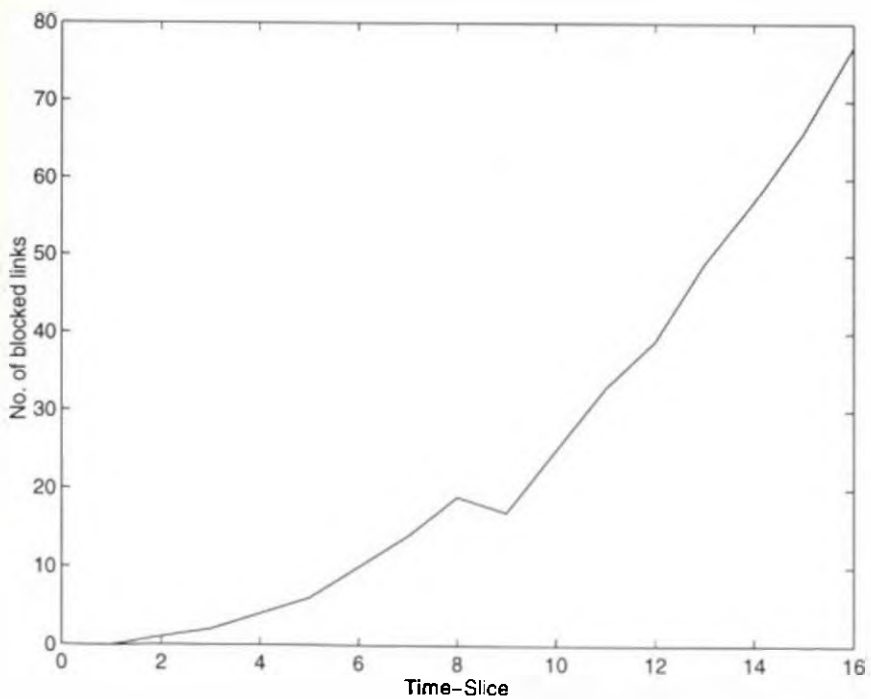


Figure 4.14: *Dispersal Mechanism III: traffic jam fails to clear*

incident is eliminated during the seventh cycle of the simulation. The dispersal pattern follows a downward trend, although fluctuations occur from one cycle to the next. The third graph, shown in Figure 4.14, highlights the build-up of the traffic jam until the extraction of the incident, which occurs in the eighth simulation cycle. Thereafter follows a re-arrangement of the vehicles within the queues, which sometimes causes a temporary reduction in the jam size. The slight reduction in jam size is due to the additional space which is created as a result of removing the obstruction. This enables the release of a small number of vehicles from the queue immediately upstream of the obstruction. This movement produces a ripple effect throughout the traffic jam but it is only temporary and the traffic jam soon begins to expand as before.

The above experiment was repeated using different seeds which were generated at random at the start of the simulation. The results indicate similar phases in the dispersal mechanism (decay, oscillation and growth). The results are presented in Appendix B.

The importance of early attention of the cause of the traffic jam is demonstrated in this simulation experiment. Postponing the removal of the obstruction by a single cycle sometimes results in the traffic jam's failure to clear.

The onset of gridlock

Spatial analysis of congestion patterns has highlighted that the formation of a gridlock cycle represents an important threshold within the evolution of the traffic jam. Once it has occurred it becomes more difficult to clear the resultant traffic jam. This effect can be observed in Figures 4.12, 4.13 and 4.14. In the first case, the gridlock cycle had formed on three of the four possible sides, and total dispersal was achieved. In the second case, the gridlock cycle had extended into the fourth side. However the cycle had not 'closed' completely, ie the fourth queue had not arrived at the starting point of the traffic jam. This occurred in the third case. With the onset of gridlock,

it was no longer possible to clear the traffic jam by removing the obstruction and more extensive forms of intervention were required. These included the application of vehicle-bans at specific network locations. These are described in more detail in Chapter 5.

4.3.5 Implications for control

There are two approaches for controlling traffic queues. The first approach considers how features in the road layout, such as the allocation of queue storage space between ahead and turning vehicles, can be exploited in order to curb the rate of traffic jam expansion and to delay the onset of gridlock phenomena. Measures such as these are static because once set, they can no longer be changed. The second approach suggests responsive measures which can alleviate congested situations where gridlock has already set in. These are known as dynamic measures. This section considers a selection of traffic management measures that can be implemented as part of a static control scheme. Dynamic control measures will be dealt with separately in Chapter 5.

Static control measures can be classified in two groups: measures which affect the rate of queue propagation and measures which affect the formation of the gridlock cycle. Traffic engineers would aim to reduce the rate of queue propagation whilst postponing the onset of gridlock for as long as possible. We use the simulation tool to point to various ways of how these objectives might be achieved in practice.

Before describing the experiments, we provide some justification for the ranges of the key parameter values over which the experiments have been conducted. These values include the proportion of turning vehicles (p), the level of demand (μ), the allocation of stopline widths (α), and the degree of interference between the segregated and reservoir areas of a link (s).

In general, the proportion of turning vehicles in a road network is relatively low and

does not regularly exceed thirty percent. Unless otherwise specified, the value of p in the experiments is assumed not to exceed this limit. Similarly, within the constraints of the saturation capacities, the level of demand has varied from $\mu = 16$ (low) to $\mu = 28$ (high) with values in between representing average levels.

The allocation of stopline widths has been controlled using the α parameter which in principle, could assume values from the continuous range $0 \dots 1$. Nevertheless, in the context of these experiments, α was limited to a discrete set of choices $\alpha \in \{0.33, 0.50, 0.66, 0.75, 0.80\}$. These were thought to reflect some of the road layouts found in practice (for example $\alpha = 0.66$ would allocate two lanes for ahead traffic to each lane of turning traffic). A similar stance was taken with the actual range assumed by the s parameter. Three values were selected from the range $0 \dots 1$. These were $s = 0$ (worst case), $s = 1$ (best case) and $s = 0.5$ (average case).

Minimising rate of traffic jam growth

The analytic model of Wright and Roberg (1997) predicted that, under idealised conditions, 'the rate of traffic jam growth was affected among other things by the severity of the original blockage, the overall level of traffic demand, the proportion of link storage area devoted to the segregated queues and the balance of storage areas between ahead and turning queues. In addition, it was postulated that for Type I traffic jams it would be desirable to aim for a layout in which the widths allocated to the segregated approach queues would be in exactly the same ratio as the demands ($\alpha = 1 - p$) together with minimal length of segregated queue storage (small σ). See Appendix C for the derivation of this result.

A series of experiments were conducted using the simulation tool, where for each run, the size of the traffic jam (measured in terms of the total number of blocked links) was recorded after a fixed period. The fixed simulation parameters were as follows: the storage capacity of a link was 50, the demand, μ , was 21 (veh per min per entry point), the saturation capacity, κ , was 50 (veh per min), the proportion

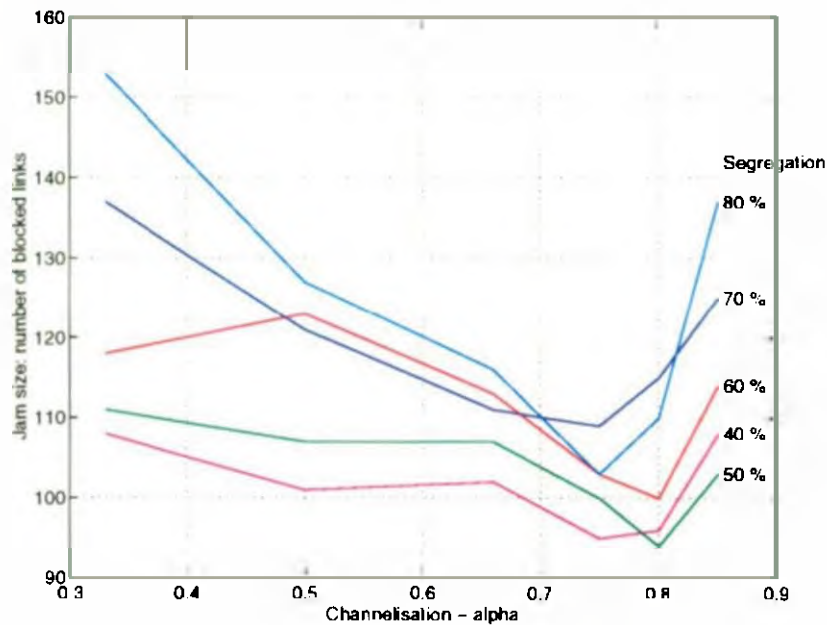


Figure 4.15: Effect of segregated queue storage on jam size for various channelisation regimes

of turning vehicles, p , was set at 0.2 and imperfect driver discipline ($s = 0$) was assumed throughout. The parameters that varied from trial to trial were the extent of channelisation α and the degree of segregation (σ). α was chosen from a discrete parameter set ($\alpha \in \{0.33, 0.5, 0.66, 0.75, 0.80, 0.85\}$) and σ varied from 0.4...0.8.

Figure 4.15 shows the effect of segregated queue storage capacity on the jam size for various channelisation regimes. The results indicate that keeping all other simulation parameters fixed, the rate of traffic jam growth was slowed when the segregated region was small. Thus, given a particular set of demands, turning proportions, road link storage capacities, saturation flows and stopline width allocations, the rate of traffic jam growth could be curbed by dedicating a small portion of the road to the segregated region.

In addition, the rate of jam growth follows a downward trend for increasing values of α , but only up to a point. As α increases beyond 0.80 spillback events become more frequent and the rate of traffic jam expansion increases. Thus, the rate of traffic jam

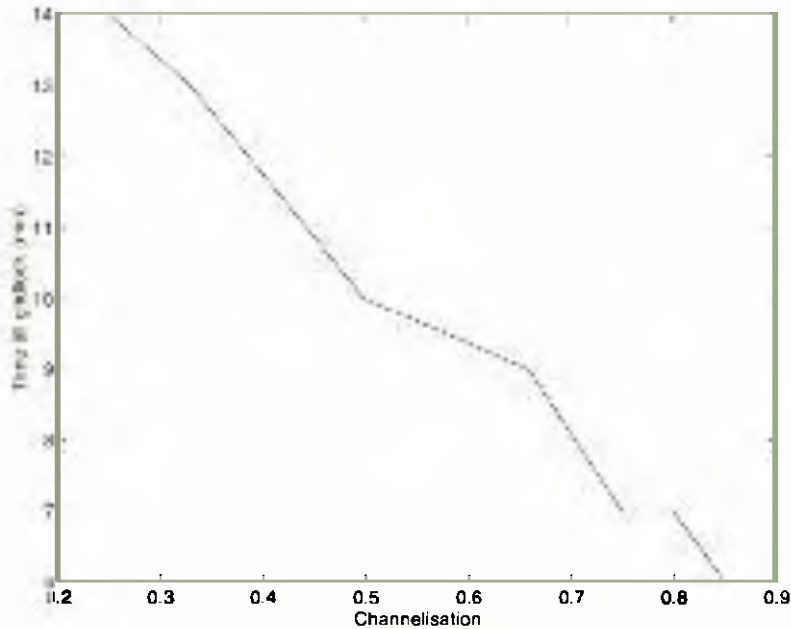


Figure 4.16: *Effect of channelisation on time till gridlock: $\sigma = 0.4$*

growth is minimised when $\alpha = 0.80$, in other words, when the amount of stopline width devoted to the *ahead* segregated approach is in the same proportion as the demand using it, ie, for small σ and when $\alpha = 0.8 = 1 - p$.

Inhibiting gridlock

The theory also pointed to possible measures for inhibiting gridlock. These included allocating a *larger* proportion of stopline width to the turning traffic. The reasoning for this is provided in Appendix C.

Another series of simulation trials were conducted using the simulation tool to test this hypothesis. In the experiments, an obstruction was introduced at the start of the simulation. The number of time-slices until the onset of gridlock was recorded for various values of α . The remaining model parameters were the same as in the previous experiment except that σ was fixed at 0.4 throughout.

The results show that for high values of α , gridlock occurs almost immediately. However, it is possible to delay the onset of gridlock by reducing α , ie by allocating a

larger proportion to the turning traffic. Thus, there is a direct conflict between the requirement to maximise gridlock time, and the requirement to minimise the overall growth rate of the traffic jam. Figure 4.16 shows the effect of channelisation on the time till gridlock for the particular parameter values listed above: however, a similar pattern was observed for other demand levels as well.

Queue interaction

One of the assumptions underlying this model is the effect of spillback of one queue on its neighbour. Unless specified otherwise, it is assumed that when a queue spills back to the reservoir area, movement is suspended into both segregated areas. In real networks, this assumption might be too crude, since skillful drivers often manouver themselves around the tail of a queue. Consequently, we next examine the sensitivity of the model to this assumption insofar as the rate of traffic jam growth and the time till gridlock are concerned.

In a limited experiment, traffic jams were simulated under identical conditions for a variety of turning proportions, $p \in \{0.2 \dots 0.7\}$. The size of the traffic jam was recorded after 25 time-slices for each level of turning. The fixed parameters were the link capacity (60 vehicles), the saturation capacity $\kappa = 100$, the level of demand $\mu = 17$ (veh per min per entry point), and $\alpha = 0.50$ (equal allocation of stopline widths). The size of the traffic jam was measured for various values of s - the queue interaction parameter - typically $s = 0, 0.5, 1$.

Figure 4.17 describes the size of the traffic jam over time for $p = 0.3$. The results show how the rate of traffic jam growth may be reduced, but not apparently by very much by increasing the value of s beyond zero. The results here seem to be relatively insensitive to queue interaction. A similar trend was observed for different proportions of turning. The results are summarised in Table 4.3, where, for each value of p , the rate of traffic jam growth is slowest when $s = 1$ and highest when $s = 0$. This can be understood intuitively. With $s = 1$, vehicles that would be restricted under $s = 0$,

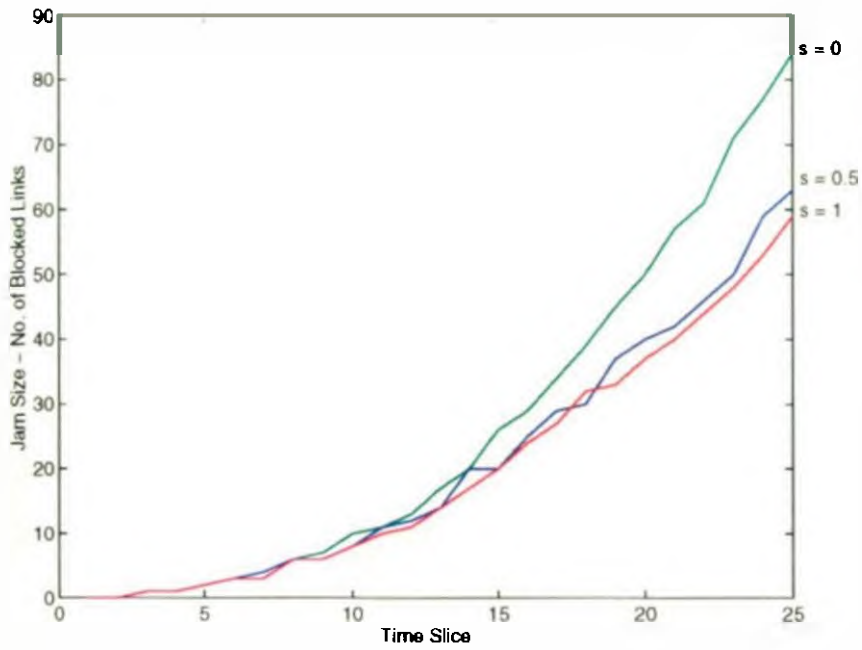


Figure 4.17: Rate of growth of traffic jam for various s

Turning p	Jam Size		
	$s=0$	$s=0.5$	$s=1$
0.2	109	74	73
0.3	103	91	81
0.4	93	68	65
0.5	82	61	52
0.6	73	46	44
0.7	41	25	25

Table 4.3: Extent of traffic jam (No. of blocked links at $\tau = \tau_{10}$) with respect to s

Turning p	t_g		
	s=0	s=0.5	s=1
0.2	17	17	13
0.3	12	12	10
0.4	14	11	10
0.5	7	7	7
0.6	7	7	7

Table 4.4: *Time to gridlock (minutes) with respect to s*

may proceed en route to their respective destinations. Consequently, less vehicles become trapped by congestion, thereby ensuring a slower rate of jam expansion. Traffic delays can therefore be reduced if drivers position themselves according to the correct exit lane.

However, with the criterion of maximising the time till gridlock the situation is reversed. Table 4.4 shows how, for various p , the time till gridlock - measured in time slices and denoted by t_g) - decreases as s increases. The results indicate that it is not always beneficial to increase the value of s beyond zero when the overall aim is to delay the onset of gridlock. Furthermore, the likelihood of gridlock formation increases proportionately with increases in the percentage of turning. Gridlock is also triggered more rapidly with higher turning proportions - seventeen time-slices are required for $p = 0.2$ whereas with $p = 0.6$, gridlock has formed within seven time-slices. Similar conclusions have been drawn in the study by Shepherd (1994).

It follows that the onset of gridlock and traffic jam expansion as a whole are to some extent independent: they may be influenced in different ways by the various network parameters. It may therefore be appropriate to view traffic delays and gridlock as two separate characteristics: strategies that yield improvements in one respect may not yield a similar improvement in the other. In particular, there are circumstances where queueing restrictions reduce the rate of growth of the jam, but can actually induce

a premature onset of gridlock conditions, and hence make dispersal more difficult. Similarly, there are circumstances which may postpone the onset of gridlock, but increase the rate of traffic jam growth, and hence the overall rate of delay. The control of traffic congestion is complicated by these conflicting aims, both of which need to be addressed in tandem.

It follows that in developing strategies for controlling traffic congestion in real networks, the solution will depend on the overall objectives. The comparative importance of these objectives will dictate the type of strategy to be implemented and some element of compromise will need to be sought. Consequently, networks will need to become more adaptable to forms of control since these will vary according to the prevailing conditions and requirements.

4.4 Traffic jams on two-way grid networks

4.4.1 Growth pattern

The development of a traffic jam on a two-way network resembles the process described for the one-way network. When an obstruction is created towards the centre of the network, initially there will be one or more queues which are effectively independent. Subsequently, these will branch out to cause interference with upstream links and eventually, the queues will encircle the cause of the obstruction leading to gridlock.

But while there are similarities in the stages leading to gridlock there are also a number of differences. In the one-way system, vehicles travelling along a link may choose from two possible movements at the downstream junction, ahead and turning. The turning movements alternate between left and right turns at successive junctions. In addition, at any given junction, vehicles may arrive via either of *two* approaches: ahead and one of left or right. It follows that when a link fills with queueing traffic, its capacity to accept vehicles will be reduced and blocking effects will be transmitted

to the two upstream links feeding into this blocked link. At the upstream junction, if the left-turning movement is blocked, then the gridlock cycle will evolve in an anti-clockwise manner, whereas if the right-turning movement is blocked, the cycle will form in a clockwise direction. (The circular analogy has been used to explain the sequence of links which block as a result of the obstruction. Each pattern produces a slightly different gridlock cycle.) Thus, traffic jam development on a one-way grid network emanates from a single loop of locked vehicles. These can be left-turning or right-turning vehicles, but not both.

Let us now consider the process of jam development in the two-way system. Here, vehicles travelling along a link may choose from three possible movements at the downstream junction: ahead, left and right. Conversely, at any given junction, vehicles may arrive via either of *three* approaches: ahead, left and right. When a link fills with queueing traffic, blocking effects will propagate to the three upstream links which feed into this blocked link. Consequently, two gridlock cycles will develop in parallel: one of left-turning vehicles, (known as a left-turning gridlock cycle), the other of right-turning vehicles, (known as a right-turning gridlock cycle). The network becomes 'gridlocked' as soon as either of these cycles has formed.

The pattern of development is dominated by a number of primary queues which subsequently branch out into secondary queues. Eventually, these extend to form larger centres of interlocking queues. The overall shape of a simulated two-way traffic jam is shown in Figure 4.18. The obstruction has been placed at the junction highlighted in purple. The gridlock cycles at the centre of the traffic jam have been highlighted in red and green respectively.

The boundary of a two-way traffic jam broadly follows the shape of a one-way traffic jam. If the tips of the two-way traffic jam shown in Figure 4.18 are joined, then the boundary exhibits the characteristic diamond shape. However, this shape is not as regular as the one-way counterpart, and there are large areas within the boundary which do not accumulate vehicles at all. Indeed, the fourth quadrant of the two-way

traffic jam appears blank in the diagram since this quadrant is the last to be affected by the jam and it takes a considerable amount of time until it fills. In addition, the overall structure of the traffic jam is more dense.

We have also simulated two-way traffic jams in which the overall turning proportion has varied whilst maintaining an equal ratio between left and right turns. This was to enable us to compare the results with those of the analytic one-way model. Later we describe the effect of unequal ratios between left and right turning traffic. The shapes correspond roughly to the one-way traffic jams simulated with $s = 1$ under similar conditions. This is not surprising since the spillback definitions rely on equivalent assumptions. Thus, in this case, it appears that the Type I (diamond shape) and Type II (octagonal and square) classification for one-way traffic jams may be extended to the case of two-way traffic jams. Traffic jams affected by ahead spillback effects will be Type I in structure, whereas those dominated by turn spillback effects will be Type II. Furthermore, it might be possible to modify dispersal strategies developed for one-way traffic jams and apply them in a similar way to tackle two-way traffic jams. We will consider the implications of this in Chapter 5.

Effect of turning proportion on jam structure

The spatial characteristics of a two-way traffic jam will be affected by network conditions in much the same way as the one-way counterpart. For example, increasing levels of demand will contribute to a faster rate of jam expansion in both cases. However, an important difference between the two types of network is the contribution of right-turning movements to the overall characteristics of the resultant traffic jam. This will be examined in two parts. In this section, we will describe the effect of the balance between left and right turns on the traffic jam shape. The contribution of the right-turn storage areas (known as right-turn bins) on the processes of growth and decay will be considered separately.

Assuming that the storage space allocated between left-turning and right-turning

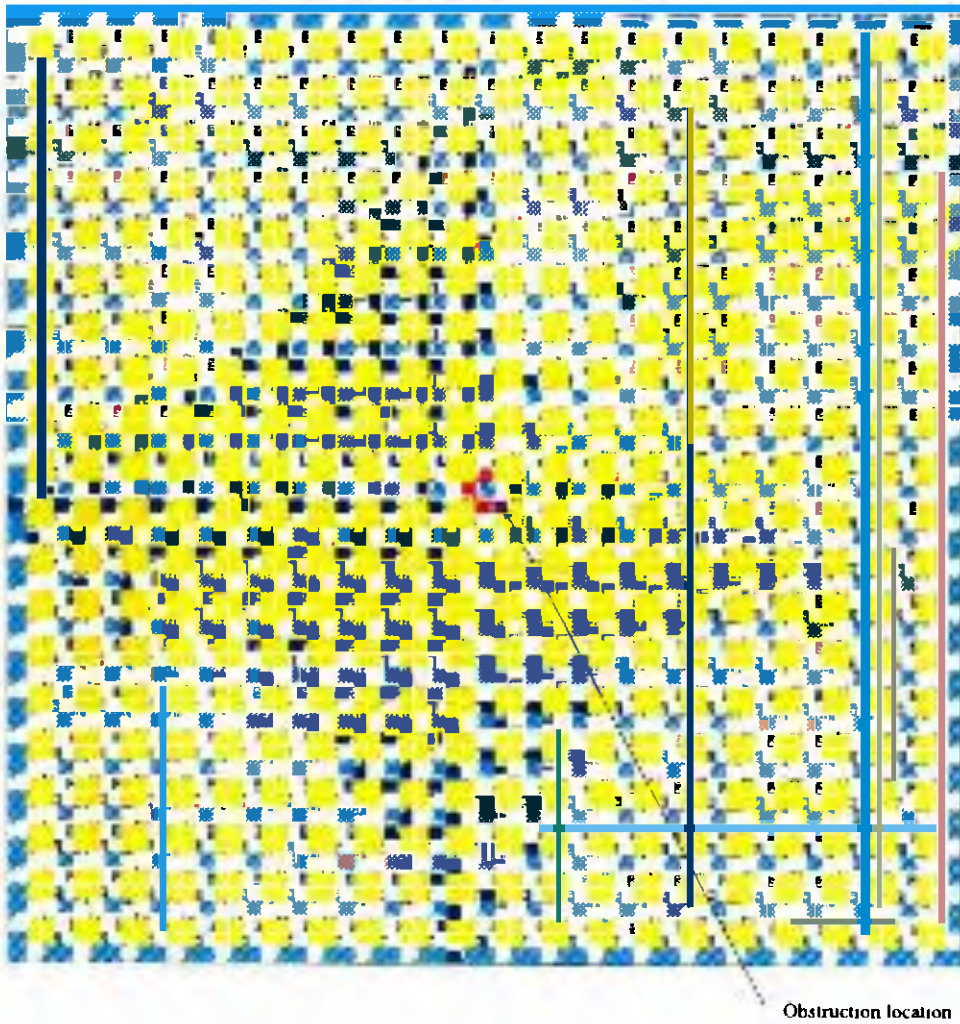


Figure 4.18: *Spatial characteristics of two-way traffic jam*

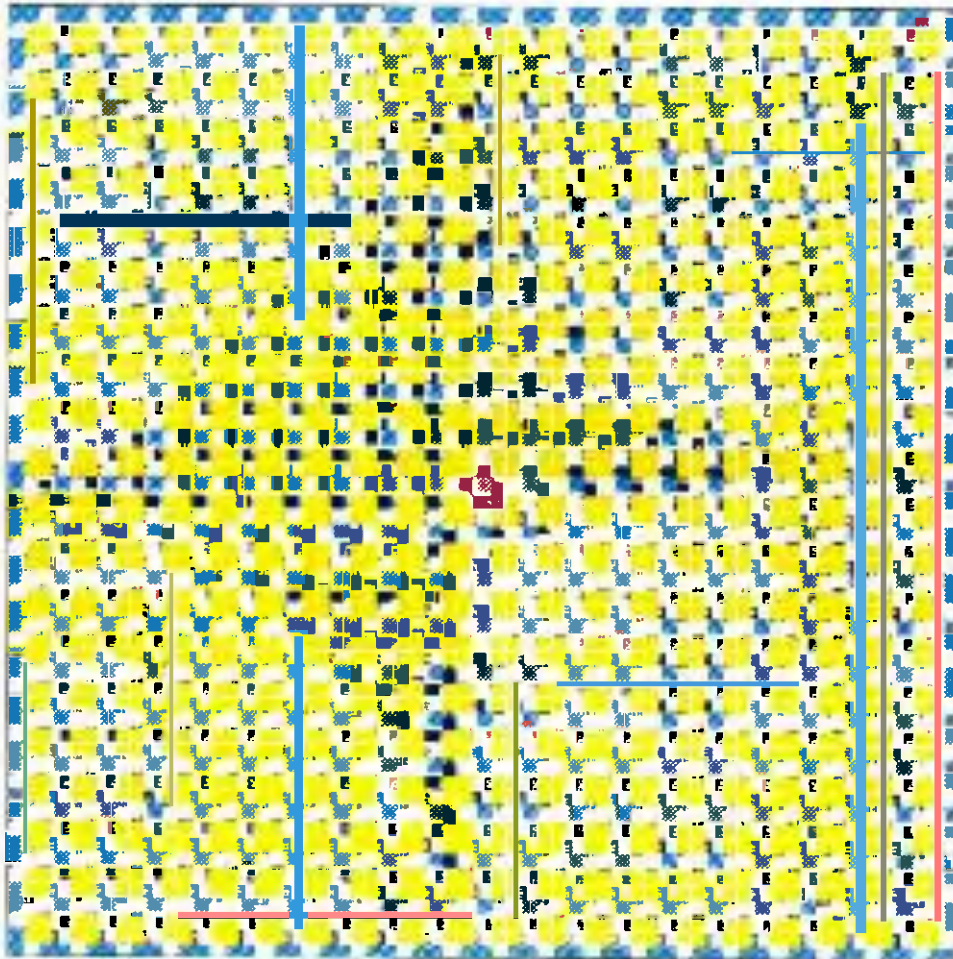


Figure 4.19: *Spatial characteristics: left-turning cycle*

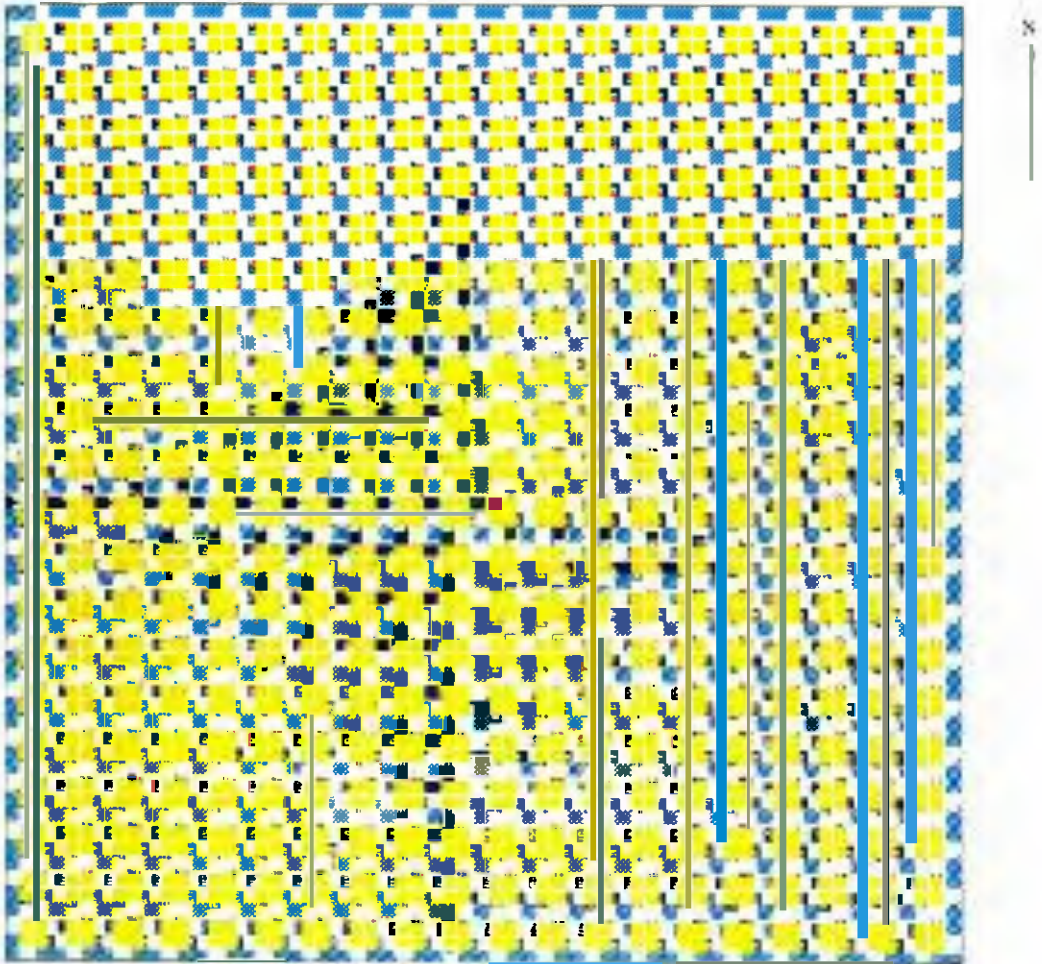


Figure 4.20: *Spatial characteristics: right-turning cycle*

queues is approximately the same, there are a number of possibilities to consider:

1. proportion of left-turning vehicles is greater than the corresponding right-turning proportion.
2. proportion of right-turning vehicles exceeds corresponding left-turning proportion.
3. left and right-turning proportions identical

It follows that with condition (1), providing the assumption regarding storage space holds, the 'left-turning' gridlock cycle will form more rapidly than the right-turning one. This can be seen in Figure 4.19, in which the 'left-turning' gridlock cycle has been highlighted in red. Furthermore, there is one quadrant which hardly accumulates any vehicles. Since the traffic jam evolves in a clockwise direction, this quadrant is affected last.

However, with condition (2) the situation is reversed: the proportion of right-turning vehicles exceeds the corresponding proportion of left-turning vehicles, and the traffic jam evolves in an anti-clockwise direction. Figure 4.20 shows the resulting traffic jam. In this case, the right-turning traffic dominates and as a result, the 'right-turning' gridlock cycle forms first. It has been highlighted in green.

Condition (3) describes the situation in which the ratio between left and right turning vehicles is identical. If this were the case, then the left-turning and right-turning gridlock cycles would evolve at equal rates and even lock at the same point in time. This would imply a double gridlock, and jam dispersal in such a situation might be more difficult. However, in practice, this situation would almost never occur. Consequently, it is only of theoretical interest.

4.4.2 Effect of right-turn bin size on jam growth and decay

The structure of the two-way network resembles the one-way counterpart except that in addition to the ordinary links which comprise the network, there are also areas

(referred to as right-turn bins) which limit the throughput of right-turning traffic at each intersection. These areas introduce a degree of irregularity to the grid structure because their storage capacity is much smaller than conventional links. This is not the case with the one-way system whose links all have the same storage capacity. At the end of each simulation cycle, the vehicles waiting in the right-turn storage bins progress towards their destination link subject to space being available downstream.

Note that the spillback assumptions for two-way links (Section 3.2.4) have not been modelled for the right-turn bins. In other words, the maximum number of vehicles that can wait in the right-turn bin area is limited to a small yet fixed number. These vehicles can form into a single line of queueing traffic but no queueing interactions are modelled within these areas. This means that the model only represents the 'best-case' scenario as there may be additional complications (such as blocking effects in the right-turn bin areas) which may cause the system to lock up faster than predicted.

The rate of traffic jam expansion does not appear to be affected by the size of the right-turn bin area. In other words, increasing the size of the area does not slow the rate of jam growth in any way. However, the rate of traffic jam *dispersal* is sensitive to the size of this area although only up to a point: given the network conditions it can be beneficial to increase the size of the bin area. This results in a faster dispersal rate and shorter delays. But, increasing the size of the area without bound does not help, particularly when the right-turn saturation flow is less than the storage capacity of the right-turn bin area. Consequently, it is advisable to set the size of this area close to the value of the right-turn saturation flow.

Figure 4.21 shows the contribution of the size of the right-turn bin on the rate of jam dispersal graphically. The total delay to drivers is smallest when the right-turn bin size is fixed at eight vehicles. (This value is close to the right-turn saturation flow). The delay increases slightly if the right-turn bin size is reduced to six vehicles. If the bin size is reduced further, to five vehicles, then the traffic jam fails to clear and the total delay to drivers increases without bound. The reason for this is that

the size of the storage bins imposes an additional limit on the throughput of vehicles from the right-turning segregated queue. If too small (as shown in this example), then the queue feeding into this area fails to clear and triggers spillback events which eventually propagate throughout the congested region.

The results in Figure 4.21 were achieved by running the simulation with the following fixed parameters: $\mu = 20$, $\alpha = 0.50$, $\kappa = 100$ and the link storage capacity was 60 vehicles, with $\sigma = 0.33$. The proportion of turning vehicles was $p = 0.3$ of which 0.14 were left-turning and the remainder right-turning. The traffic jam was grown over a fixed period of fifteen time-slices, following which the obstruction was removed.

Thus, while the rate of traffic jam expansion is not affected by the capacity of the right-turn storage bin, the size may be critical where traffic jam dispersal is concerned. This trend was observed for other parameters and network conditions.

However, whilst the rate of dispersal appears to be faster if the size of the right-turn bin area is increased, a degree of caution is required. This is because the effects of queueing within the right-turn bin areas have not been modelled directly and so, by allocating too much storage space, one might actually promote blocking effects and gridlock within the intersection itself. This might cause the system to lock up faster than expected and make dispersal much more difficult.

This analysis suggests that junction layout may play an important role where traffic jam dispersal is concerned and that inappropriate storage allocation may generate situations which are out of proportion to the original incident. In addition, whilst a particular road layout might be suitable when the overall objective is to reduce the spread of congestion, additional space might need to be provided when the system has broken down in order to aid dispersal. This should be considered in the design stages.

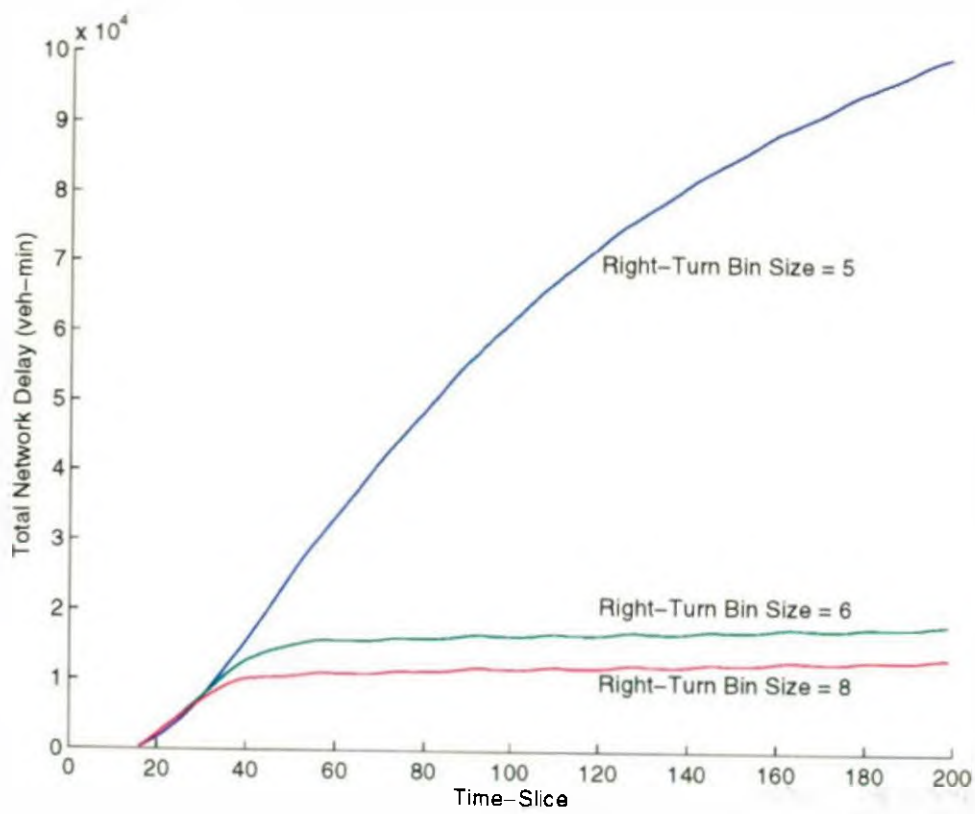


Figure 1.21: *Effect of right-turn bin size on rate of traffic jam decay*

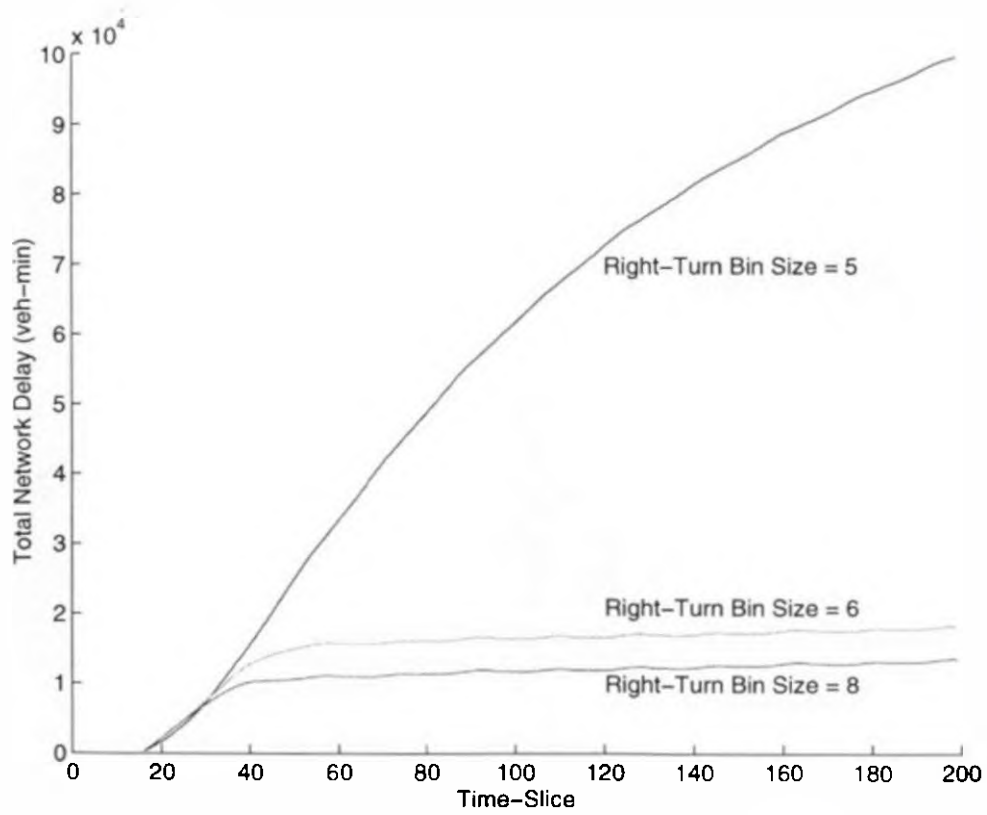


Figure 4.21: *Effect of right-turn bin size on rate of traffic jam decay*

4.5 Conclusions

This chapter considered the fundamental attributes of traffic jams in one-way and two-way rectangular grid networks, and developed a framework for their classification. Three stages characterised the behaviour of traffic jam phenomena. These were the development phase, the dispersal phase, as well as a starvation mechanism which interacted with the primary stages of growth and decay.

The growth phase was initiated by creating an obstruction towards the centre of the network. This led to one or more queues which were effectively independent. Subsequently, these branched out to cause interference with upstream links, and eventually, the queues encircled the cause of the obstruction leading to gridlock.

The overall pattern of traffic jam development was the same for one-way and two-way grid networks. But some differences were observed, for example, traffic jam development on a one-way network emanated from a single loop of locked vehicles, whereas with the two-way network, the development focussed on two loops which formed in parallel. The problem of gridlock in a two-way network was potentially more severe than in the one-way network since there were twice as many gridlock loops to deal with. In addition, whilst the size of the right-turn bin areas did not affect the rate of traffic jam growth, their size was critical where traffic jam dispersal was concerned.

The boundary of a traffic jam in a one-way grid network broadly resembled a polygonal shape. Depending on the network conditions, this was characterised as either diamond shaped (Type I), octagonal or square shaped (Type II) in plan. Similar shapes were observed for the two-way network.

Simulation experiments were conducted to assess the role of network conditions on the structure of the developing traffic jam. This highlighted the interaction between traffic jam growth and starvation - where vehicles were unable to gain access to parts of the jam due to its development. Starvation affected the shape of the traffic jam

envelope (boundary starvation) as well as the internal regions of the traffic jam. This was referred to as queue starvation.

Starvation altered the jam structure in three ways, firstly by producing 'anti-queues', secondly by creating indentations in the shape of the jam boundary and thirdly by creating a halo of reduced traffic flows outside the congested area. The extent of starvation was quantified using a fractal analogy.

The removal of the original obstruction led to one of three distinct phases which characterised the dispersal mechanism. These were (a) traffic jam disperses at varying rates, (b) traffic jam growth accelerates and decelerates at regular intervals and (c) traffic jam fails to clear without some form of external intervention.

The sensitivity of the network to gridlock was demonstrated. Furthermore, it was shown that it was best to postpone the onset of gridlock for as long as possible in order to maximise the chances of achieving total dispersal. However, there appeared to be a direct conflict between the requirement to maximise gridlock time, and the requirement to minimise the overall growth rate of the traffic jam and hence reduce delays to drivers.

Two approaches for controlling traffic queues were outlined. These were referred to as 'static' or 'dynamic' in nature. A selection of traffic management measures which could be implemented as part of a 'static' control scheme were examined. However, these were limited since they could not be applied once gridlock had already set in. This suggested the need for responsive measures which could alleviate congested situations where gridlock had already occurred. This is the subject of the next chapter.

Chapter 5

Controlling traffic jams

5.1 Summary

The previous chapter dealt with the growth patterns of traffic jams, emphasising the relationship between traffic jam structure and inherent dispersability. Building on this, we now turn to the problem of dispersal, focussing in particular on traffic jams in one-way grid networks. Using the simulation tool described in Chapter 3, we develop and test different queue control strategies for protecting such networks from gridlock and clearing them once they have broken down.

The approach is empirical and can be described as follows. Initially, the simulation tool is used to assess the impact of different forms of intervention on traffic jams in one-way rectangular grid networks. This leads to a general control principle which is then tested under a variety of network conditions. The application of counter-measures reveals some unexpected outcomes, including various forms of traffic jam migration. The chapter concludes with a qualitative discussion of the control ideas and their role in tackling traffic jams in other types of road network.

5.2 Background

Existing traffic systems fail to perform satisfactorily under prolonged congestion or over-saturated conditions. The problem has become acute because many networks are operating at saturation level. This means that any slight perturbation in the system leads to the rapid escalation of queues. If uncontrolled, these queues develop into large traffic jams which ultimately paralyse the operation of the network.

The past few decades have seen the development of a range of congestion control measures. In most cases, a combination of measures tailored to the road network configuration, the local pattern of demand and the particular control objective, seems to be the best strategy.

5.2.1 Objectives for control

Congestion control strategies can be classified in three groups: strategies which prevent the onset of congestion, strategies which alleviate the initial stages of congestion (primary congestion), and methods which lead to an expeditious recovery from severely congested situations (secondary congestion). Their success depends largely on the control objectives. Unfortunately, methods which postpone the onset of saturated conditions are not always suitable for handling saturation once it has been attained. Similarly, recovery from gridlock may not always be achieved using measures developed for preventing the spread of congestion.

Useful reviews of objectives and associated control methods can be found in Pignataro et al (1978), Huddart and Wright (1989), Shepherd (1994) and Quinn (1992). The authors provide different classifications for the measures, for example, Pignataro et al (1978) consider minimal responsive signal policies, highly responsive signal policies and non-signal treatments applied in a signalised environment. Huddart and Wright (1989) and Shepherd (1994) describe the measures as 'static' or 'dynamic' ones. Static (or pre-planned) measures are not responsive to traffic: they may be signal plans based

on average flows or physical changes to the road layout or markings, eg channelisation. Dynamic measures are responsive since they vary according to the conditions on the network.

5.2.2 Types of control

Static control methods

Several examples of static measures are described by Quinn (1992) and also by Shepherd (1994). These include congestion pricing which frees the roads for those who pay and, could in principle, provide a better service for bus passengers, through investment of the proceeds in public transport. Pre-planned signal settings aimed at minimising delay or at maximising capacity represent another form static control. The use of coordinated traffic signals for linear corridor systems has led to the idea of a 'reverse' signal progression which is particularly effective in highly congested conditions. Static control may also be imposed to regulate demand via traffic metering and restraint methods (Rathi, 1988, 1991).

However, static signal control plans are restrictive since they are mainly designed for fixed-time operation, in which the control algorithm changes at preset times to suit the anticipated traffic demand. But if the algorithms are applied too late, or if unexpected congestion occurs, queues build up quickly leading to catastrophic collapse. This suggests the need for dynamic or on-line control which responds to changes in the traffic situation.

Dynamic control methods

Dynamic control systems require a supply of traffic data, usually obtained from loops or detectors embedded in the road surface. There are many ways of treating the data and various control algorithms, see Shepherd (1994) for a resumé on existing traffic systems with dynamic capabilities.

Dynamic control measures may take on various forms. These include local-scale measures as well as area-wide measures. Local measures apply to a single intersection or pairs of intersections, whereas area-wide measures extend to cover larger sections of the urban road network.

Traffic metering combines both local and area-wide control. The purpose of metering, as applied to urban street networks, is to maintain a level of traffic density within the control area that is acceptable to drivers. This is usually achieved by temporarily storing vehicles at the periphery of the control (metered) area. These vehicles may be re-routed or accumulated around the controlled area. If alternative routes do not exist, or if there is insufficient buffer storage area, then long queues will begin to form. Ultimately, these may develop into gridlock conditions so that the net effect of the control scheme is to transfer the congestion elsewhere. Consequently, the availability of sufficient storage space is a pre-requisite for this form of control.

Five major projects for controlling congestion through metering-based schemes are reported in the literature. These are: Nottingham Zones and Collar Project (Vincent and Layfield, 1977); Southampton Bus Demonstration Project (Dept of Environment, 1976); Bordeaux Traffic Control Project (Franceries, 1977); Dusseldorf (OECD, 1978) and New York's Central Business District (CBD) area in Manhattan (Rathi, 1988, 1991).

In general, the European experience with metering control schemes has been successful, although the Nottingham experiment is the exception. The New York schemes are particularly interesting because they have been developed for a one-way grid environment. The schemes involve two forms of metering strategies: internal and external. Internal metering includes

- Critical Intersection Control (CIC) which is aimed at congested conditions that are local in character, i.e. queues do not extend into upstream intersections,
- arterial strategies that control the flow along congested arterial roads; and

- grid strategies that control the flow along major arterials and along minor cross-streets in a way that prevents 'gridlock' conditions.

Internal metering is applied within the control area and is based on a 'spillback avoidance' approach rather than the conventional 'progressive movement' approach which is inappropriate when standing queues form and do not clear during green signal indications. But, once spillback has occurred, the success of the proposed strategy is limited.

External metering, on the other hand, aims to limit the inflow of traffic into the control area and thereby improve the overall quality of traffic flow within the control area. A number of external metering control strategies are described by Rathi (1991). These include uniform metering (various levels) at all entry points of the control area; variable metering for the control area, and metering of traffic at specific locations only. The experiments were conducted using the TRAF-NETSIM simulation model and the results indicate comparative success with various combinations of the strategies. Depending on the particular control objective, the strategies improved trip productivity within the control area and reduced vehicle hours of travel. The variable metering control strategy was most effective in reducing travel time.

However, like the internal metering strategies, the external control schemes can only be applied to reduce or prevent the onset of severe congestion. But once gridlock conditions arise, the strategies are no longer applicable. In addition, the methods have been developed for a dedicated road network. This makes the generalisation of the results more difficult.

Gating is another example of dynamic control which may also be applied either locally, to protect a particular junction, or on an area-wide scale. With gating, certain links are identified as critical or bottleneck links. The gated links are those links which have been designated to store the queues which would otherwise block the critical links. When the critical link is too busy, the green time is reduced on the

gated links.

However, the traditional approach to gating is limited because it selects one or more critical intersections and uses a crude on-off logic with an abrupt cut in green time to store the traffic on a predetermined link. A less crude gating approach has been developed by Shepherd (1995) who describes various 'auto-gating' strategies that can be implemented to control arterial signalised networks. With auto-gating, each intersection meters the traffic into the downstream links. As the inbound flow increases and queues form, then the inbound green times are cut *gradually* at each intersection. This has the desired effect of queue relocation, avoiding blocking-back. The 'auto-gating' principle is demonstrated using various ramp-metering strategies which are applied to control over-saturated conditions in grid networks. However, the strategies can only be used to reduce the effects of blocking-back thereby preventing the onset of gridlock. But once gridlock occurs, the 'auto-gating' principle fails to remedy the situation and the system breaks down.

This is a hallmark of the traffic management measures described so far, because whilst most can increase capacity and hence postpone the onset of severe congestion, they are not suitable once gridlock has set in. In this chapter, we propose a range of effective mechanisms for treating traffic jams where gridlock has already occurred.

5.2.3 Recovery from severe congestion

It is convenient to consider the recovery from gridlock in three stages, each of which leads on to the next. The stages are detection, treatment and evaluation. The detection stage includes establishing the cause of the congestion (for example, using closed circuit television monitoring, infra-red technology or cellular phone callers), ascertaining its severity and determining whether gridlock has occurred. Following this, an appropriate control scheme may be implemented to treat the affected regions. This is followed by an evaluation stage during which the effectiveness of the treatment is monitored and subsequently assessed.

Detection

In Section 4.3.4 we ascertained that the formation of a gridlock cycle represented an important threshold within the evolution of the traffic jam and later showed how certain static control measures became increasingly ineffective once gridlock has occurred. It follows that gridlock detection forms an integral element of the control process.

However, whilst many traffic systems can detect the formation of queues at specific intersections, few can determine if gridlock has occurred. In other words, the systems detect queues but cannot easily recognise patterns or chains of queues.

One exception is the SAGE system being developed in France, Foraste and Scemama (1987). This, in part, is due to expert system methodology which underlies it. The expert system comprises a knowledge base of facts and rules. The system uses these facts to compose 'congestion chains' via inference and deductive reasoning. Thus, a gridlock can be recognised via a sequence of logical conditions which lead to the required inference when analysed recursively, (Scemama, 1995). This information is then communicated to the traffic controller in visual form, who may then choose to implement the suggested form of control. A history of congestion patterns can also be generated and stored so that the system could eventually learn and *predict* from situations.

Treatment

There appears to be three approaches for tackling severe congestion. These are summarised by Huddart and Wright (1989) who suggest that (a) the control system be altered to disperse or free critical queues, (b) reserve capacity be provided to relieve congested links, or (c) the level of demand be reduced albeit temporarily.

Several practical schemes are reviewed by Quinn (1992). These include reduced cycle times for the period of recovery, sterilisation of spare capacity as well as the use of

driver information systems.

Short cycle times allow a high saturation flow to be maintained throughout the green period, enable the clearing of junctions blocked by turning traffic and provide more frequent opportunities for pedestrians to cross.

The provision of extra storage space for queues may be achieved for example, by allowing general traffic to use bus lanes in emergency conditions. Variable message signs could be used to inform drivers of the measure being implemented.

Driver information systems (eg dynamic route guidance) aim to direct people to their destination using the best route under prevailing conditions. This implies that drivers will be diverted around incidents and congestion. However, when a traffic jam is detected, it may not be possible to transfer traffic to alternative routes without causing a jam elsewhere, or imposing unacceptable environmental or safety costs to local residents.

Gray and Ibbestson (1991), describe several traffic control strategies used by traffic engineers for the treatment of recurrent congestion. The authors identify 'metering' as being a valuable tool for limiting the extent of damage while extended greens were used to recover from severe congestion.

Evaluation

Evaluation takes place during and after the implementation of control strategies. There are two types of evaluation: short and long term. Traffic engineers need continually to review the success or failure of a particular control strategy during and after the control period. This should be supplemented with a long term perspective on the overall changes induced by the specific measures. For example, a selection of control measures may appear to clear a traffic jam in the short term, but in the long term, the treatment fails as the jam recurs elsewhere in the network. This phenomenon will be referred to here as jam migration. We will discuss this in more detail

later in the chapter.

In the next section, we describe several counter-measures which can be implemented to control traffic jams where gridlock has already occurred. First, the principal control mechanisms are introduced and experimented with under different network conditions. A particular combination of these measures leads to the formulation of an integrated control strategy whose performance is then evaluated under a variety of network conditions and assumptions. The issues of traffic jam migration and vehicle diversion are raised in this context. A discussion of how the control strategy may be applied to treat gridlock in other types of network brings this chapter to its conclusion.

5.3 Counter-measures for traffic jam control

The form of a traffic jam in a simple, idealised, one-way road network is characterised by gridlock cycles which develop at specific locations within the jam area. These cycles persist even when traffic demand falls away at the end of the peak period. External measures are required to break the interlocking queues apart in order to restart vehicular movement. In real life, these measures take the form of a set of rules, which channel vehicles away from the sensitive locations of the network. However, if demand is heavy, control measures such as these may only lead to a temporary halt in jam growth, which is later resumed.

5.3.1 Types of intervention

The strategies involve the application of vehicle bans to a number of critical junctions in the network. The bans come in two forms: turn or ahead. Turn bans are imposed on selected links to break gridlock cycles at the nucleus of the traffic jam. Ahead bans are implemented around the traffic jam envelope to reduce input into critical sections of the road.

Experiments with bans have lead to a number of possible arrangements which may be applied simultaneously or in isolation. The control plans can be divided into three groups: (i) nucleus intervention aimed at the gridlock cycles in the traffic jam centre, (ii) boundary intervention directed at the envelope of the traffic jam, and (iii) a combination of both.

Nucleus intervention

Nucleus intervention is directed at, or near, the point of the original obstruction. With this arrangement, turn bans are superimposed at specific junctions to resolve the conflicts which arise when a gridlock cycle forms. This can be seen in Figure 5.1 where the four locking turns are forced apart by restricting the movement along these critical links to *ahead* only. The strategy will be referred to as the *block* strategy owing to the arrangement of the bans around four sides of a city block.

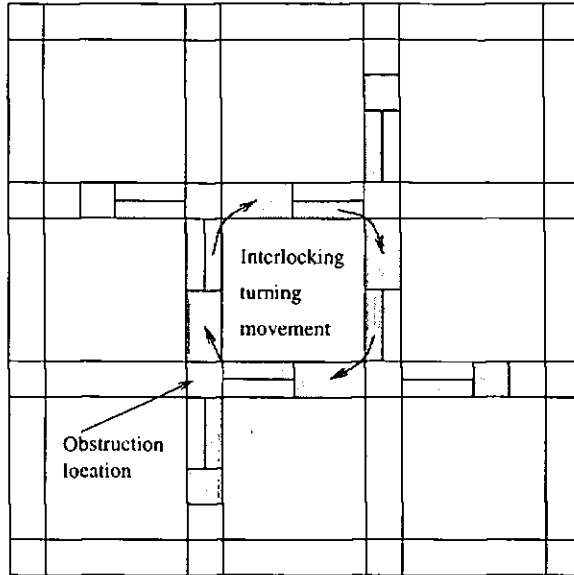
If the demand is fairly light or if the traffic jam is in its early stages of development, it may be sufficient to ban a small number of critical turning movements to clear the whole jam. However, experiments have demonstrated that this expedient often forces the original gridlock cycle to migrate to a new location, close by in the network. This problem has been dealt with by using multiple *block* strategies which have treated gridlock migration phenomenon with increasing success. But multiple *block* strategies become less effective when demand increases and when the jam has already assumed area-wide proportions. This is explained in more detail in Section 5.4.

Boundary intervention

An alternative strategy (referred to as the *envelope* strategy) sets a cordon of *ahead* bans around the periphery of the traffic jam. The ahead bans can be used in two ways. Vehicles using the banned junctions can either be queued outside the congested region or else be re-routed away from the jam.

The *envelope* strategy resembles the external metering approach described by Rathi

No treatment: gridlock forms around city-block



Turning prohibited along links L1, L2, L3 and L4

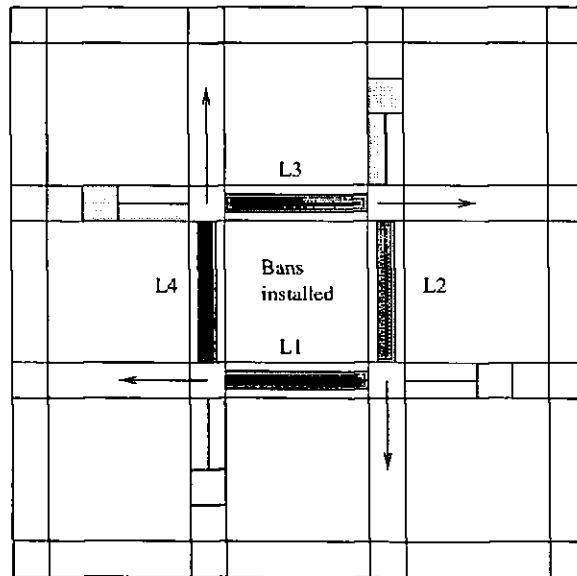


Figure 5.1: Application of the 'block' strategy

(1991) although there are some differences. For example, one of the pre-requisites of external metering control is the availability of a sufficient buffer (storage) for the metered vehicles at the periphery of the control area. This effect was not simulated in the aforementioned study and hence not accounted for in the results.

By contrast, the boundary intervention strategies developed using our simulation tool take the queueing conditions along the periphery of the control area into account. The severity of the resulting congestion will depend amongst others, on the availability of reserve storage, on how long the control scheme is set in place and also on the level of throughput allowed.

A combined approach

Applying the *block* and either of the *envelope* strategies in parallel enables a quick and effective dispersal of jams on the idealised network. The *envelope* strategy protects the emerging jam from excessive demand whilst the single or multiple *block* strategies are applied directly to gridlock cycles at the heart of the jam. The shielding of the cordon minimises the tendency of the gridlock cycles to reform at nearby locations and the reduction in the overall jam size and delay time can be dramatic within a few time slices of the simulation process. The combined approach has been extensively tested and is dealt with separately in Section 5.4.2.

5.3.2 Notation and assumptions

We have used the symbols shown in Table 5.1 to denote the different control strategies which have been developed using the simulation tool. The notation is intended to be informative, for instance a *block* strategy is represented by the letter B . The subscript is used to distinguish between a single *block* strategy (B_1) and multiple *block* strategies (B_3, B_4). Similarly, the two versions of the *envelope* strategy are denoted by the letters Q , where vehicles are queued on the boundary to the control area, and R , for the re-routing option.

Strategy Type	Strategy Definition	Notation
Nucleus	Block	B_i
Boundary	Envelope: queueing	Q_j
Boundary	Envelope: re-route	R_j
Combination	Block and Envelope: queueing	$B_i Q_j$
Combination	Block and Envelope: re-route	$B_i R_j$

Table 5.1: A summary of notation used to describe control strategies

The extent of the cordon can be altered according to the size of the growing traffic jam. The subscript used in the definition of the *envelope* strategy describes the size of the cordon. For example, Q_4 implies a boundary queueing strategy, centered on the original obstruction, which extends over four blocks in each direction.

The application of strategies in parallel is described using a combination of the above symbols, for example, $B_4 R_6$ represents the situation in which four *block* strategies are implemented in parallel with a re-routing envelope encompassing six blocks in each direction. Other combinations can be defined intuitively.

It is assumed that the obstruction has been removed prior to the application of external counter-measures. We will use *none* to describe the situation in which no control has been implemented other than the removal of the original cause, and *fail* when the suggested control scheme fails to clear the jam. In all other cases, an estimate is made of the total delay incurred by the presence of the traffic jam.

To ascertain whether traffic congestion has sufficiently cleared it is possible to estimate the amount of spare space capacity available in the network. However, with a traffic jam, this measure is of limited use since (a) there are areas in the traffic jam which do not accumulate vehicles at all ('anti-queues') and (b) there may be some spare capacity on a blocked link which cannot be used due to the structure of the queues occupying the link. Consequently, we have developed two alternative criteria to signify

the demise of a traffic jam.

The first was qualitative and consisted of a graphical analysis of the jam structure, whereas the second was quantitative. For the spatial analysis, the extent of the traffic queues was observed and any potential gridlock cycles were identified. Providing these were sufficiently fragmented then it was concluded that the strategy had been successful in clearing the traffic jam. The second criterion was then used to quantify the strategy's performance. This was measured in terms of the total delay incurred during the dispersion operation and was usually supplemented with a second statistic: the length of the dispersion period.

The dispersion period was measured as the number of time-slices required until the network re-established its steady state. The dispersion period commenced with the removal of the original obstruction and terminated when the number of vehicles waiting in the system lay within two standard deviations of the average number of vehicles waiting in the system when in the steady state (\bar{w}). We have used the statistical convention of two standard deviations because it often took a substantial amount of time for the number of waiting vehicles to converge on the original estimate \bar{w} , in spite of the fact that the system had already re-achieved its steady state.

Finally, the simulation model assumes that the proportion of turning does not change as a result of an incident. In practice, drivers will react to congestion and may alter their original route. However, in the isotropic network which has been presented here, it is unclear how drivers would respond because the routing is not goal-driven. In applying the strategies to non-idealised networks, the problem of self-diversion will have to be addressed specifically. Before doing this, it will be necessary to introduce routing patterns in the model.

Having outlined the principal mechanisms of the control strategies, we now turn to their application and investigate their effectiveness in treating area-wide traffic jams which are affected by gridlock conditions.

5.4 Results

In Chapter 3, we described the operation of the simulation tool in three stages. These included a stabilisation period, a jam evolution period and an optional dispersion stage. In this section, we use the one-way simulation model to explore the impact of the different types of intervention strategies on the dispersal process of a traffic jam. The control of two-way traffic jams is dealt with in a separate section.

The results are based on a series of simulation experiments in which a traffic jam was allowed to develop for a fixed number of time-slices. The obstruction was then removed and different combinations of *block* and *envelope* strategies were subsequently introduced. The traffic pattern was identical for each simulation experiment.

The series of experiments was repeated for a range of network conditions. These highlighted three parameters which had an appreciable impact on jam development and dispersal. These were the level of demand (μ), the proportion of turning vehicles (p), and the extent of channelisation (α). The results of the simulation trials are summarised in Table 5.2 and in Figures 5.2 and 5.3. These show the respective performances of different types of control for increasing levels of demand, turning proportion and for various channelisation regimes. Where the strategy was successful, its performance was recorded in terms of the total delay incurred. The total delay was calculated using the methods described in Appendix B. The delay estimate is supplemented with the length of the dispersion period which appears in brackets beneath the total delay.

But if the strategy failed to clear the traffic jam, then the experiment was halted prematurely and a 'failed' outcome was recorded. The formation of a gridlock cycle on the network was used as an indicator of whether the traffic jam was likely to clear.

5.4.1 Performance of intervention strategies

Increasing demands

The results in Table 5.2 indicate higher delays for each increase in the level of demand. *This trend was observed regardless of random fluctuations in the demand pattern. See Appendix B for details.* In other words, as the road network approaches saturation level, it becomes increasingly more difficult to disperse the traffic jam. The time taken for the network to return to its steady state also increases with the level of demand, μ . The proportion of turning, p , was fixed at 0.25 and the stopline widths were allocated equally between ahead and turning vehicles. Similar effects were found with other turning proportions yet we have recorded the results for $p = 0.25$ since in real networks, the proportion of turning does not regularly exceed this percentage.

By considering each column of entries it is possible to compare the effect of the various types of strategy on the overall delay. For relatively low demand levels ($\mu = 18, 19$ veh per min per entry point) there is no apparent need for the implementation of external counter measures, and the traffic jam disperses once the original obstruction has been removed. Counter-measures do not help, on the contrary, (particularly versions of the *envelope* strategy) they usually increase delay. This is because of the additional restrictions imposed through the implementation of unnecessary control strategies.

With average levels of demand ($\mu = 23, 25$ veh per min per entry point) it may no longer be sufficient to remove the obstruction and further intervention is required to clear the traffic jam. This is mainly due to the irreversible gridlock cycles that have formed at the jam centre. Consequently, some form of nucleus intervention is required in order to re-start vehicle movement. In some situations, single or multiple *block* strategies may be sufficient.

However, repeated simulations have shown that a more effective strategy is to combine the *block* strategies with an appropriate form of boundary intervention. This strategy

has been used with increasing success as demand levels approach saturation. This is because with higher levels of demand (eg $\mu = 27$), the rate of growth of the traffic jam is more rapid, and the jam assumes an area wide structure. The *block* strategies operating at the nucleus of the jam cannot cope with the excessive demand caused by the large number of queues in the structure. A shielding mechanism is required so that vehicles are deflected away from the congested area for a limited time period. During this time period the *block* strategies may be applied to the nucleus of the jam, and the cordon placed around the jam acts as a barrier.

The cordon can be set up so that it neatly encompasses the jam at whatever node it originates. The cordon can be altered in size. Invariably, a larger cordon implies a higher number of restrictions (and hence heavier delays) being imposed on the network, so that the overall aim is to define a cordon which is as small as possible. As with the *block* strategies care must be exercised when fitting cordons to emerging jams. An ill-fitting cordon which doesn't capture the complete jam produces irreversible gridlock cycles outside the boundary. This occurs if the cordon is left in place for too many simulation time-slices and is heightened in situations of severe demand. Effectively, the congestion problem is transferred from the jam centre to the periphery, and the strategy fails to clear the traffic jam. The strategy also fails as the network approaches saturation level and no spare capacity is available.

Finally, looking at the results in rows four and five of Table 5.2, one can observe that boundary intervention on its own is not advisable since it does not actively clear the traffic jam. In fact, boundary intervention on its own only works where gridlock has not already occurred. Under these circumstances, there is no apparent need for the strategies since the traffic jam clears itself as soon as the obstruction has been removed.

However, with the onset of gridlock, boundary strategies are important because they protect the control area from excessive demand, and can be extremely powerful when applied together with *block* strategies.

Parameters						
μ	18	20	23	25	27	29
\bar{w}	6035.4	6598.6	7483.2	8000.2	8487.7	9663.5
s_w	42.4	47.1	58.2	53.9	62.7	86.7
Strategy	Total Delay (veh-min)					
A	2550.2 (14)	5403.6 (19)	FAIL	FAIL	FAIL	FAIL
B₁	2550.2 (14)	5788.8 (27)	17123.0 (33)	FAIL	FAIL	FAIL
B₃	2581.2 (15)	5767.2 (27)	16455.1 (33)	FAIL	FAIL	FAIL
C₆	2847.2 (11)	5258.4 (14)	FAIL	FAIL	FAIL	FAIL
Q₆	5925.6 (20)	10444.8 (32)	FAIL	FAIL	FAIL	FAIL
B₃C₄	2882.9 (11)	5460.0 (15)	9982.0 (28)	FAIL	FAIL	FAIL
B₃C₆	2950.2 (11)	5409.6 (14)	9376.6 (28)	19915.0 (35)	49450.5 (99)	FAIL
B₃Q₆	6354.4 (19)	10821.6 (27)	18354.0 (28)	38801.0 (46)	FAIL	FAIL
B₃C₈	2514.6 (11)	4827.2 (14)	9569.8 (28)	24052.5 (45)	49236.7 (99)	FAIL

Table 5.2: Total delay(veh-min): Increasing Demands

The results in Table 5.2 correspond to an experiment in which the seed value remained constant throughout the simulation tests. We have also investigated the effects of random seeds on the total delay for various control strategies. In general, the pattern of delay followed a similar trend: lower demands implied lower delays although some variation was observed in the actual delay experienced with a particular strategy. The details are provided in Appendix B.

Increasing turning proportion

The effect of turning proportion on the performance of different control strategies has been evaluated via a series of simulation experiments in which all the parameters of the model remained fixed except for the turning proportion which varied from one experiment to the next. In particular, the level of demand was maintained at ($\mu = 25$ veh per min per entry point). The experiments were similar to the ones described in the previous section: a traffic jam was simulated for a fixed number of time-slices, which was followed by the removal of the obstruction and the implementation of counter-measures. These were limited to the multiple *block* strategy (B_4) and combinations of the multiple block strategy B_4 and re-routing *envelope* strategies of various sizes (B_4R_4 , B_4R_6 and B_4R_8). These appeared to be particularly effective in clearing the resultant structures.

The results of the experiments are shown in Figure 5.2. The data for the figures is provided in Appendix B. The graphs show that the strategies perform less satisfactorily as the turning proportion increases. However, unlike with the previous set of experiments, it is difficult to measure this effect. This is because the behaviour of the strategies does not follow a systematic trend with respect to increases in p . Even when the interval between successive p values was reduced further, it was not possible to identify a particular pattern of behaviour.

This may be due to the discrete nature of the software. At each stage in the simulation, the number of turning vehicles is subject to numerical rounding. Thus, one

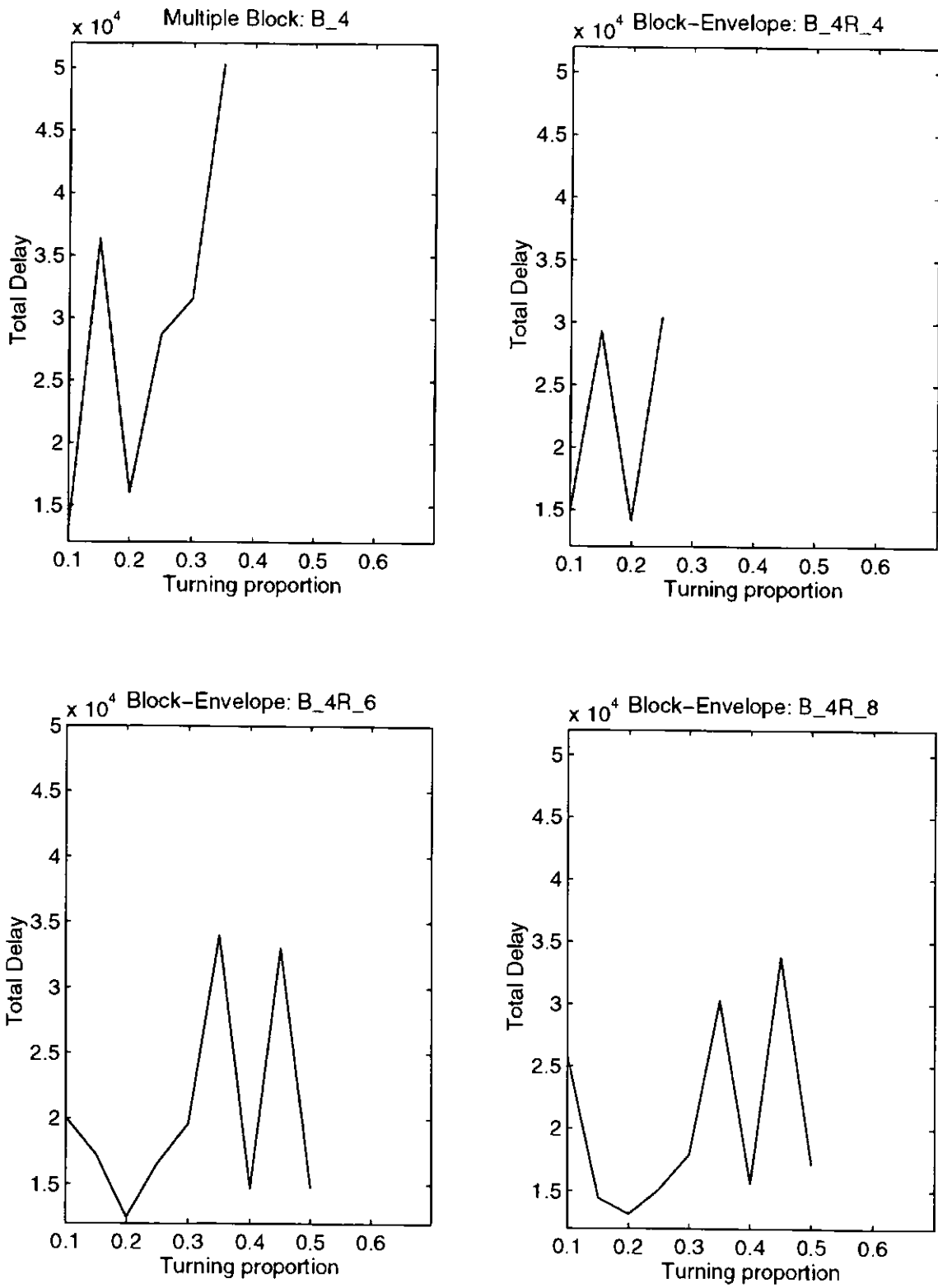


Figure 5.2: The performance of control strategies for increasing turning proportions

combination of network parameters may induce circumstances which are more suited to the proposed control measures. Hence the delay might be reduced. This may not be the case for a higher proportion of turning vehicles.

Referring to the individual plots in Figure 5.2, one can observe that the delays incurred using strategy B_1 are higher than for the remaining graphs. This suggests that a protecting envelope of restrictions improves the quality of flow in the congested area. This hastens the dispersal process and leads to a reduction in the overall delay.

Referring to the top-right figure (B_1R_4), one can observe that this strategy fails to clear most of the simulated traffic jam structures and only works for a small range of turning proportions ($p \leq 0.25$). By contrast, B_4 on its own works well for a wider range ($p \leq 0.4$). This is because the cordon of restrictions imposed via the application of R_4 did not always capture the traffic jam entirely. This led to additional queueing and gridlock problems near the jam boundary. Ultimately, this paralysed the operation of the strategy. This case serves as an example of how inappropriate treatment can sometimes be extremely damaging and not worthwhile.

By contrast, strategies B_1R_6 and B_1R_8 were suited to the structure of the traffic jam and boundary queueing problems did not occur. Furthermore the strategies were effective for a wide range of turning proportions as shown in the bottom two graphs. Both indicate similar patterns of delay.

Whilst it is not possible to predict how the delay will vary with respect to changes in turning proportion, the results show that for a fixed level of turning, the combination of *block* and *envelope* strategies is most effective and produces the shortest delays.

Further experiments have been conducted to examine the effect of random seeds under increasing turning proportions. The results of these experiments are presented in detail in Appendix B. However, the random fluctuations in the delay patterns which were observed in this case were also found in the extended set of experiments. In other words, no conclusive reason could be identified for these fluctuations.

The effect of channelisation

The effect of channelisation on the rate of traffic jam growth and dispersal was examined in Section 4.3.5 of the previous chapter. The analysis found that in order to minimise the rate of traffic jam growth it would be desirable to aim for a layout in which the stopline widths were allocated to the segregated queues in exactly the same ratio as the demands. But this was in direct conflict with the requirement for postponing the onset of gridlock, which recommended that a *larger* proportion of stopline width be allocated to the turning traffic.

These suggestions formed part of a selection of *static* control measures which could be implemented in advance to minimise traffic jam expansion or to aid dispersal. In this section, we evaluate the effect of channelisation on the *dynamic* control strategies proposed in this chapter.

Figure 5.3 summarises the results of an experiment which considered the performance of the B_4Q_6 control strategy on the dispersal process of a traffic jam for various channelisation regimes. (The data for the figures is provided in Appendix B.) The strategy involved the simultaneous application of four *block* strategies together with the *queueing* version of the *envelope* strategy, where vehicles queue at the jam boundary until the restrictions are relaxed.

In the experiment, the proportion of turning varied from $p = 0.10$ to $p = 0.65$. Different stopline arrangements were investigated by changing the value of the parameter α . For example, $\alpha = 0.50$ represented the situation in which the stopline widths were allocated equally between ahead and turning vehicles. The remaining parameters were fixed throughout the simulation experiment.

The results in Figure 5.3 indicate that the strategy is most effective when $\alpha = 0.50$ (top-left figure). In addition, the success of the control strategy can be reduced substantially through interference with the stopline width allocations. This is what one would expect in practice: allocating proportionately more stopline width to the ahead

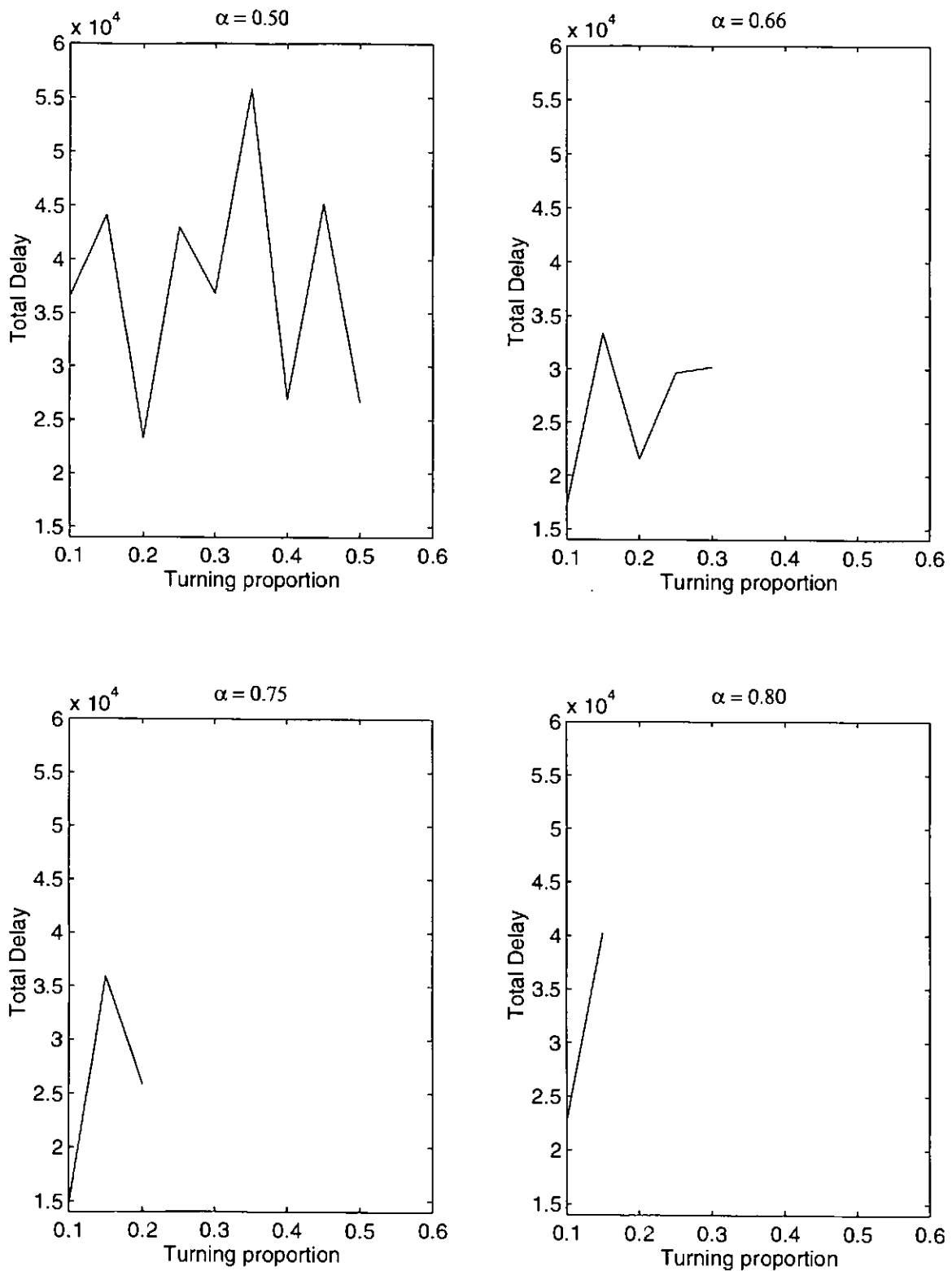


Figure 5.3: The performance of B_4Q_6 for various channelisation regimes

traffic simultaneously reduces the amount of stopline width devoted to the turning traffic. This is helpful when the turning proportions are low, for example $p = 0.10$, but counter-productive as p increases because, the reduction in turning stopline width triggers spillback more rapidly and induces a premature onset of gridlock phenomena. This makes dispersal more difficult.

This experiment was repeated using random seeds. The results for the extended set of simulation trials are presented in Appendix B. The results indicate a similar trend to the ones shown here. With this particular control strategy, B_4Q_6 , the delays involved are relatively high. Looking across the vertical axes of the respective figures one can observe that the overall delay can be reduced and often minimised by slightly changing the link storage arrangement. The results show that the total delay can be reduced by setting $\alpha = 0.75$. In this example, the turning proportions are low, ($p < 0.20$), and the majority of vehicles at a particular intersection will move ahead. The optimal exploitation of the road link storage capacity would be one which allocates a higher proportion of space to the ahead queues, so that the widths allocated to the segregated queues are in proportion to the demands. However, in reality, this configuration would hardly ever be used and so the result is only of theoretical interest.

However, unnecessary interference may sometimes be damaging as can be seen in the bottom-right graph. Here, too little space has been devoted to the turning movement so that spillback is more rapid and the total delay increases as a result of changing α from $\alpha = 0.75$ to $\alpha = 0.80$. In other words, the benefit of allocating proportionately more stopline width to the ahead traffic (by devoting 80% of the width of the road rather than 75% does not help and may cause unwanted damage. This is due to the fact that the remaining storage space for the turning traffic (a mere 20%) is insufficient relative to the demand. This triggers premature spillback in the turning segregated area and higher rather than shorter delays. It follows that the benefits of such changes are small, suggesting that interference with the α parameter is best avoided.

Finally, referring to Figure 5.3 one can observe that in general, the control strategy appears to be suited to Type I traffic jams and works best when $\alpha = 0.50 < 1 - p$. But when the proportion of turning increases beyond this threshold (Type II traffic jams), the strategy become less effective and fails to clear the jam. Since the situation of $p > 0.50$ would almost never occur, this result is only of theoretical interest.

Queueing instead of re-routing

The effects of boundary queueing may be reduced by replacing the queueing restriction with a *re-routing* option. Thus, on approaching the controlled area, vehicles are diverted or channelled away towards less congested regions in the network. However, in the simulation model, once re-routed, vehicles do not adhere to their original planned routes. This implies that the queueing option involves heavier delays to drivers caught in the congestion. However, re-routing takes drivers away from their destinations. In reality, they would drive around until allowed to go where they wanted and the delays might be large. But this effect was not simulated in the model and hence not accounted for in the results.

Figure 5.4 compares the delays incurred using the re-routing and queueing alternatives (B_4R_6 and B_4Q_6) under a particular configuration of network parameters.

However, the results from the re-routing option might be misleading since the results do not include the delay that might be incurred through driving around until reaching an intended destination. Hence the actual delay might be higher than the value shown in Figure 5.4. The results from the queueing option, therefore, represent more realistically what might happen under the particular assumptions made. This pattern has been observed using other combinations of control strategies.

5.4.2 The integrated solution

The analysis of control strategies for different demands, turning proportions and stopline width allocations has established that treatment which is aimed solely at the

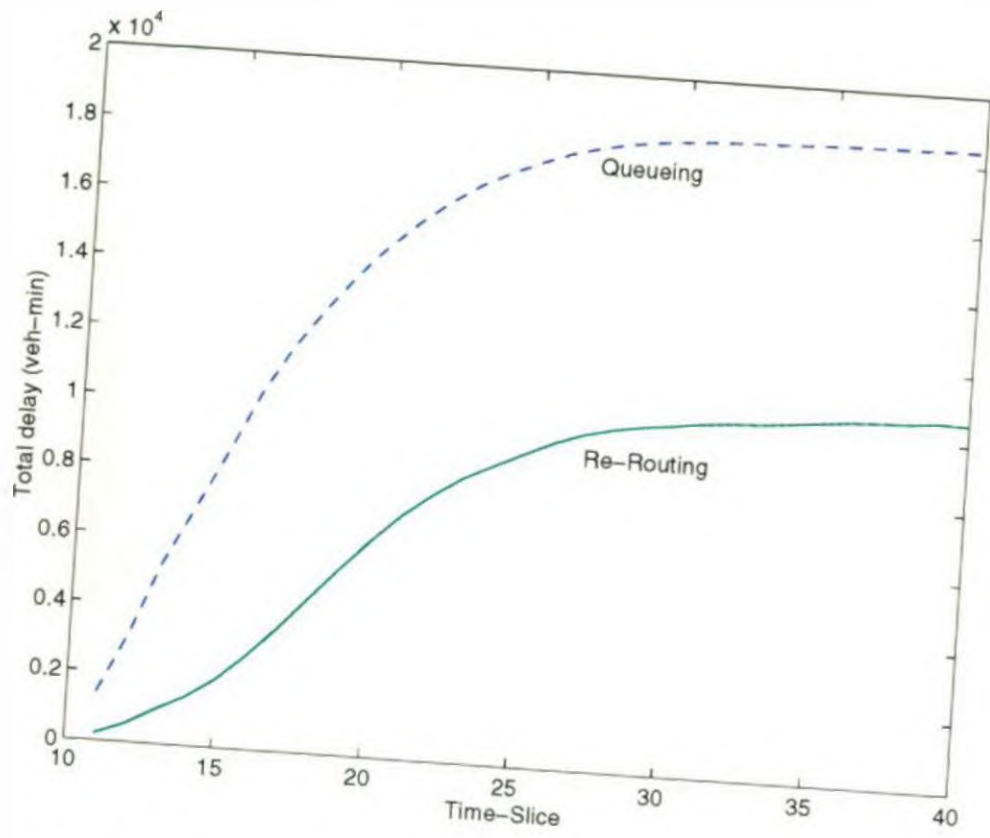


Figure 5.4: *Queueing vs re-routing: a comparison in delay times*

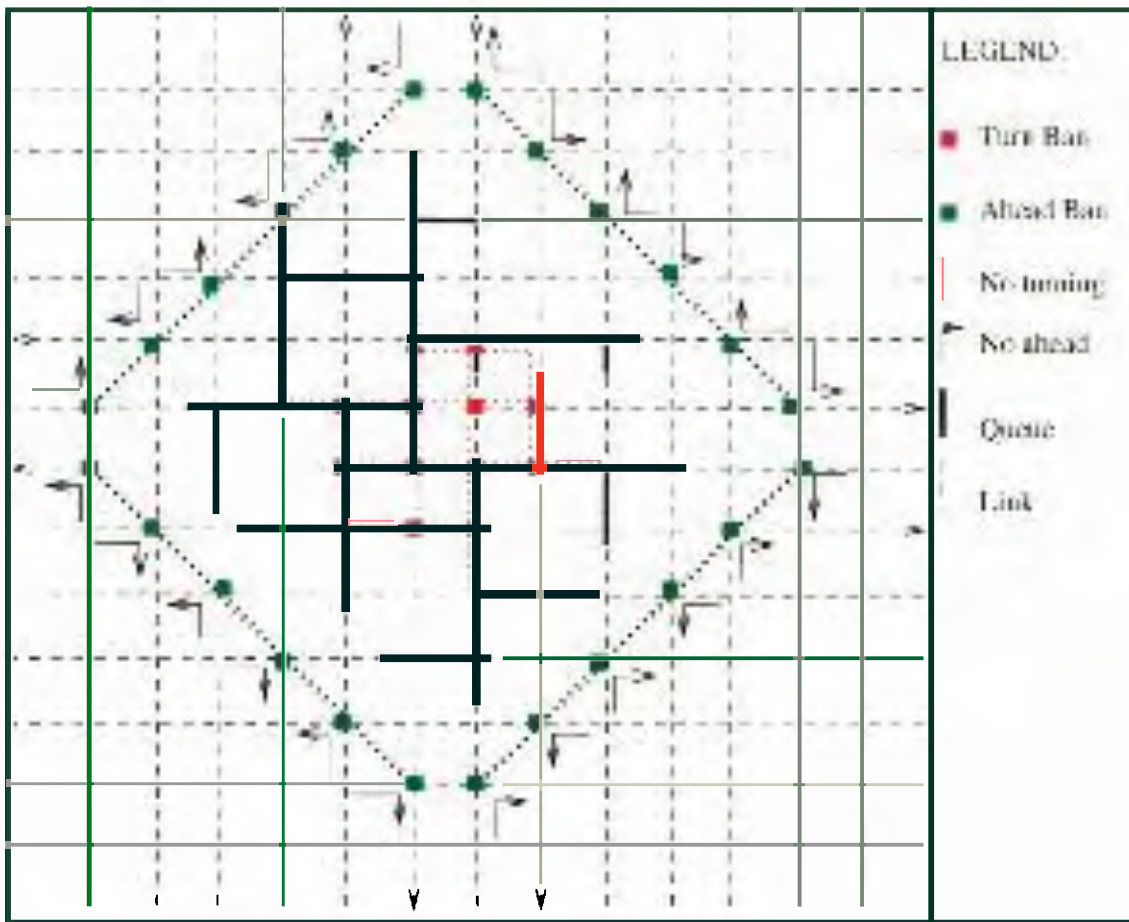


Figure 5.5: An integrated control scheme

traffic jam boundary (*envelope* strategy) will only work providing the gridlock cycle at the centre of the jam has not already formed and sufficient storage area is available for the vehicles waiting at the periphery. Likewise, the success of *block* strategies (turn bans directed at the gridlock cycles) on their own is not always guaranteed.

Experimentation with various combinations of *block* and *boundary* strategies have yielded a control plan which will be referred to as the integrated control scheme. The integrated strategy is based upon a particular arrangement of bans which is illustrated graphically in Figure 5.5. In terms of the notation developed in Section 5.3.2, this arrangement corresponds to B_4R_j or B_4Q_j depending on the type of ahead ban (re-routing or queueing) being used.

Sensitivity analysis

So far, we have observed that the three parameters μ (the level of demand), p (the proportion of turning vehicles) and α (the allocation of stopline widths between ahead and turning vehicles) each had an appreciable impact on the dispersion process of a traffic jam. Let us now consider the effectiveness of the integrated control scheme under the combined influence of increasing levels of demand, turning proportions and different stopline width constraints.

In the experiment, traffic jams were simulated under a range of conditions. Demand varied from $\mu = 18$ to $\mu = 28$. The proportion of turning vehicles ranged from $p = 0.1$ to $p = 0.9$. (Note that in practice, $p = 0.3$ represents the plausible maximum. The entire range of values of p was considered for theoretical completeness.) Different stopline width allocations were investigated by varying α as described above. Following the period of growth which was limited to ten time-slices to prevent the traffic jam from overflowing the system boundary, the obstruction was removed. Thereafter followed the implementation of the integrated control scheme (B_4R_6). With this strategy, turn bans were imposed at particular locations in the jam centre, whereas ahead bans were implemented along the jam boundary. These restrictions channelled vehicles away from the jam following which they continued to progress through the system as before. Finally, the outcome of the experiment was recorded, either in terms of the total delay (See Appendix B for details) incurred if the strategy had been successful, or as a 'fail' otherwise.

With the integrated control scheme, the *boundary* restrictions remained in place for a short time (typically three time-slices) but the *block* strategies stayed in longer. There are two reasons for this. Firstly, to exploit the time lag that exists between the removal of the gates and the arrival of the released cordon traffic at the jam. Thus, it is possible to relax the boundary restrictions before the traffic jam has cleared completely. Secondly, the early removal of the boundary restrictions minimises the tendency of queueing along the periphery and reduces the chances of new gridlock

cycles forming there.

The results of the experiment are summarised graphically in Figure 5.6, and can be interpreted as follows. The figure is divided into five graphs. Each graph shows the operative domain of the control strategy for a particular arrangement of stopline widths. The operative domain is the coloured region. This is where the strategy has successfully eliminated the traffic jam. The areas outside the coloured region correspond to regions where the strategy has failed to clear the jam. The total delay incurred for a particular combination of turning and demand levels is indicated by the intensity of the colour: lower delays are indicated by shades of blue and higher delays are shown in shades of purple. The objective is to identify a plane with maximum area with the lowest corresponding delay.

This choice of display has been selected because it could be assist traffic controllers in deciding which strategy to adopt. In other words, an operator might simulate the effects of a particular strategy under specific conditions and view the results graphically rather than consider tabulated columns of data. The most effective strategy would correspond to the plane which extends over the largest area with lowest delays. This decision could be made quickly and the associated control strategy would subsequently be implemented.

In this example, one can observe that the shapes of the planes display similar characteristics with respect to the changes in α . However, as α increases, the integrated strategy becomes less effective, as indicated by the reduction in the planar area. This means that regardless of other parameters, allocating more space to the ahead segregated queues is not to be encouraged if the overall aim is to disperse traffic jams. However, in the region where the control policy successfully treats the traffic jam (coloured areas), the results show that, the total delay may be reduced slightly by carefully increasing α beyond 0.50. However, the benefits of such changes are small, suggesting that interference with the α parameter is best avoided. This (Note, we are not so much interested in the exact shape of the areas, but more in the size of the

coloured regions. We are interested in observing as large a region as possible with lowest corresponding delays.)

The results for $\alpha = 0.33$ deserve specific attention. This layout allocates two lanes of turning traffic for each lane of ahead traffic) and would almost never be used in practice. Even so, the results indicate a relatively large feasible region where the strategy is successful. This confirms the hypothesis suggested in Chapter 4. that allocating a *larger* proportion of stopline width to the turning traffic would aid dispersal. However, by comparison with the results for $\alpha = 0.50$, one can observe that for high levels of demand and low turning proportions the strategy fails when $\alpha = 0.33$ but works when $\alpha = 0.50$. In addition, the delays incurred when $\alpha = 0.33$ (for example $\mu = 24, p = 0.3$) are much higher than the corresponding delays when $\alpha = 0.50$ as indicated by the respective colours. A similar effect can be observed for other combinations of these parameters.

5.4.3 The complexity of congestion phenomena

In Chapter 4, we described the transition from traffic jam growth to dispersal in three phases. The transition was induced by removing the original obstruction with no other forms of intervention being installed. This led either to the controlled dispersion of the traffic jam, or to an oscillatory state, in which the traffic jam expanded and contracted over a short period or to a 'failure' phase where the traffic jam did not disperse at all and continued to grow as before. This was due to the closure of the gridlock cycle.

However, further experiments with control strategies has revealed some additional, unexpected outcomes which were not apparent in the aforementioned study. For example, in certain situations, the installation of a single *block* strategy forced the original gridlock cycle apart, but eventually the traffic jam developed around a new gridlock cycle at a nearby location. We refer to this as traffic jam migration. Other forms of traffic jam migration have also been observed using the queueing and re-

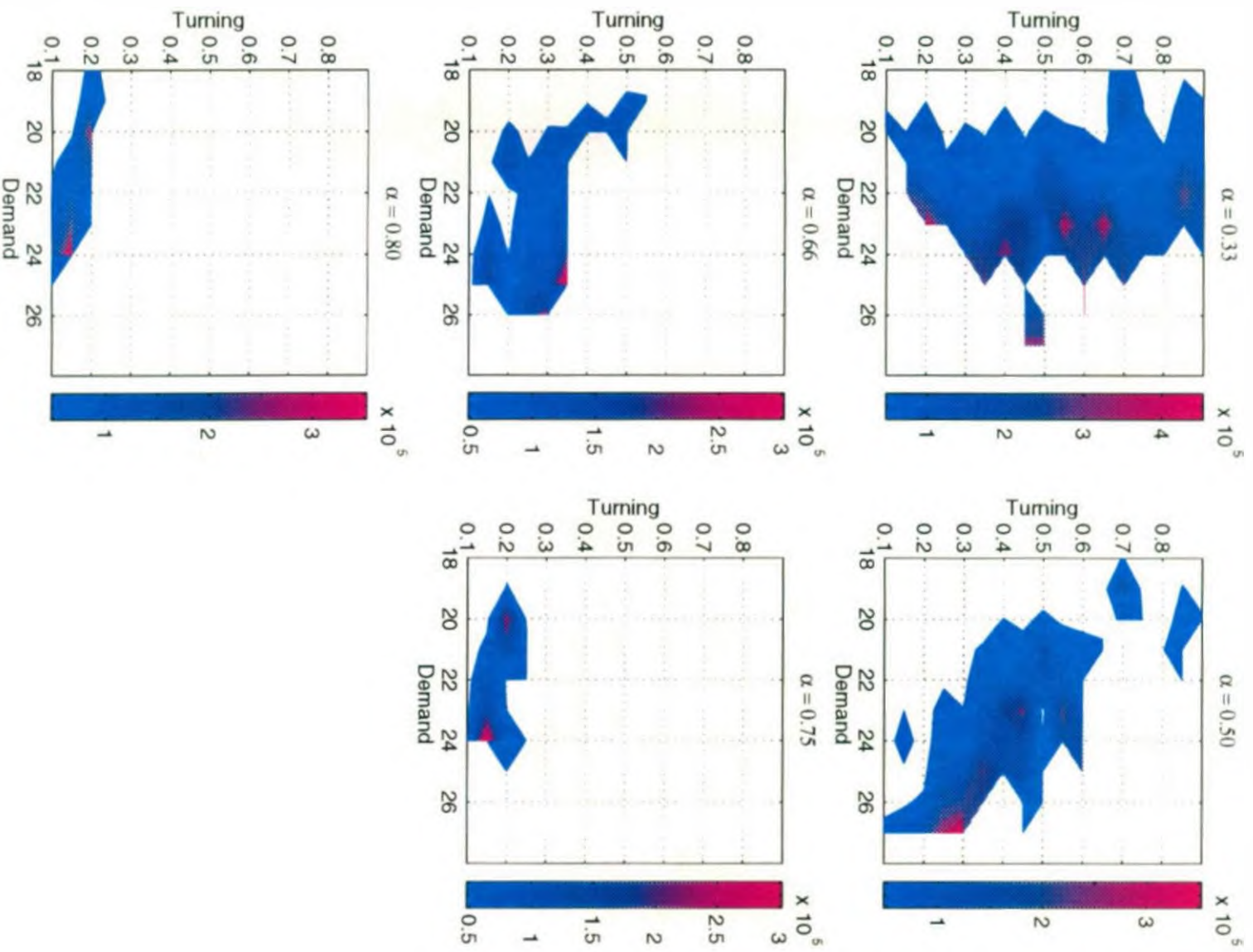


Figure 5.6: Operative domain of the integrated control scheme with respect to demand, turning proportion and stopline widths (Colour indicates total delay measured in veh-
 $\times 10^5$ s)

routing versions of the boundary strategy. These will be described later on.

But unlike the traffic jam growth process, which develops in a predictable way, traffic jam dispersal phenomena are less straightforward. This is due mainly to the variety and unpredictability of dispersal outcomes which may arise. These can be summarised as follows.

- (a) traffic jam clears completely within a few time-slices
- (b) traffic jam expands and contracts at regular intervals. This period of oscillation is usually followed by dispersal or resumed growth. The fluctuations are due to the migration of the original gridlock cycle to a nearby vicinity: in other words the congestion transfers itself and does not always clear.
- (c) traffic jam assumes a static nature - it neither expands nor contracts
- (d) traffic jam migrates to a fresh location
- (e) single jam splits into two mini-jams which:
 - resonate until one dies out followed by the other; or
 - resonate until one grows and other dies out; or
 - resonate until both become gridlocked and extend over the network.
- (f) traffic jam continues to grow regardless of the control measures introduced
- (g) traffic jam develops over network without an obstruction ever having been present (the spontaneous traffic jam)

We have already dealt with categories (a), (b) and (f) in Chapter 4 and subsequently in Appendix B. In the next section, we describe some of the unexpected traffic jam dispersal phenomena (c), (d), (e) and (g) which have arisen during the experiments with the various control strategies. The phenomena are important because they highlight the significance of the online graphical display in analysing congestion patterns.

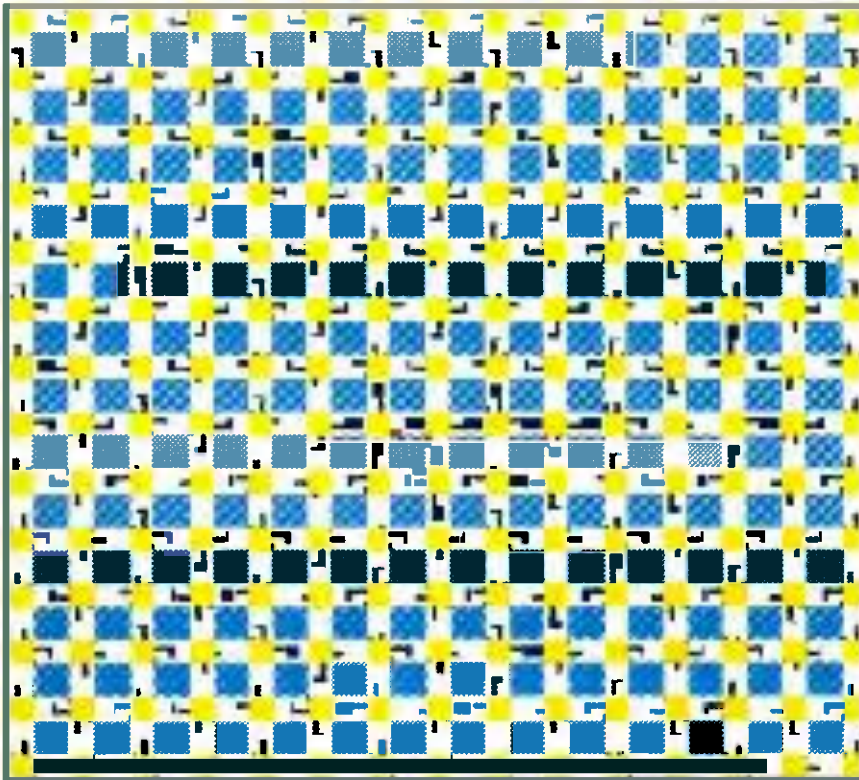


Figure 5.7: *Snapshot of 'static' traffic jam after treatment*

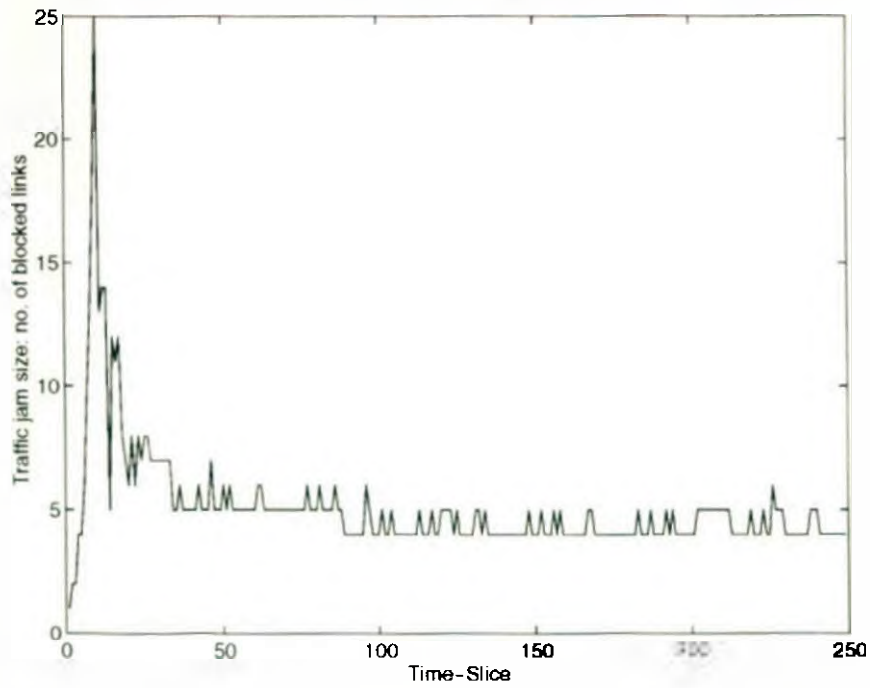


Figure 5.8: *The prolonged dispersal process of a 'static' traffic jam*

We begin with the static traffic jam which is a degenerate case on the boundary between successful dispersion and failure. Although it is not a distinct phenomenon as such, nevertheless, it highlights some interesting points about traffic jam dispersal. Later, we describe the spontaneous traffic jam and discuss various forms of traffic jam migration.

The static traffic jam

The static traffic jam arises in situations where the road network is operating at or close to saturation level. Under these conditions, the application of counter-measures may result in partial success. Traffic jam growth is contained, ie the congestion does not spread over the network, yet, the original traffic jam does not entirely dissolve.

A typical snapshot of a 'static' traffic jam can be seen in Figure 5.7. The traffic jam was simulated with relatively high levels of demand ($\mu = 24,25$), low turning proportions ($p = 0.2$) and with the channelisation parameter set at $\alpha = 0.33$. The implementation of the *integrated* control scheme led to the picture shown in Figure 5.7 which was taken during the fortieth time-slice. Queues can be seen both at the centre of the traffic jam, as well as along the periphery of the control area.

But, as shown in Figure 5.8, there is little variation in the jam size from one time-slice to the next. Although not shown on the graph, the traffic jam had still not cleared completely by the two thousandth time-slice. This demonstrates how slow the dispersal process of static traffic jams may be.

It follows that the delay experienced by drivers trapped in static congestion remains approximately constant, with little variation from one time-slice to the next. However, since the traffic jam never appears to clear completely, the total delay increases without bound each time-slice.

Migrating traffic jams

An essential aspect of good congestion management is the evaluation stage of traffic control. Evaluation takes on two forms: short and long term. In the short term, traffic engineers need to disentangle the traffic queues. However, the situation should also be monitored with a long term perspective. This is because a control strategy might only relieve the situation temporarily, in other words the traffic jam recurs elsewhere in the network.

Figure 5.9 highlights this idea graphically. With no intervention, the traffic jam continues to grow regardless of the fact that the original obstruction has been eliminated. By contrast, applying the single *block* strategy (B_1), forces the original gridlock cycle apart and the traffic jam appears to clear. This accounts for the drop in the jam size until the fifteenth cycle. (In other words, the strategy achieves partial success.) However, the relief is shortlived, as the period of decay is immediately followed by renewed growth. Effectively, the congestion has been transferred to a new gridlock cycle from which the traffic jam evolves.

This shows that reactive control may not always be sufficient when clearing traffic jams. In other words, traffic controllers need to be able to *predict* where the congestion is most likely to go and implement control measures in advance. In this example, the traffic jam is eliminated completely by installing the multiple *block* strategy (B_3). This strategy has the desired effect because it tackles the original gridlock cycle but more importantly, it reduces the possibility of fresh gridlock cycles forming in the vicinity. This is because the mechanism of the *block* strategy releases vehicles into potentially sensitive locations, ie areas where partial gridlock cycles have already developed. The excess traffic can cause these cycles to lock up completely. If this happens, the treatment is rendered useless. Multiple *block* strategies address this problem since they ensure that both the original cycle is eliminated and also, that no new ones are completed as a result of the treatment.

Unfortunately, it is not always easy to reliably predict where the traffic jam is likely to go. Figure 5.10 shows a more complex form of traffic jam migration. Again, a traffic jam was simulated under particular conditions with the demand level set at $\mu = 25$, the turning proportion $p = 0.24$ and $\alpha = 0.75$. This road layout allocates the stopline widths in proportion with the demands. The top part of the figure shows the traffic jam prior to the installation of the *integrated* control scheme. In the bottom figure, the traffic jam has split into two small traffic jams which interact with each other as follows. Whilst one grows, the other contracts. Then the jams reverse roles and the process repeats itself. However, neither traffic jam dominates. Eventually, one of the traffic jams dies out followed by the other, see Figure 5.11.

The behaviour of the paired traffic jams may be explained as follows. The arrangement of the mini-jams on the grid may imply a mutual attraction so that the output from one mini-jam serves as the input to the second mini-jam and vice versa. Neither jam dominates since the input total is balanced by the output total and so the jams expand and contract at regular intervals. This phenomenon seems to represent a special case of the 'oscillatory' phase of traffic jam dispersal.

We have also observed other situations, where this period of mutual oscillation is followed by renewed growth of one or both of the traffic jams. Ultimately, the operation of the network is paralysed by the rapid escalation of queues.

Spontaneous traffic jams

The graphical display of traffic jams has highlighted an interesting phenomenon which arises when $\alpha = 0.33$. This layout allocates one third of the stopline width to the ahead queues and the remainder to the turning queues, and would almost never be used in practice. Nevertheless it gives rise to spontaneous traffic jams which develop over the network without an obstruction ever having been installed. Instead, the conditions for a steady-state system are satisfied, i.e. the total input to the network is balanced by the total output and there is little variation from one time-slice to

the (See Appendix A for more details.) However, local pockets of traffic queues are distributed in a number of locations in the system. These do not dissipate freely as expected (to create a truly isotropic network), but sometimes develop, albeit slowly, often one at a time, until one or more traffic jams have formed and extended over the entire network.

The aftermath of congestion

Simulation experiments have also revealed a *hysteresis* effect which affects the process of dispersal. By this we mean that the road network may be affected by congestion long after it has been cleared. In some cases, the road network returns to its steady state (as defined at the start of the simulation) within a short time period, but in others, it is necessary to wait a long time before the return of 'normal' conditions - where the number of vehicles waiting in the system corresponds to \bar{w} . Thus, although the network conditions may appear to be isotropic, the impact of congestion may still be felt through the excess number of cars (over \bar{w}) traversing the system. A similar effect is described in OECD (1981).

5.4.4 Extending the control principles

The dispersal of traffic jams on one-way grid networks has focussed on the implementation of banning mechanisms at specific locations in the congested region. The research pointed to an *integrated* control strategy for traffic jam elimination. We now consider how the control strategy could be applied to treat gridlock in other types of road network.

Returning to the relationship between a traffic jam's structure and the way it disperses, we define a *friction point* as one where a number of traffic streams converge from more than one approach. These streams are competing for the vacant road space in the same exit link. Examples of friction points in one-way and two-way grid networks are shown in Figure 5.12.

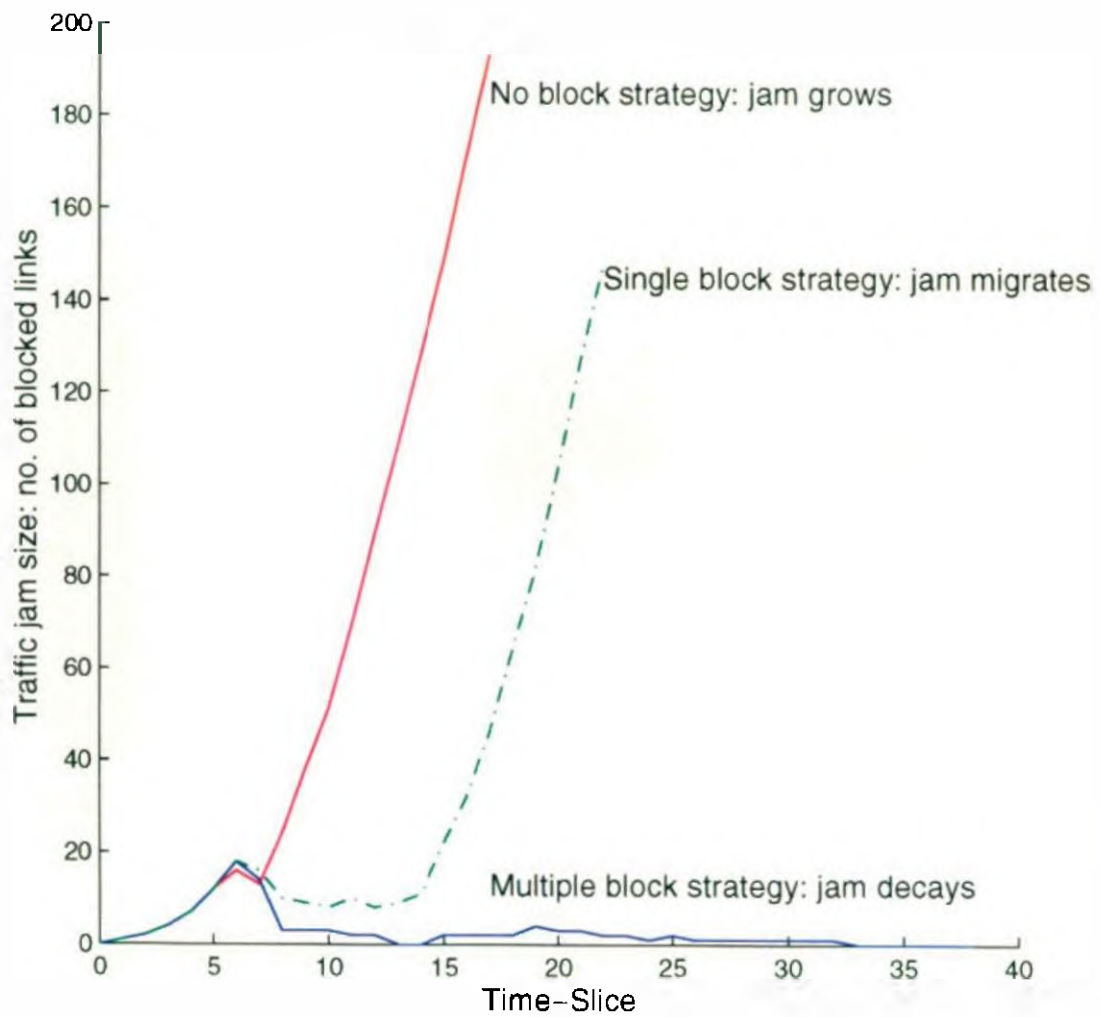
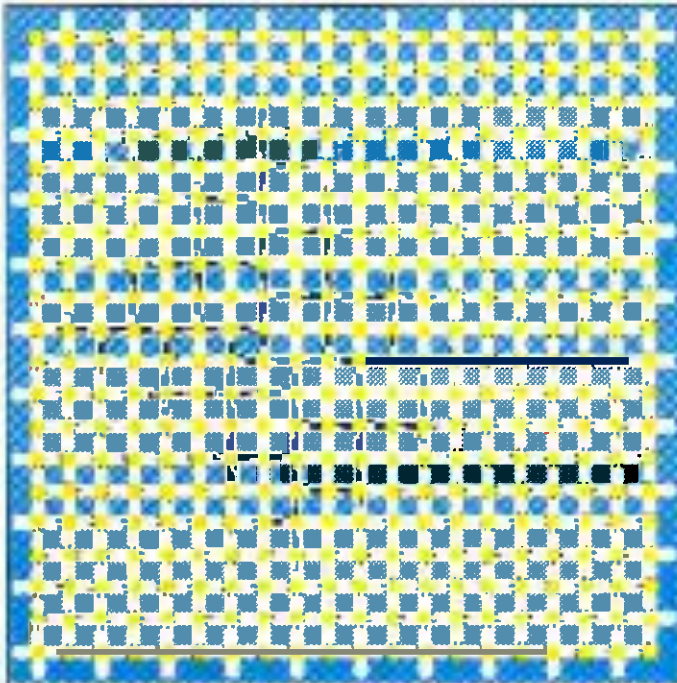
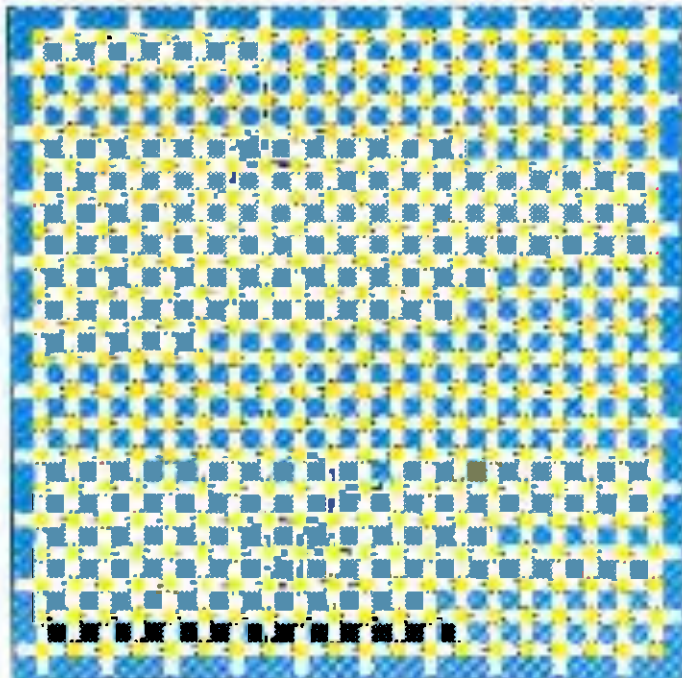


Figure 5.9: *Traffic jam migration*



Traffic jam before treatment (Cycle = 13)



Traffic jam after treatment (Cycle = 26)

Figure 5.10: *Traffic jam splits in two parts following the implementation of the integrated control scheme*

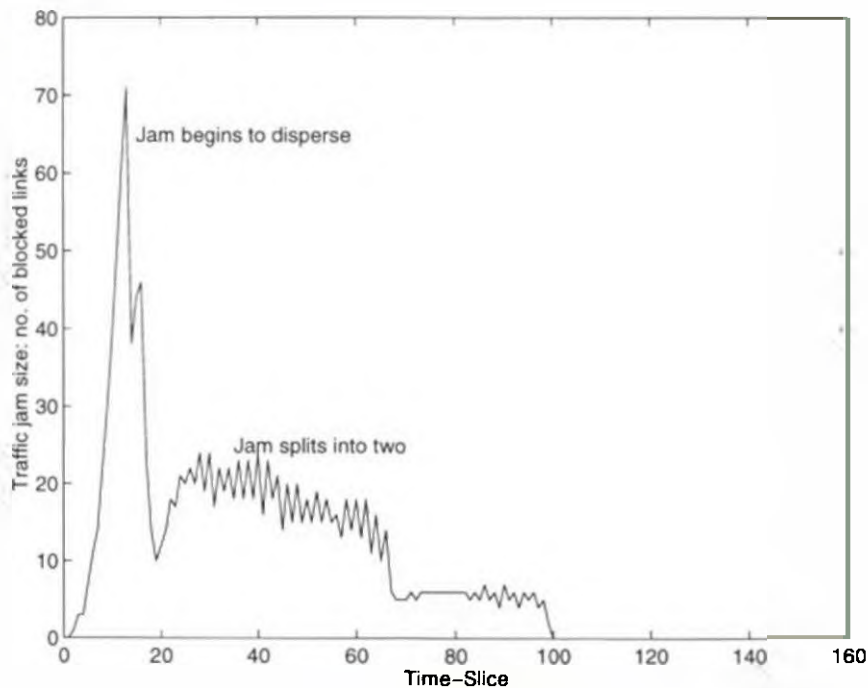


Figure 5.11: *Decay process of resonating traffic jam*

In order to re-start vehicle movement, it is necessary to separate out the converging streams otherwise they themselves lead to fresh gridlock cycles. Each *friction* point which forms may be viewed as an obstruction whose severity depends on the particular traffic conditions. It follows that each *friction* point represents a potential location for a gridlock cycle, and, that the number of *friction* points increases proportionately with the size of the traffic jam.

This explains the operation of the *block* strategy in one-way grid networks. Imposing a turn ban on link **L2** in Figure 5.12 forces the queueing vehicles to go ahead. This then allows the ahead queue in **L3** to dissipate using the spare capacity created in **L1**. This idea is applied to each of the four interlocking turns which comprise the gridlock cycle, and eventually leads to the dispersal of the traffic jam.

A similar approach can be applied to clear traffic jams in two-way rectangular grid networks and also for networks which do not possess a grid topology. These cases are more complicated because there are *more* than two streams of traffic converging on

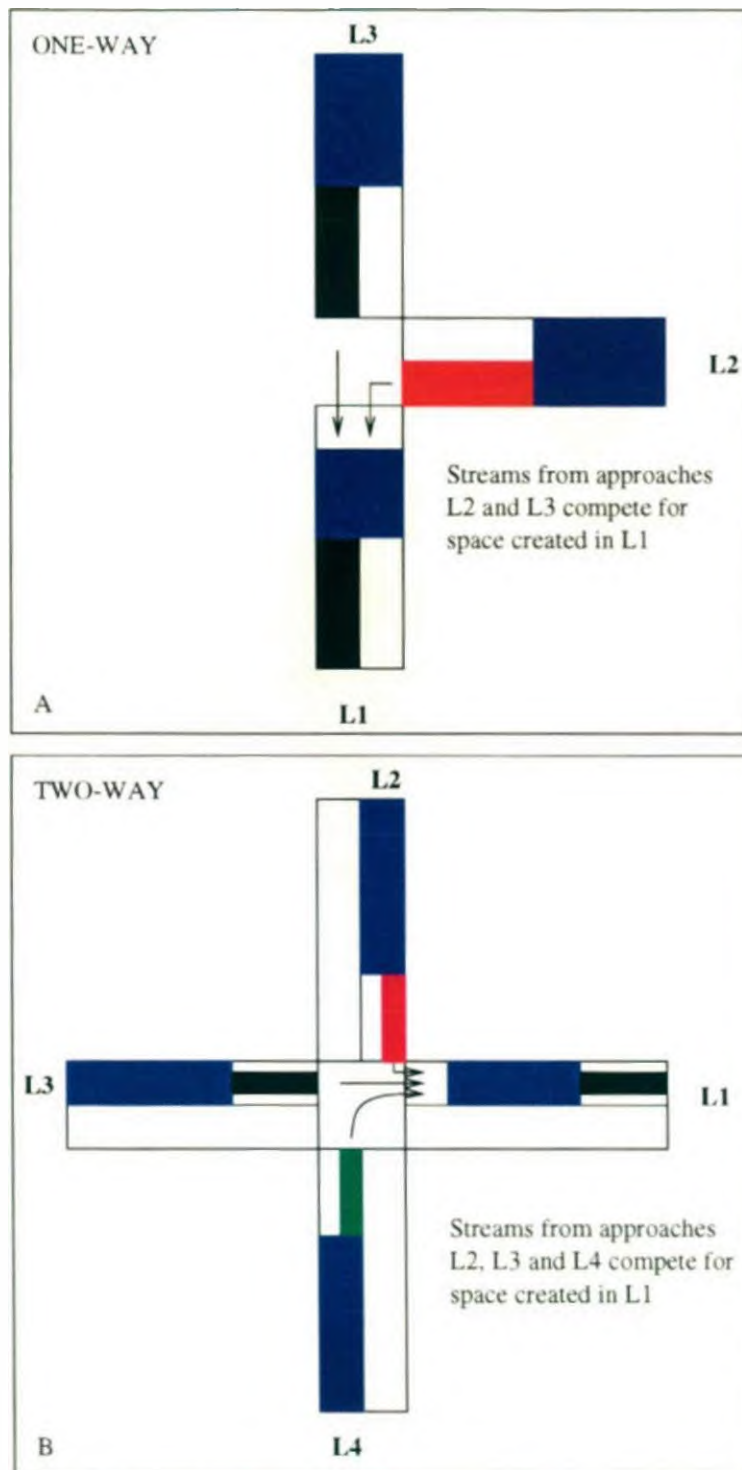


Figure 5.12: Friction points in grid networks

each *friction* point. In the one-way network, each *friction* point involves streams of traffic arriving from two approaches (ahead and either of left turning or of right turning vehicles, but not both). With the two-way network, there are three approaches to consider: ahead, left turning and right turning. All three streams compete simultaneously for the vacant road space in the same exit link.

In addition, the one-way traffic jam evolves from a single gridlock cycle (either of left turning or of right turning vehicles, but not both). With the two-way network, jam development focusses on *two* gridlock cycles: one of left turning vehicles and one of right-turning vehicles. The increase in the potential number of *friction* points and the number of competing streams increases the likelihood of gridlock formation, and makes jam dispersal more difficult.

However, regardless of the underlying network structure, the principles of control remain similar, namely, treatment of the traffic jam internally whilst simultaneously reducing the incoming flow into the congested region. The implementation of the control techniques still requires further research since these would need to be tailored to suit the traffic jam under consideration. Nevertheless, the mechanisms summarised in Table 5.3 do yield qualitative insights which can be used to develop a set of guidelines for clearing gridlock-based traffic jams.

5.5 Conclusions

This chapter began with a survey of various control measures which had been developed to tackle urban congestion. The strategies were classified in three groups: strategies which prevented the onset of congestion, strategies which alleviated congestion once it had occurred, and methods which could lead to an expeditious recovery from gridlock or severe congestion. However, the study found that whilst most traffic congestion control measures which are currently being used could increase capacity and hence postpone the onset of severe congestion, they were not suitable once

Detection	
1	identify and remove source of obstruction
2	locate any gridlock cycles
3	determine the extent of their development
Treatment	
4	separate any interlocking turning movements (gridlock cycles) using <i>ahead</i> only restrictions
5	define a cordon which encloses the currently developed traffic jam
6	reduce the incoming flow into the congested region using an appropriate system of gates or diversion mechanisms
7	identify all potential <i>friction</i> points
8	tackle <i>friction</i> points one by one from the centre of the traffic jam outwards using mechanism similar to criterion 4
Evaluation	
9	remove the cordon of gates once congestion at centre dissolves
10	remove internal restrictions, and continue to monitor events until system has returned to original free-flow state

Table 5.3: *Guidelines for clearing gridlock-based traffic jams*

gridlock had set in.

The recovery from severe congestion (gridlock) was considered in three stages, each of which led to the next. The stages were detection, treatment and evaluation. Detection included establishing the type of congestion, determining its severity and ascertaining whether gridlock had occurred. This represents an important factor in the control process because, whilst many existing traffic systems can detect the formation of queues at individual or paired intersections, few were capable of recognising patterns or chains of queues, and hence were deficient in treating gridlock phenomena.

The detection stage was followed by the implementation of appropriate control measures. This was known as the *treatment* stage. The performance of the measures was subsequently evaluated both in the short and long term.

The treatment of traffic jams focussed on the installation of bans at specific network locations. The bans came in two forms: turn or ahead. Turn bans were imposed on selected links to break gridlock cycles at the nucleus of the traffic jam. This was known as *nucleus* intervention and was implemented via the *block* strategy. By contrast, ahead bans were implemented around the traffic jam envelope to reduce input into critical sections of the road. These ahead bans were used in two ways. Vehicles using the banned junctions were either queued outside the congested region or else re-routed away from the jam. This type of control was referred to as *boundary* intervention.

A selection of dynamic strategies involving various arrangements of bans was developed for each type of intervention. These included the *block* strategy which was applied to fragment gridlock cycles that had formed at the jam centre, and the *envelope* strategy which protected the congested region from excessive demand. The simulation tool described in Chapter 3 was used to examine the individual and combined performance of these measures under a range of network conditions. This led to the formulation of *integrated* control scheme for the elimination of traffic jams in

one-way traffic management strategies. Simulation experiments involving this and other strategies found that dispersal was more difficult under increasing levels of demand and turning proportion and that it was best to allocate equal stopline widths between ahead and turning vehicles. grid networks.

The results are still at an early stage and are not immediately applicable in practice. Nevertheless, they do yield some qualitative insights into the way gridlock-based traffic jams disperse and it is hoped that these could be incorporated into existing and future traffic management strategies.

Whilst in certain situations counter measures dispersed the traffic jam, they didn't always work. In some cases, the traffic jam appeared to clear but later on, it recurred elsewhere in the network (jam migration). The complexity of such phenomena highlighted the importance of monitoring the effects of control on a local level whilst simultaneously maintaining a long term perspective. In addition, *reactive* control was sometimes insufficient for treating traffic jam migration problems, whereas advance implementation of control measures at potentially sensitive locations could sometimes help.

Building on the results of the study of one-way traffic jams, the chapter concluded with a brief discussion of how the control ideas could be extended to treat traffic jams in other types of road network. The principles of control were to remain similar, namely, treatment of the internal parts of the traffic jam whilst simultaneously reducing the flow into the congested region. However, the mechanics of control would need to be tailored to suit the specific requirements of the traffic jam under consideration. This topic will form part of future work, dealt with in the final chapter.

Chapter 6

Conclusions

The aim of this thesis was to investigate the characteristics of traffic jams caused by incidents in rectangular grid networks and to suggest effective control mechanisms which would aid their dispersal. In this final chapter, the principal elements of traffic jam growth and decay are summarised; some limitations and possible extensions of the research are then discussed.

6.1 Traffic jam development and dispersal

The characteristics of area-wide traffic jams (as opposed to queues on isolated sections of road) have been investigated using a combination of theoretical and simulation models which have been applied to one-way and two-way rectangular grid networks. This enabled us to identify those features of traffic jams which would apply independently of the underlying network configuration.

The approach was first to develop a framework which explained the structure of traffic jams and use this to formulate control strategies for their successful elimination. The main results are summarised below.

6.1.1 Fundamental concepts

The behaviour of traffic jam phenomena was described in two phases: development and dispersal. The development of a traffic jam was initiated by creating an obstruction towards the centre of the network. This led to one or more queues which were effectively *independent*. Subsequently, these branched out to cause interference with upstream links, and eventually, the queues encircled the cause of the obstruction leading to gridlock.

This was identified as an important stage in terms of jam development because once this threshold had been reached, it became increasingly more difficult to disperse the resultant traffic jam. Furthermore, it was shown that it was best to postpone gridlock for as long as possible in order to maximise the chances of achieving total dispersal. However, there appeared to be a direct conflict between the requirement to maximise the time to gridlock, and the requirement to minimise the overall growth rate of the traffic jam and hence reduce delays to drivers.

The dispersal phase was activated by removing the original obstruction. This led to one of three distinct phases. These were (a) traffic jam disperses at varying rates, (b) traffic jam growth accelerates and decelerates at regular intervals and (c) traffic jam fails to clear.

A secondary characteristic of traffic jams was the *starvation* mechanism which interacted with the primary stages of growth and decay. Starvation occurred in the internal regions of the traffic jam (queue starvation) as well as along its boundary (boundary starvation).

The overall pattern of traffic jam development was the same for one-way and two-way grid networks. The pattern was affected, amongst others, by the queue propagation times associated with the ahead and turning queues. Whilst the theory pointed to two possible spatial configurations for traffic jams on one-way grid networks (Type I and II), simulation has highlighted a broader spectrum of structures. The research

found that while Type II traffic jams developed at a slower rate than Type I traffic jams, the latter type of jam was easier to disperse. This was mainly due to the queue structure of the jam. Type II traffic jams were compact and developed in tight gridlock cycles and queue starvation effects were less noticeable. Type I traffic jams, on the other hand, were more spread out and incorporated a substantial proportion of 'anti-queues' at specific locations. These played an important role in the jam dispersal process.

But some differences between one-way and two-way traffic jams were observed, for example, traffic jam development on a one-way network emanated from a single loop of locked vehicles, whereas with the two-way network, the development focussed on two such loops which developed in parallel. The problem of gridlock in a two-way grid network was therefore potentially more severe than in the one-way network. Also, on the two-way network, the size of the right-turn storage areas (known as right-turn bins) was critical where traffic jam dispersal was concerned, although this was not the case with the rate of traffic jam growth.

In addition, when a link blocks back to the upstream junction, a point of friction is created between the streams of traffic requiring this junction. In a one-way network, the friction is created between two streams: ahead traffic and one of left or right turning traffic. With the two-way network, there are three such streams (ahead, left and right) competing at each blocked junction. This makes the dispersion of two-way traffic jams more complicated.

Two approaches for controlling traffic queues were outlined. These were referred to as 'static' or 'dynamic' in nature. A selection of traffic management measures which could be implemented as part of a 'static' control scheme were examined. However, these were limited since they could not be applied once gridlock had already set in. This led to the development of a range of responsive measures for clearing gridlock-based traffic jams.

6.1.2 Control principles

The past few decades have seen the development of a selection of control measures to treat urban congestion problems. These included strategies which prevented the onset of congestion, strategies which alleviated congestion once it had occurred and methods which could be applied once gridlock had already set in. This research has been concerned mainly with the development and testing of various queue control strategies for protecting networks from gridlock and dissipating traffic jams once they have occurred.

Three stages, each one leading to the next, formed the basis for the control of traffic jams affected by gridlock. These were detection, treatment and evaluation. Detection included establishing the cause of congestion, determining its severity and ascertaining whether gridlock had occurred. This was an important part of the control process because, whilst many existing traffic systems were capable of detecting queues at individual or at paired intersections, few could recognise patterns or chains of queues, and hence were deficient in representing gridlock phenomena.

Following the detection stage, various control measures were implemented. This was known as the *treatment* stage, and was usually followed by an *evaluation* stage in which the impact of the measures was assessed in the short and long term.

The treatment of traffic jams in one-way grid networks focussed on the installation of bans at specific network locations. The bans came in two forms: turn or ahead. Turn bans were imposed on selected links to break gridlock cycles at the nucleus of the traffic jam. This was known as *nucleus* intervention and was implemented using the *block* strategy. By contrast, ahead bans were implemented around the traffic jam envelope to reduce input into critical sections of the road. These ahead bans were used in two ways. Vehicles using the banned junctions were either queued outside the congested region or else re-routed away from the jam. This type of control was referred to as *boundary* intervention.

A selection of dynamic strategies involving various arrangements of bans was developed for each type of intervention. These included the *block* strategy which was applied to fragment gridlock cycles that had formed at the jam centre, and the *envelope* strategy which protected the congested region from excessive demand. The simulation tool described in Chapter 3 was used to examine the individual and combined performance of these measures under a range of network conditions. In particular, the study found that although *boundary* intervention was useful in that it improved the quality of traffic flow within the congested region, it did not help if gridlock had already occurred.

This led to the formulation of the *integrated* control scheme for the elimination of traffic jams in one-way grid networks. The integrated control scheme combined two control plans to provide a comprehensive strategy for clearing one-way traffic jams. The strategy involved the application of turn bans at the jam centre to treat gridlock cycles that had formed there, together with a cordon of ahead bans which were implemented along the jam boundary, to reduce flow into the congested area. Simulation experiments involving this and other strategies found that dispersal was more difficult under increasing levels of demand and turning proportion and that it was best to allocate the stopline widths between ahead and turning vehicles in proportion to the respective turning movements.

Whilst in certain situations the proposed control mechanisms dispersed the traffic jam, they didn't always work. This was mainly due to the mismatch between the selected strategy and the traffic jam structure being cleared. In some cases, the traffic jam appeared to clear but later on, it recurred elsewhere in the network (jam migration). The complexity of such phenomena highlighted the importance of monitoring the effects of control on a local level whilst simultaneously maintaining a long term perspective. In addition, *reactive* control was sometimes insufficient for treating traffic jam migration problems, whereas advance implementation of control measures such as *block* strategies at potentially sensitive locations could sometimes

help.

Building on the results of the study of one-way traffic jams, we then considered how the control principles developed for one-way traffic jams could be applied to tackle traffic jams in other types of road network. This could be achieved in theory by treating the internal parts of the traffic jam whilst simultaneously reducing the flow into the congested region. However, the practical implementation of these principles would need to be tailored according to the traffic jam under consideration. This represents a future goal for this research.

6.2 Possible extensions of the research

The research in this thesis was based on the formulation of a working theory which explains the evolution and decay of traffic jams in idealised rectangular grid networks. Whilst the approach sacrificed a degree of realism and the results could not be applied immediately to real networks, nevertheless, they yielded qualitative insights concerning the congestion problem in general terms. In particular, they predicted certain features of jam behaviour that were not accessible in any other way.

In the next section, some possible extensions of the research are discussed. These can be divided into two main areas: theoretical studies concerning the role of gridlock and practical developments to extend the results to real networks.

6.2.1 Theoretical studies

Predicting the likelihood of gridlock

The formation of gridlock was identified as an important stage in terms of jam development because once this threshold had been reached, it became increasingly more difficult to disperse the resultant traffic jam. There are a number of approaches which can be adopted in order to ascertain or predict the likelihood of gridlock occurring. These include (a) spatial analysis of congestion patterns (b) statistical surveys or

(c) specific tests such as the one developed using the simulation tool, in which, if the traffic jam continued to grow regardless of the original obstruction having been removed, then gridlock was deemed to have set in.

However, the latter technique would only be feasible within a simulated environment and could not be implemented in practical situations. This highlights the need for a set of criteria which could be used to help traffic engineers predict and decide whether gridlock had occurred. The expert system methodology developed by Scemama (1995) points in this direction because it uses a knowledge base of rules and facts to compose 'congestion chains' via inference and deductive reasoning. By combining this with the spatial analysis of congestion patterns, it may be possible to establish whether a gridlock had formed at a particular location.

But traffic engineers would also want to be able to predict the likelihood of gridlock. By this we mean that it would be useful if, given particular input conditions such as traffic demands and turning proportions, one could develop a set of guidelines which would determine whether gridlock was likely to occur and if so, how severe the consequences would be. This is a difficult problem because gridlocks of the type considered in this research are comparatively rare events and there is little relevant data available. Even so, the data described by Allsop and Bell (1995) might offer scope for a preliminary analysis of this type.

Comparisons between one-way and two-way grid networks

The development of control strategies for traffic jams in one-way and two-way grid networks raises the question as to which network configuration is to be preferred where the objective is to minimise traffic jam development and inhibit gridlock.

Our earlier analysis suggests that there is no simple solution since there are obvious differences between the two types of network. For example, with the two-way network there is the complication of right-turning storage areas and the vehicles associated with them.

Nevertheless, the models developed in the context of this research could be used to provide a partial solution to this problem. For example, we have seen that the growth of a two-way traffic jam emanates from two loops of locked vehicles which form simultaneously whereas one-way traffic jam growth focusses on a single loop. Consequently, the problem of gridlock in a two-way network is potentially more severe than in the one-way network.

Building on these ideas, two comparable systems could be set up with the one being a two-way grid and the other a one-way grid. However, the problem of comparability should not be underestimated and the results of the study will be restricted by the idealised nature of the modelling tools being used for the analysis.

6.2.2 Practical developments

A number of simulation tools, mostly developed over the past few decades, exist for the formulation, verification and implementation of traffic management strategies. However, none of these models can fully meet the requirements of simulating large-scale traffic networks in faster than real time. At the operational level, the models are currently undergoing extensive enhancement to incorporate state-of-the-art technologies such as mainline traffic control, real-time route guidance and incident management.

In order to extend some of the recommendations of the thesis to real urban networks, an interface will need to be developed between one of these commercially available, state-of-the-art models and the simulation tool described in Chapter 3. One possibility is the MITSIM model currently being developed in the USA, Yang and Koutsopoulos (1996). This model is suitable for modelling dynamic traffic networks and incorporates functionalities such as advance traffic control, route guidance and incident detection systems. The model satisfies many of the congestion modelling criteria described in Chapter 2 of this thesis. For example, MITSIM represents networks and queues in detail and is fully microscopic. It includes a graphical user interface for

animating vehicle movements and can be used to develop and test general control strategies. Furthermore, it possesses a flexible, modular software structure which can be adapted to suit specific needs.

Thus, the limitations imposed by the idealised networks used in our research will need to be modified in order to cope with traffic conditions in modern urban networks. For example, so far, in our simulation tool, origins and destinations have not been specified for individual vehicles. This will need to be incorporated in the simulation model so that variable turning proportions may be calculated for all the links comprising the network. It will also be necessary to develop the model so that it could handle more realistic features such as variable link lengths and types, pedestrian interaction, on and off-street parking facilities and driver behaviour information.

However, as with many commercially available models, the network size and the number of vehicles allowed in MITSIM are limited by the available computer memory, and the running time increases proportionately with increases in network size and the number of existing vehicles. This problem will need to be addressed in order to model the growth and decay of area-wide traffic jams. Assuming this can be overcome, the resultant traffic simulator could then be used (a) to determine how queues propagate and disperse in irregular network types and (b) to implement strategies for breaking the resultant gridlocks which may subsequently form.

Although the principles will remain the same, nevertheless, there are likely to be some differences in detail. In particular, the envelope of the traffic queues is likely to depend on the position of the original obstruction and the traffic flows there, and the traffic locking mechanism may be more complicated than the simple patterns simulated in this project.

In addition, the problem of re-routing of vehicles deflected away from the congested region will need to be addressed specifically. Ultimately, it is envisaged that the simulator could be used to help select appropriate control strategies to disperse incident-

induced traffic congestion on generic network types.

Bibliography

ABBESS C R, WRIGHT C C and HUDDART K W (1991). "Controlling area-wide traffic congestion", *Final report to the Rees Jeffreys Road Fund, Middlex. Polytechnic Consultancy Services Organisation*, (unp).

ABBESS C R and ROBERG P (1995). "Fractal structure of traffic jam images", *Proceedings IMA Conference: Complex Stochastic Systems and Engineering* (TITTERINGTON D M, Editor), **54**, 251-281.

ALLSOP R and BELL M (1995). "Loss of capacity at signal-controlled junctions due to blocking back by downstream queues", *Report to EPSRC on Research Grants GR/J 43042 and 44681*, Nov 1995.

BATTY M, LONGLEY P and FOTHERINGHAM S (1989). "Urban growth and form: scaling, fractal geometry, and diffusion-limited aggregation", *Environment and Planning*, **21A**, 1447-1472.

BATTY M and LONGLEY P (1994). "Fractal Cities: A geometry of form and function", Academic, San Diego, 1994.

BIHAM O and MIDDLETON A A (1992). "Self organisation and a dynamical transition in traffic flow models", *Physical Review*, **46A**(10), 6124-6127.

DAGANZOS C F (1994). "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory", *Transportation Research*, **28B**(4), 269-287.

DAGANZOS C F (1995). "The cell transmission model, Part II: Network traffic", *Transportation Research*, **29B**(2), 79-93.

DAVIES R and O'KEEFE R (1989) *Simulation modelling with Pascal*, Prentice Hall, Englewood, N.J.

DEPARTMENT OF THE ENVIRONMENT (1976). "Bus Demonstration Project", *Summary Report No. 8*, London, UK, 1976.

DEPARTMENT OF TRANSPORT (1992). *Transport Statistics Great Britain 1992* HMSO.

EDIE L C (1974). "Flow theories", *Traffic Science*, (GAZIS D C, Editor), 1-109, New York, Wiley 1974.

FISHMAN G S (1973) *Concepts and Methods in Discrete Event Simulation*, em John Wiley, NY

FORASTE B and SCEMAMA G Inrets (1987). "An expert system approach to congestion", *Recherche Transports Sécurité*, English issue, No 2, July 1987, (Translation from RTS September 1986).

FRANCERIES C, *Un centre ville sans embouteillage - Une realite a Bordeaux*, Paris, France: Revue Generale des Routes et des Aerodromes, 1977.

GAFARIAN A V, ANCKER C J, MORISAKU T (1978) Evaluation of commonly used rules for detecting steady-state in computer simulation. *Naval Res. Log. Quart.* **25**, 511.

GAZIS D C, HERMAN R and POTTS R B (1959). "Car Following theory of steady-state flow", *Operations Research* **7**, 499.

GLASS C J and NI L M (1994). "The turn model for adaptive routing", *Journal of the Association for Computing Machinery*, **41**(5), 874-902.

GRAY B M and IBBETSON L (1991). "Operational traffic control strategies for congestion", *Traffic Engineering and Control*, **32**(2), 60-66.

HÉRAN F and TOSTAIN P (1995). "A comparison between the development of the principles of the organisation of industrial production and of the design of urban road

systems", *Recherche Transports Sécurité*, English issue, No 11, 41-54, Sept 1995.

HOLDEN H and RISEBRO N H (1995). "A mathematical model of traffic flow on a network of unidirectional roads", *SIAM Journal of Mathematical Analysis*, **26**(4), 999-1017.

HOLMES B (1994). "When shock waves hit traffic", *New Scientist*, 25 June, 36-40.

HOLMES G N and LEONARD D R (1992). "The frequency and importance of incidents which cause congestion in urban areas", *TRL Contractor Report*, **342**.

HUDDART K W and WRIGHT C C (1989). "Catastrophic congestion and some possible ways of preventing it", *Proceedings TRAFFEX Seminar on 'Congestion, control and parking enforcement'*, (Brighton, April 1989), PTRC Education and Research Services Ltd, London, 1989.

HURDLE R (1992). "A theory of traffic flow for congested conditions on urban arterial streets II: Four illustrative examples", *Transportation Research*, **26B**(5), 397-415.

JOHNSON B A and AKCELIK R (1994). "Review of analytical software for applicability to paired intersections", *Proceedings of the Second International Symposium on Highway Capacity*, (AKCELIK R, Editor) Vol **2**, 347-367.

KENDALL M G and GIBBONS J D (1990). "Rank Correlation Methods", 5th Ed. Edward Arnold, London

KENDALL M G, STUART A and ORD J K (1983). "The advanced theory of statistics", Vol **3** (4th Ed). Charles Griffin, London.

KENDALL M G and ORD J K (1990). "Time-Series", 3rd Ed. pp 18-26, Edward Arnold, London.

LEONARD D R, GOWER P and TAYLOR N B (1989). "CONTRAM: The structure

of the model", *TRRL Research Report* **178**.

LEONARD D R (1992). "The meaning of the term 'congestion' in urban areas", *TRL Working Paper*, **WP/TD/38**.

LIGHTHILL M J and WHITHAM G B (1955). "On kinematic waves: II. A theory of traffic on long crowded roads", *Proceedings Royal Society*, **A229**, 318.

LINDLEY J A (1987). "A methodology for quantifying urban freeway congestion", *Transportation Research Record*, **1132**, 1-12.

LIU G (1991). "Congestion, flow and capacity", *Proceedings of the first symposium on Highway capacity and level of service*, (BRANNOLTE, Editor), Balkema, Rotterdam, 245-251.

LOGIE D M W (1979). "TRAFFICQ: a comprehensive model for traffic management schemes", *Traffic Engineering and Control*, **11**, 516-518.

LONGLEY D A (1968). "A control strategy for computer controlled traffic networks", *Transportation Research* **2**, 391-408.

MANDELBROT B B (1983). "The fractal geometry of nature", W H Freeman, New York, N.Y., 1983.

MASEK J G and TURCOTT D L (1983). "A diffusion limited aggregation model for the evolution of drainage networks", *Earth to Planetary Science Letters*, **119**(3), 379-386.

MAURO V (1991). "Evaluation of dynamic network control: Simulation results using NEMIS urban microsimulator", *presented at TRB annual meeting*, Washington DC, January 1991.

MAY A D, PHIU-NUAL K, MONTGOMERY F O, PAKSARSAWAN S and VAN VUREN T (1993). "Application of SATURN in Bangkok", *Traffic Engineering and*

Control, **34**(1), 20-27.

MEAKIN P (1983). "Cluster growth processes on a two dimensional lattice", *Physical Review*, **28B**(12), 6718-6732.

MOGRIDGE M J H (1988). "Jams and superjams: a systems approach to congested traffic networks", *Civil Engineering Systems*, **5**, 181-193.

MOGRIDGE M J H (1990). "Travel in towns: jam yesterday, jam today and jam tomorrow ?", Macmillan, London 1990.

MOLERA J M, CUSETA J A, MARTINEZ F C, and SANCHEZ A (1993). "Phase transitions in two-dimensional traffic flow models, *Physical Review*, **48E**(6), 4175-4178.

MONGEOT H (1993). "Predicting the effects of traffic incidents in urban networks, *MSc Thesis*, University of Southampton.

MORSE D R, LAWTON J H, DODSON M M and WILLIAMSON M H (1985). "Fractal dimension of vegetation and the distribution of anthropod body lengths", *Nature*, **314**(6013), 731-735.

NEELAMKAVIL F (1987). "Computer simulation and modelling", Wiley, GB, 1987.

NEWELL G F (1971). "Applications of queueing theory: Monographs on applied probability and statistics", Chapman and Hall, 1971.

NEWELL G F (1993). A simplified theory of kinematic waves: I: general theory, II: queueing at freeway bottlenecks, III: Multidestination flows, *Transportation Research*, **27B**(4), 281-314.

OECD (1978). "Integrated Urban Traffic Management", OECD Report, Paris.

OECD (1981). "Traffic control in saturated conditions", OECD Report, Paris.

PAKSARSAWAN S (1994). "The development of queueing simulation procedures for

traffic in Bangkok", *PHD thesis*, ITS, University of Leeds.

PIGNATARO L J, MCSHANE W R, CROWLEY K W et al, (1978). "Traffic control in oversaturated street networks", *NCHRP Report 194*, Transportation Research Board, Washington DC, 1978.

PROSSER N and DUNNE M (1994). "A procedure for estimating movement capacities at signalised paired intersections, *Proceedings of the Second International Symposium on Highway Capacity*, (AKCELIK R, Editor), Vol 2, 473-492.

QUINN D J (1992). "A review of queue management strategies", *Traffic Engineering and Control*, **33**(11), 600-605.

RATHI A K (1988). "Development and evaluation of a congestion based control scheme for high traffic density sectors", *Transportation Research*, **22B**(2), 81-101.

RATHI A K (1991). "Traffic metering: an effectiveness study", *Transportation Science*, July, 421-440.

RATHI A K and SANTIAGO A J, (1990). "The new NETSIM program", *Traffic Engineering and Control*, **31**(5), 317-320.

RICHARDS P I (1956). "Shock waves on the highway", *Operations Research* 4, 193.

ROUPHAIL N M and AKCELIK R (1992). "A preliminary model of queue interaction at signalised paired intersections" *Proceedings of the 16th ARRB Conference*, Vol 5, 325-345, Australian Road Research Board Ltd., Vermont South, Victoria, 1992.

RUFFORD N (1987). "London heading for a superjam", *New Civil Engineer*, 6 August, 28-29.

SCEMAMA G (1995). "CLAIRE: An independent, AI-based supervisor for congestion management", *Traffic Engineering and Control*, **36**(11), 604-612.

SHEFFI Y (1985). "Urban transportation networks: Equilibrium analysis with math-

emational programming methods", *Prentice Hall*, New Jersey, 1985.

SHEPHERD S P (1994). "Traffic responsive control in over-saturated conditions using state-space methods", *PHD thesis*, ITS, University of Leeds.

SOUTHAMPTON UNIVERSITY (1992). "CONTRAMI: Modelling the effects of incidents in urban networks", *Final report by the TRG to the Transport Research Laboratory*, February.

VAN VLIET D et al (1982). "SATURN - a modern assignment model", *Traffic Engineering and Control*, **23**(4), 578-581.

VAN VUREN T and LEONARD D R (1994). "Urban congestion caused by incidents", *Traffic Engineering and Control*, **35**(7/8), 422-433.

VAUGHAN R, HURDLE V F and HAUER E (1984). "A traffic flow model with time dependent O-D patterns", *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*, (VOLMULLER J and HAMMERSLAG R, Editors), 155-178, VNU Science Press, Utrecht, 1984.

VAUGHAN R and HURDLE V F (1992). "A theory of traffic flow for congested conditions on urban arterial streets I: theoretical development", *Transportation Research*, **26B**(5), 381-396.

VINCENT R A and LAYFIELD R E (1977). "Nottingham Zones-and-Collar-Study - overall assessment", *Laboratory Report LR805*, Transport Road Research Laboratory, Crowthorne, 1977.

WARDROP J G (1952). "Some theoretical aspects of road traffic research", *Proceedings of the institution of Civil Engineers*, **1**(2), 325-378.

WHITE C, TAYLOR N and HOUNSELL N B (1994). "CONTRAM- A computer suite for modelling road congestion", *Traffic Technology International '94* UK and International Press, 106-111.

- WILSON R E (1995). "Mathematical modelling of road traffic networks", *PHD thesis*, University of Oxford, (to appear).
- WILLIAMS R J (1996a). "Some recent developments for queueing networks", *Probability Towards 2000*, (ACCARDI L and HEYDE C C, Editors) Springer, 1996, (to appear).
- WILLIAMS R J (1996b). "On the approximation of queueing networks in heavy traffic", *Stochastic networks: theory and applications*, (KELLY F P, ZACHARY S, ZIEDINS I, Editors), Oxford, 35-36.
- WITTEN T A and SANDER L A (1981). "Diffusion Limited Aggregation: a kinetic critical phenomenon", *Phys. Rev. Lett.* **47**, 1400-1403.
- WRIGHT C C and HUDDART K W (1989a), "Breaking the grip of gridlock", *Surveyor*, **172**(5052), 16-17.
- WRIGHT C C and HUDDART K W (1989b), "Strategies for urban traffic control". 'Transport and Society' Discussion Paper 6, Rees Jeffreys Road Fund, (unp).
- WRIGHT C C and ROBERG P (1998). "The conceptual structure of traffic jams", *Transportation Policy*, to appear.
- YANG Q and KOUTSOPOULOS H N (1996). "A Microscopic Traffic Simulator for Evaluation of Dynamic Management Systems", (pre-print).

Published and Conference Papers

ROBERG P (1994). "Development and dispersal of area-wide traffic jams", *Traffic Engineering and Control*, **35**(6), 379-385.

ROBERG P (1995a). "A distributed strategy for eliminating incident-based traffic jams from urban networks", *Traffic Engineering and Control*, **36**(6), 348-354.

ABBESS C R and ROBERG P (1995b). "Fractal structure of traffic jam images", *Proceedings IMA Conference: Complex Stochastic Systems and Engineering* (TITTERINGTON D M, Editor), **54**, 251-281.

ROBERG P and ABBESS C R (1998a). "Diagnosis and treatment of congestion problems in central urban areas", *European Journal of Operational Research*, January.

ROBERG P (1998b). "The importance of detection in controlling traffic jams in rectangular grid networks", *Transportation Research C (Special Issue)*, (RITCHIE S, Editor), to appear

Appendix A

This appendix describes the technique which has been implemented in the simulation model to detect the steady state of the system. The appendix begins with a brief survey of a selection of standard methods available for this purpose and then contrasts them with the technique used in this research. Following this, a number of simulation experiments are conducted in order to determine the effect of random seeds on the steady state detection algorithm.

A.1 Introduction

When activated, most systems pass through what is commonly known as a *transient state* before settling down into a *steady state*. A simulation is said to be in its steady state if its current behaviour is independent of its initial starting conditions. Most simulation programs will attain a steady state if they are run in for a long enough period.

With a steady state simulation, the starting conditions (the initial values of the simulation parameters) will influence the time taken to reach the steady state. If results are collected over the entire simulation period, then the results from the run in period will influence the overall results and the parameter estimates. Consequently, for accurate results, it is advisable to ensure that the simulation should be in a steady state during the period when statistics are collected. This can be achieved in a number

of ways, Davies and O'Keefe (1989).

1. Start the simulation in an empty state and run the simulation for sufficiently long time so that the effect on the mean and variance of the statistics being collected is small.
2. Start the simulation nearer to its steady state by priming the model with information from a 'real-life system'. This requires good knowledge of similar system's characteristics.
3. Simulation is run up to the steady state once and its finishing point is used as a starting point for subsequent runs.
4. A simulation is run from empty up to the steady state. Statistics are collected from the simulation once it has reached the steady state. Results from the run in period are discarded.

With the first three methods, there is a danger that the simulation is biased. There is also the danger that a simulation which starts in a steady state with one set of values may not start in the same steady state if these values are changed. The last approach is thus generally considered more reliable. This is the approach we have adopted here.

A.1.1 Location and application of steady state

In the original simulation model, traffic flow was deterministic and, providing the difference between the system input and output remained constant, it was possible to establish the road network's steady state. In other words, the flows of traffic in the road network converged to a state in which there is no gradient and no extrema and the difference between system input and output remained constant from one time-slice to the next. Once this condition had been established, it was possible to conduct experiments regarding traffic jam development and dispersal relative to the steady state.

However, in the stochastic equivalent, it is not sufficient to examine the difference between the total input to the system and the corresponding output since the sequence will not necessarily converge.

In order to discuss possible methods for locating the steady state of a stochastic system, we first provide a definition for the steady state. Suppose that a stochastic process, say $\{X_t, t = 1, 2, 3 \dots\}$, is being observed for which a set of initial conditions, denoted by I , exist at $t = 0$. Within the context of our simulation, X_t could represent the difference between system input and output and I could describe the initial values of the simulation parameters. We suppose that the first moments of these random variables exist and tend to an asymptotic limit, independent of I , ie

$$\lim_{t \rightarrow \infty} E(X_t | I) = \mu_{\infty}$$

where μ_{∞} is defined as the steady-state mean. We are interested in determining the minimum t , call it t^* , for which the equilibrium condition holds.

Several methods for determining an estimate of t^* exist in the literature. Gafarian et al (1978) discuss the applicability of five common heuristic rules for predicting the approximate end of the run in period. Each rule is introduced and then evaluated for effectiveness according to criteria such as accuracy, precision, generality, cost and simplicity.

For example, one particular rule is the 'crossing-the-mean test', Fishman (1973). This test specifies that a running cumulative mean be computed as the data are generated, and a count be made of the number of crossings of the mean looking back from the beginning. When the number of crossings reaches a pre-specified number, then one has reached the steady state. Thus, with a series of observations $\{x_1, x_2 \dots x_n\}$, define

$$\omega_j = 1 \quad \text{if } x_j > \bar{x}_n; \quad x_{j+1} < \bar{x}_n \quad \text{or} \quad x_{j+1} > \bar{x}_n$$

$$\omega_j = 0 \quad \text{otherwise.}$$

Note that

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$$

The method then computes

$$\Omega_n = \sum_{j=1}^{n-1} \omega_j$$

which gives the number of times that the series $\{x_1, x_2 \dots x_n\}$ crosses the mean. One computes $\Omega_1, \Omega_2, \dots \Omega_t$, whence for the first time, the number of crossings is greater than or equal to a pre-assigned number. The intuitive notion is that the larger this number is, the greater is our confidence that bias due to initial conditions has been resolved. For example, if one starts with an empty system one would expect the cumulative average \bar{x}_n to increase and level out to its equilibrium value. When this happens, the individual x_j 's, measured in the near equilibrium state would be scattered on either side of the equilibrium value and thus contribute to an increasing Ω_n . However, during the early stages, the individual x_j 's would fall above this cumulative mean, and therefore add nothing to the Ω_n .

The test is popular as it wastes no data. However, Gafarian et al (1978) observe that for certain conditions, the rule tends to overestimate the time taken to reach the steady-state value. This may sometimes present a problem in terms of the number of excess observations required before the steady state is detected.

In their evaluation of the remaining four tests, the authors concluded that none were satisfactory and they did not recommend their use for practitioners. In particular, they found that a number of the rules (Conway and Modified Conway) badly underestimated the steady-state, were costly in the sense of wasting data and no practical procedure could be identified for determining either the length or the number of exploratory runs. The 'crossing-the-mean' test was found to be biased and for certain parameter values was not always precise. For a more comprehensive evaluation of these rules, see Gafarian et al (1978).

As a result, we have considered a combination of two other tests described by Kendall

and Ord (1990). The tests are similar to the 'crossing-the-mean' test although there are some differences. The first test establishes that there is no trend in the data, whilst the second test determines the lack of turning points. The authors describe the individual application of each test to different data sets, for example, in a data set describing sheep population in the UK over a certain time period, the authors fail to detect any trend in the sheep series.

Our technique *combines* these two tests to provide a method for establishing the steady state condition for the traffic simulation network. The method has the advantage that it is general (it makes no assumption about the underlying series) and is easy to apply. Moreover, the steady state is detected extremely quickly and the cost in terms of computer time is negligible.

However, there are a number of restrictions, for example, to apply the tests, the sequence needs to be partitioned into a number of samples. The outcome of the tests may sometimes be affected by an inappropriate sample size. The method for locating the steady-state of the system is described next.

A.1.2 Technique for establishing steady state

Suppose the stabilisation phase involves n time slices, where n needs to be determined. At the end of each time slice τ_i , where $i \in \{1 \dots n\}$, the program records the total input to the system along with the corresponding output. The difference between the input and the output can be denoted by $y_1 \dots y_n$. The series forms a sequence which can be tested for stationarity as follows.

The sequence $y_1 \dots y_n$ is partitioned into k samples of consecutive intervals each of size ν . Each sample is considered separately as a time series and is tested for trend and absence of turning points. The sample size, ν , is chosen empirically, not so long as to cause the stabilisation process to terminate prematurely, but not so short as to be over-sensitive to local stationarities and turning points.

If a particular sample $\{y_j | j = (k-1)\nu + 1, \dots, k\nu\}$ satisfies the conditions for stationarity then the system has achieved its steady state. Hence $n = k\nu$. Once n has been determined, it is possible to compute the average number of vehicles waiting in the steady-state network. This is denoted by \bar{w} and is defined as

$$\bar{w} = \frac{1}{\nu} \sum_{j=(k-1)\nu+1}^{k\nu} w_j$$

where w_j is the number of vehicles waiting in the system at the end of time-slice j . The standard deviation s_w of the mean \bar{w} may be defined intuitively.

As stated above, the conditions for stationarity have been adapted from a set of statistical tests described by Kendall and Ord (1990). The first test examines the time-series data for trend by comparing successive terms and examining them for increases or decreases. The test counts the number of points of increase in the series by employing a 'difference-sign' test. This will be described later.

The second test examines the data for turning points. A turning point is defined either as a 'peak' when a value is greater than its two neighbouring values, or a 'trough' when the value is less than its two neighbours. A simple test is given by counting the number of peaks and troughs.

Test 1: testing for trend

Consider a time-series consisting of n values y_1, y_2, \dots, y_n . The test for trend is achieved as follows. Let

$$q_{ij} = 1 \quad \text{if } y_i > y_j \quad \text{where } j > i$$

$$q_{ij} = 0 \quad \text{otherwise.}$$

Then let

$$Q = \sum_{i < j} q_{ij} \tag{A1}$$

Under the *null hypothesis* that the series is random, we can deduce that

$$P(q_{ij} = 1) = P(q_{ij} = 0) = \frac{1}{2}$$

Since there are $\frac{1}{2}n(n-1)$ pairs in eq(A1), it follows that

$$E(Q) = \frac{1}{4}n(n-1)$$

If Q is less than the expected value, this indicates a rising trend, whereas $Q > E(Q)$ suggests a falling trend. Note that Q is known as the number of discordances in the series and is related to the rank correlation coefficient known as *Kendall's* τ , Kendall and Gibbon (1990). Consequently, we may set

$$\tau = 1 - \frac{4Q}{n(n-1)}$$

Under the null hypothesis (ie, no trend), $E(\tau) = 0$ and

$$\text{var}(\tau) = \frac{2(2n+5)}{9n(n-1)}$$

It follows that

$$z = \frac{\tau}{\sqrt{\text{var}(\tau)}} \simeq N(0,1)$$

Test 2: testing for turning points

Again, consider a time-series consisting of n values y_1, y_2, \dots, y_n . Three consecutive points are needed to define a turning point; thus $y_i > y_{i-1}$ and $y_i > y_{i+1}$ defines a peak at time i whereas $y_i < y_{i-1}$ and $y_i < y_{i+1}$ defines a trough at time i . (This is similar to the 'crossing-the-mean' test.)

Then for $i = 2, 3, \dots, n-1$ we define the indicator variable U_i as follows:

$$U_i = 1, \quad \text{when there is a turning point at time } i$$

$$U_i = 0, \quad \text{otherwise}$$

Kendall and Ord (1990) show that the number of turning points is given by

$$p = \sum_{i=2}^{n-1} U_i$$

and

$$E(p) = \sum E(U_i) = \frac{2}{3}(n - 2)$$

By an extension to this argument (see Kendall et al. 1983, Section 45.18) we find that the variance of p is

$$\text{var}(p) = \frac{16n - 29}{90}$$

As n increases, the distribution of p approaches normality. Consequently, we may carry out the test using

$$z = \frac{p - \frac{2}{3}(n - 2)}{\sqrt{v(p)}} \simeq N(0, 1)$$

When the null hypothesis is true, (ie, there are no turning points) then the distribution of z is approximately the standard normal.

ALGORITHM

The simulation model is run for a fixed number of time slices, typically 200 although this value may be increased further depending on the outcome of **Test 1**. The interval length is also fixed, typically at 40 although this may be increased depending on the outcome of **Test 2**. The difference between the total input to the system and the corresponding output is then tested for absence of trend using **Test 1**.

IF the null hypothesis is rejected (ie, trend is detected) increment the duration of the simulation and repeat **Test 1**

ELSE

Perform **Test 2**;

IF the null hypothesis is rejected (ie, there are turning points) increment the interval pattern and repeat **Test 2**

ELSE

accept sequence as steady and record the mean and standard deviation of the total number of vehicles in the system.

A.1.3 Convergence of system to steady state using random seeds

In Chapter 3 we showed a single example of the gradual approach of the road network to its steady state. This experiment was repeated fifteen times using different seeds which were generated at random by the program. The results indicate a similar trend and the steady-state of the system is detected within a short period of time. Table A.1 summarises the results of the experiment in which all the parameters remained fixed for each trial except for the seed which changed from run to run. The parameters were $\alpha = 0.50$, $p = 0.4$, $\mu = 31$, $\sigma = 0.4$, $\kappa = 50$, $\tau = 1.00$ and the storage capacity of a link was fifty vehicles.

For each trial, the system was deemed to be steady after one hundred and twenty time-slices and the average number of waiting vehicles in the steady state system was estimated to be in the region of 9400 vehicles. Each of the time series exhibited similar characteristics although some variation was observed from one time-slice to the next. A sample time-series is shown in Figure A.1. This is similar to the series shown in Figure 3.8 of the thesis. The first twenty five data items (describing the difference between system input and output) from a selection of simulations are provided in Table A.2 to show the variation described above.

A.2 Summary

In this appendix, we have provided a brief survey of some of the standard techniques available for detecting the steady-state of a stochastic system. We then described how a similar set of tests can be adapted in order to locate the steady-state of the traffic simulation model. The application of this method using random seeds demonstrated the robustness of the procedure to fluctuations in the arrival patterns. In the next appendix, we describe how to calculate delays in the system relative to its steady state, and test the validity of the proposed algorithms under various assumptions.

Test No.	Seed Value	\bar{w}	s_w
S1	46359	9436.56	53.12
S2	46629	9450.53	50.95
S3	46791	9490.83	39.30
S4	46937	9462.11	52.75
S5	43737	9436.92	64.20
S6	43777	9461.65	64.05
S7	43819	9417.30	42.30
S8	43856	9426.55	89.72
S9	43895	9474.67	78.34
S10	43938	9446.42	91.20
S11	44027	9450.30	71.18
S12	44067	9440.85	51.81
S13	44107	9417.22	89.20
S14	44147	9442.53	56.48
S15	44578	9515.20	31.60

Table A.1: *Convergence to steady state using different seeds*

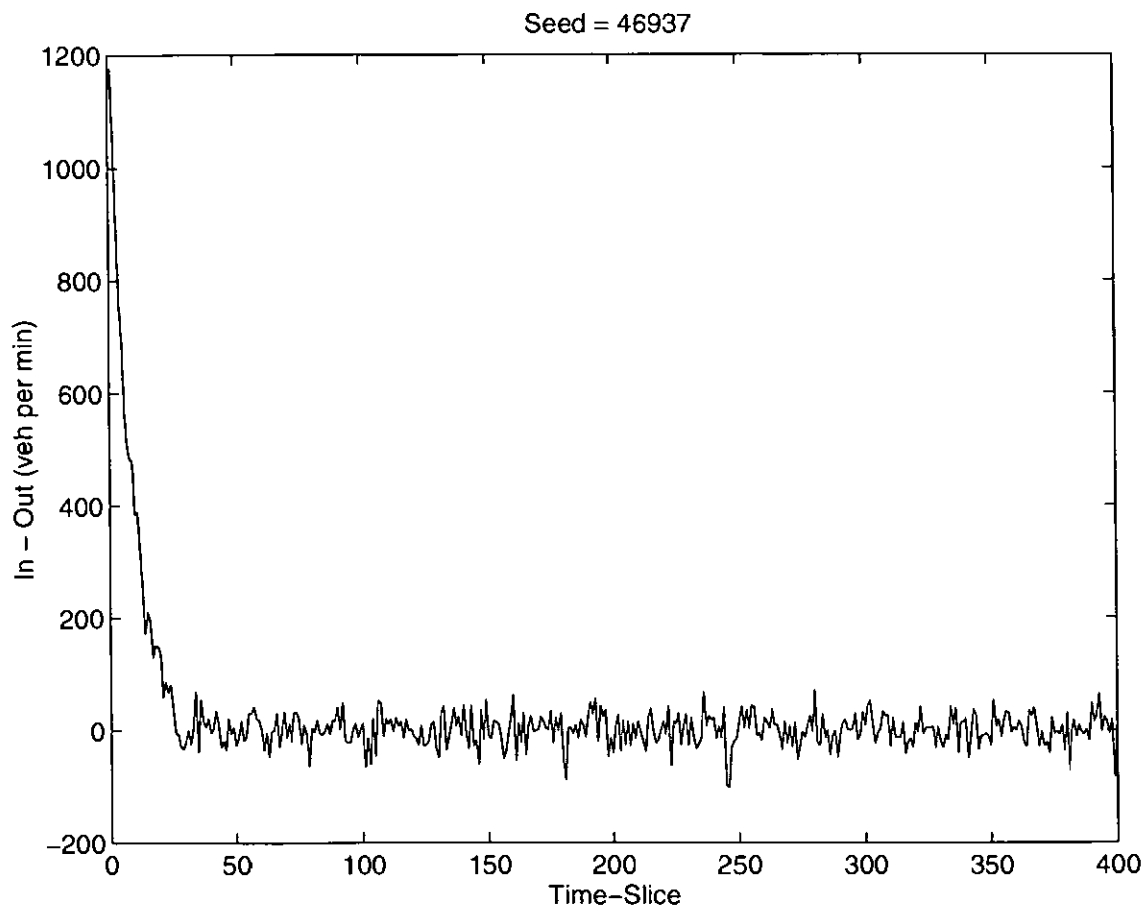


Figure A.1: *The gradual approach to stationarity using a random seed*

τ	S1	S2	S3	S4
1	1173	1160	1153	1177
2	1040	1057	989	1033
3	909	935	875	897
4	745	772	740	761
5	651	661	677	694
6	547	622	617	564
7	511	501	565	506
8	518	397	494	481
9	401	461	435	481
10	326	422	411	383
11	352	331	332	388
12	305	318	272	319
13	252	278	287	251
14	231	217	259	171
15	201	159	180	210
16	187	155	192	194
17	151	163	157	129
18	120	103	165	151
19	147	153	148	150
20	126	118	109	135
21	97	98	117	58
22	93	110	49	86
23	101	143	78	66
24	81	41	38	81
25	95	59	58	37
\vdots	\vdots	\vdots	\vdots	\vdots

Table A.2: *Time-Series data (Input - Output) for steady state convergence using different seeds*

Appendix B

This appendix provides the algorithms which have been implemented to measure the total delay incurred by the system as a result of a traffic jam. Following this, a number of simulation experiments are described. These include the evaluation of the effect of random seeds on the patterns of jam dispersal described in Chapter 4 of the thesis. In addition, the effect of random seeds on the patterns of delay for various combinations of demand, turning proportion and channelisation parameters is also described. Whilst some variation was observed from one simulation trial to the next, the overall patterns remained unchanged with respect to variations induced by the random seed.

B.1 Background

In order to compare the efficiency of various strategies for controlling traffic jams a statistical measure is required. The total delay incurred by the road network is often deemed a suitable measure. In most traffic models, the total delay can be estimated without too much difficulty since the routes of drivers are goal driven and vehicles travel from particular origins towards specific destinations.

In the simulation tool described here, vehicles are not individually tracked and as a result we needed to develop a measure which would enable us to estimate the delay incurred by the traffic jam. This could then be used to quantify its severity.

Two methods have been developed for calculating the total delay incurred as a result of a traffic jam. The first method compares the total number of vehicles in the system in the steady state with the corresponding number in the congested state. The estimated delay denoted by Δ_i^* , represents the total delay in the system incurred as a result of the traffic jam, assuming that the original obstruction has been removed. This method will be referred to as the **wait totals** method.

However, this method is deficient in two ways.

1. There is a time lag between the placing of the obstruction and the observation of an increase in delay. This is caused by the fact that excess levels of vehicles in obstructed queues are compensated by lower levels in queues downstream.
2. Having delay dependent on the total number of vehicles in the system may cause an *underestimation* of the actual delay particularly when control strategies rely on rapid expulsion of vehicles from the system. In addition, random fluctuations in the demand levels may generate some undersaturated links which would also reduce the delay experienced in the network.

A second method has been devised to cope with the aforementioned problems. This method uses a direct approach to the excess queueing of vehicles at each link in the system. The average size of a queue which has formed on a link when the road network is in the steady state is recorded prior to the simulation phase. A comparison is then made between this value and the size of the queue during the congested state. The difference between these two values represents the excess queueing due to congestion and as a result forms a measure of delay. The total delay, Δ_i^* , represents the total delay incurred in the system as a result of the traffic jam on the network. This method is described later in the Appendix and will be referred to as the **positive excess queue method**.

This method addresses the problems associated with the first method but in so doing, it may overestimate the actual delay experienced by drivers. This is because

random fluctuations in the demand levels may cause some links to become temporarily oversaturated. This excess will be considered as delay caused by congestion. The implications of this are that delay may, in some cases, be over estimated using this technique.

Hence, whilst the first method may underestimate the total delay, the second may overestimate it. Consequently, the *actual* delay incurred by the presence of a traffic jam is estimated as the mean of these bounds. We shall use this statistic to compare the relative efficiency of certain control strategies in various congested situations.

This appendix introduces two algorithms which have been developed to measure the total delay incurred as a result of a traffic jam in a rectangular one-way grid network.

B.1.1 Delay - wait totals method

The algorithm, referred to at the end of **Chapter 4**, which has been used to estimate the total delay, Δ_i^* , is described in this section.

In a 20×20 one-way road network there are 40 entry points. During each time slice, the sum of the vehicles arriving at each entry point represents the system input for that time slice. The system input over all entry points at time slice j is denoted by u_j . The corresponding system output over all entry points is denoted by v_j . In addition, let W be the total number of vehicles waiting in the system. We assume that W is approximately normal with mean ω and standard deviation σ_w .

Let $E(U) = \nu$ and $\nu = 40\mu$ where μ is the mean demand per entry point defined by the demand parameter. The mean value of input will remain close to the steady state throughout the simulation. Also, let τ be the cycle time length for a single time-slice.

Now in a statistical sense, the true mean number of vehicles in the network (ω) can be estimated by the actual number of vehicles when the system has stabilised (w_0). In other words, w_0 is an estimator of the parameter ω . Similarly, $\nu = 40\mu$ is the true

mean number of vehicles entering the network (according to this design) and $\sum u_i$ is the actual sum of entries in i cycles. This estimates $i\nu$ in a statistical sense.

Hence, if w_0 is the total number of vehicles in the system just before the obstruction is placed, then the number of vehicles in the system at the end of time slice i will be given by:

$$w_i = w_0 + \sum_{j=1}^i u_j - \sum_{j=1}^i v_j$$

This can be approximated by:

$$w_i \simeq \omega + i\nu - \sum_{j=1}^i v_j$$

where ω represents the true value for the number of vehicles waiting in the system.

Rewriting and approximating again we obtain:

$$w_i - \bar{w} \simeq \sum_{j=1}^i (\nu - v_j)$$

Since $\delta\Delta_i = \tau \sum_{j=1}^i (\nu - v_j)$ it follows that the contribution to the total delay from time slice i is:

$$\delta\Delta_i \simeq \tau(w_i - \bar{w})$$

Hence the total delay in the system at time slice i is given by:

$$\Delta_i \simeq \tau \sum_{j=1}^i (w_j - \bar{w}) \quad (\text{veh} - \text{min})$$

If the implementation of a control measure in time slice s successfully treats the traffic jam so that the system recovers its steady state at time slice i , it is possible to compute the total delay incurred from time slice s to i via the definition that follows. (The total delay in this case has been computed relative to time slice s and is hence denoted by Δ_i^*).

$$\Delta_i^* = \tau \sum_{j=s}^{s+i} (w_j - \bar{w}) \quad (\text{veh} - \text{min})$$

B.1.2 Delay - Positive excess queue method

This method, referred to at the end of Chapter 4, is known as the positive excess queue algorithm because it uses a direct approach to estimating the excess queueing at each node in the system. The algorithm can be described as follows:

We first define $\eta = \omega/r$ as the expected number of vehicles in each queue in the steady state. The number of links in the system is defined by r . Let \bar{w} be the mean number of vehicles in the system. η can be estimated using $\bar{w}/r = \bar{f}$ (say).

The size of a queue in link l at the end of time slice i is denoted by f_{il} . Let L be the set of all links with queues in the system, with $L = \{l \mid l \in 1 \dots r\}$. P is the set of all links $P \subset L$ with $f_{il} > \bar{f}$. Then the contribution of time slice i to the total delay is given by:

$$\delta\Delta_i = \tau \left(\sum_{l \in P} (f_{il} - \bar{f}) - K \right) \quad (\text{veh} - \text{min}).$$

where K is the expected value of $\delta\Delta_i$ when the system is in the steady state. In other words K is the expected value of $\sum_{l \in P} (f_{il} - \bar{f})$. when the system is in the steady state and can be estimated empirically during the RUNUP stage via

$$K = \frac{1}{\nu} \sum_{i=(k-1)\nu+1}^{k\nu} \left(\sum_{l \in P} (f_{il} - \bar{f}) \right)$$

where ν is the sample size (defined in Appendix A). (Note that $K \simeq \frac{s_w}{\sqrt{2\pi}}$ using the mean deviation for the normal distribution). The total delay at time slice i is thus given by:

$$\Delta_i = \tau \sum_{j=1}^i \left(\sum_{l \in P} (f_{jl} - \bar{f}) - jK \right) \quad (\text{veh} - \text{min}).$$

As in the previous method, the total delay incurred from a time slice s may be defined intuitively.

This method of delay estimation has been implemented in the simulation model by considering each of the phases (east-west and north-south) separately and summing

the results. Thus, the average number of vehicles in the isotropic system (\bar{w}) and the corresponding average queue length (\bar{f}) have been calculated separately for each phase of the simulation via $\bar{w} = \bar{w}_e + \bar{w}_n$ and $\bar{f} = \bar{f}_e + \bar{f}_n$. The reason for this is to account for the slight asymmetry in the simulation caused by processing the East-West phase prior to the North-South phase. (The subscripts e and n respectively refer to the separate east-west and north-south phases of the simulation. Thus, \bar{w}_e refers to the total number of vehicles waiting on eastbound or westbound links in the network. The other symbols can be defined intuitively).

B.1.3 Comparison between two methods

The positive excess queue method differs from the wait method in one respect, this being the fact that whilst the wait method takes the starvation effect into account, the positive excess method only considers the effect of those links containing an excess of vehicles above the average.

If this condition is relaxed, ie *all* differences between the queue lengths and the corresponding average are considered regardless of whether they are positive or negative, then the two methods are equivalent. This can be seen in the sample output from the program.

Useful Statistics:

μ	no. of cycles	entry points
23	32	40

Average Delay using WAIT method 0.190 veh-mins

Average Delay using QUEUE (+ queues) method 0.395 veh-mins

Average Delay using QUEUE (+/- queues) method 0.190 veh-mins

The question arises as to which of the two methods is to be preferred. On the one hand, the wait totals method provides a global measure of the total delay in the

system, yet the statistic may be flawed, when strategies rely on the rapid expulsion of vehicles. In this case the actual delay might be severely underestimated. By contrast, the positive excess method would accurately estimate delay in these situations, yet, the total delay is a system characteristic and the global effects of starvation are ignored using this technique. Thus, one can view the wait-totals method as a lower bound for the total delay and the positive excess method as an upper bound. We have found that the average between these two bounds provides a more realistic reflection of the delay patterns. We have adopted the *average* of the upper and lower bounds in calculating the total delay incurred by the system.

B.2 The effect of random seeds on the dispersal pattern of a traffic jam

In Chapter 4 we described the dispersal mechanism according to three distinct phases. These were decay, oscillation and growth. In a series of simulation experiments, a traffic jam was grown for a fixed number of time-slices. This was followed by the removal of the obstruction. The timing of the removal varied and appeared to have a critical impact on the dispersal outcome.

The results of the experiment can be summarised as follows: With the elimination of the incident in the sixth cycle of the simulation the traffic jam clears within a short period of time. However, when the removal of the incident is delayed a further cycle, the dispersal pattern is not smooth but undergoes a series of oscillations before dying out. The third category highlights the build-up of the traffic jam until the extraction of the incident, which occurs in the eighth simulation cycle. Thereafter follows a re-arrangement of the vehicles within the queues, which sometimes causes a temporary reduction in the jam size.

The above experiment was repeated using different seeds which were generated at random at the start of the simulation. The results indicate similar phases in the

dispersal mechanism (decay, oscillation and growth). The results are provided in Figures B.1, B.2 and B.3.

Figure B.1 shows four examples of traffic jams in which the incident was removed in the sixth cycle of the simulation. The traffic jams were grown with identical simulation parameters to the one shown in the thesis. The value of the seed varied from one example to the next as indicated. The results show a similar although not identical pattern of jam decay.

Figure B.2 shows a further six examples of traffic jams in which the extraction of the incident was delayed until the seventh cycle of the simulation. This led to the oscillatory phase where the traffic jam undergoes a series of fluctuations before settling into a distinct pattern of *decay* or *growth*. In a series of twenty simulations where the seed varied between each experiment, in most cases the traffic jams oscillated and then began to decay, whilst for the remainder they oscillated and then began to grow. The oscillations appear to arise as a result of the original gridlock cycle migrating to a nearby location. This new cycle does not always lock up entirely causing fluctuations in the traffic jam size. If the cycle locks up the traffic jam begins to expand whereas if the cycle disintegrates the traffic jam clears. This suggests that the oscillatory phase is a delicate one which represents the boundary between success or failure with regard to dispersal.

The results for the transitional state were not unexpected. In the original model where traffic demands were uniform (ie, no stochastic element was present) similar results were achieved. It was demonstrated that when the removal of the obstruction induced the fluctuating phase, it was not possible to predict the final outcome of the dispersal pattern. In some cases, the traffic jam eventually cleared, whereas in others, the jam appeared to disintegrate but then began to grow. As a result, we defined the oscillatory phase as a transitional state on the *border* between growth and decay. This pattern continued with introduction of stochastic input and the use of random seeds. Again, in the majority of cases, the traffic jams oscillated and then decayed

but there were some examples when the traffic jam subsequently began to grow.

Figure B.3 shows a further four examples of traffic jams where the incident was removed in the eighth cycle of the simulation. The graphs show how the extraction of the incident may in some cases (although not always) lead to a temporary reduction in jam size which is followed by renewed growth. This pattern was observed using different random seeds.

B.3 The effect of random seeds on the level of delay

In Chapter 5, Section 5.4 we described an experiment which explored the impact of several control strategies on the dispersal process of a traffic jam. In the experiment, a traffic jam was grown over a fixed number of time-slices, following which the obstruction was removed. The control strategy was then implemented over a further time period. An estimate was made of the total delay incurred as well as the length of the dispersion period. However, the results correspond to a series of tests in which the value of the seed was maintained at a constant value of 5. In this section, we describe the effects of three different seeds on the total delay and the length of the dispersion period for a selection of control strategies and network conditions.

Tables B.1 and B.2 summarise the effects of random seeds on the total delay and the length of the dispersion period under increasing levels of demands. Table B.1 shows the results using the B_3 strategy whereas Table B.2 describes the results using the B_3R_6 strategy. These two strategies were selected since they appeared to be most representative and useful in clearing the jammed structures.

Referring to Table B.1 one can observe that changing the value of the seed does not alter the operation of the strategy in a substantial way. The B_3 strategy works well under low and moderate demands ($\mu = 18 \dots 23$) but fails as the level of demand

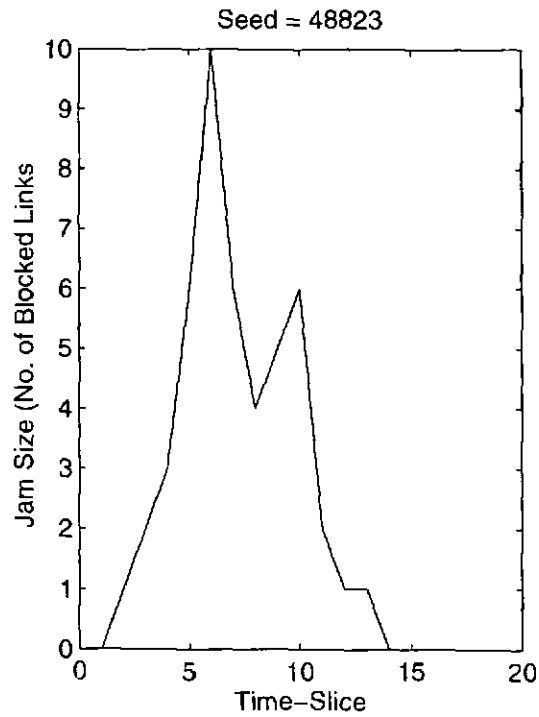
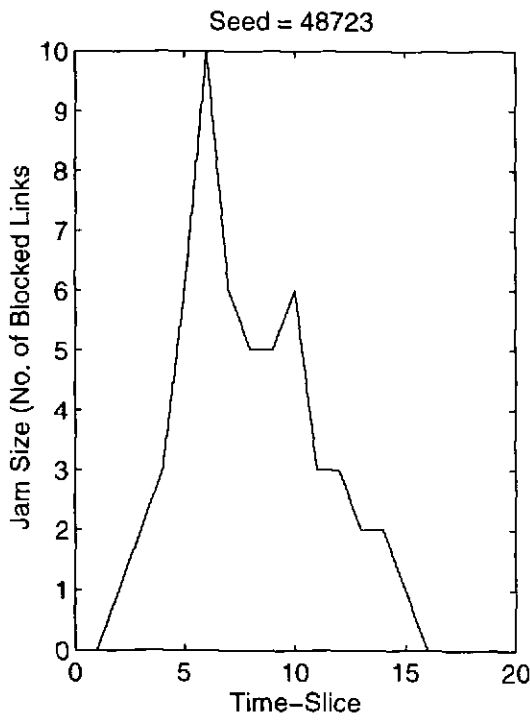
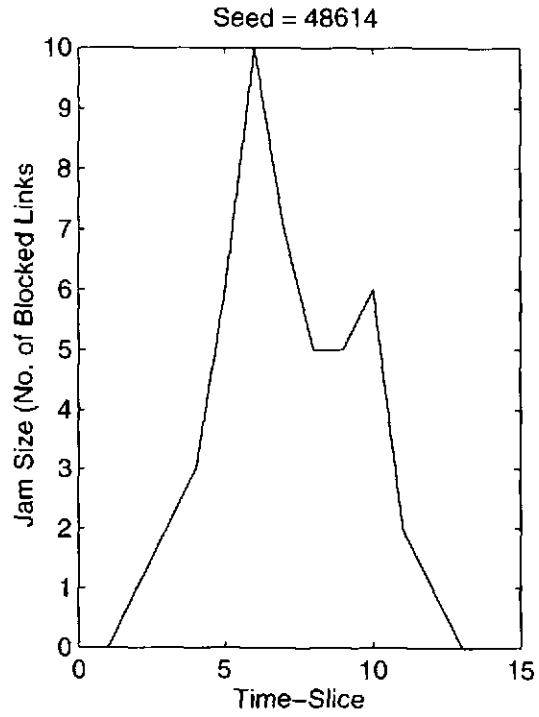
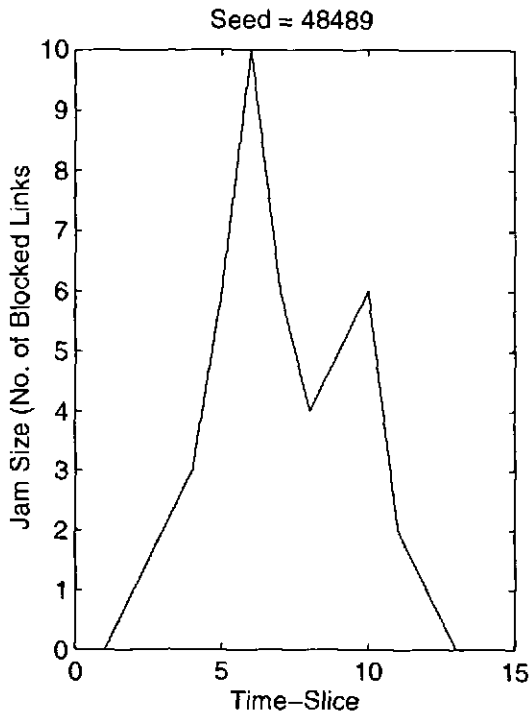


Figure B.1: *Dispersal Mechanism I: Traffic jam clears*

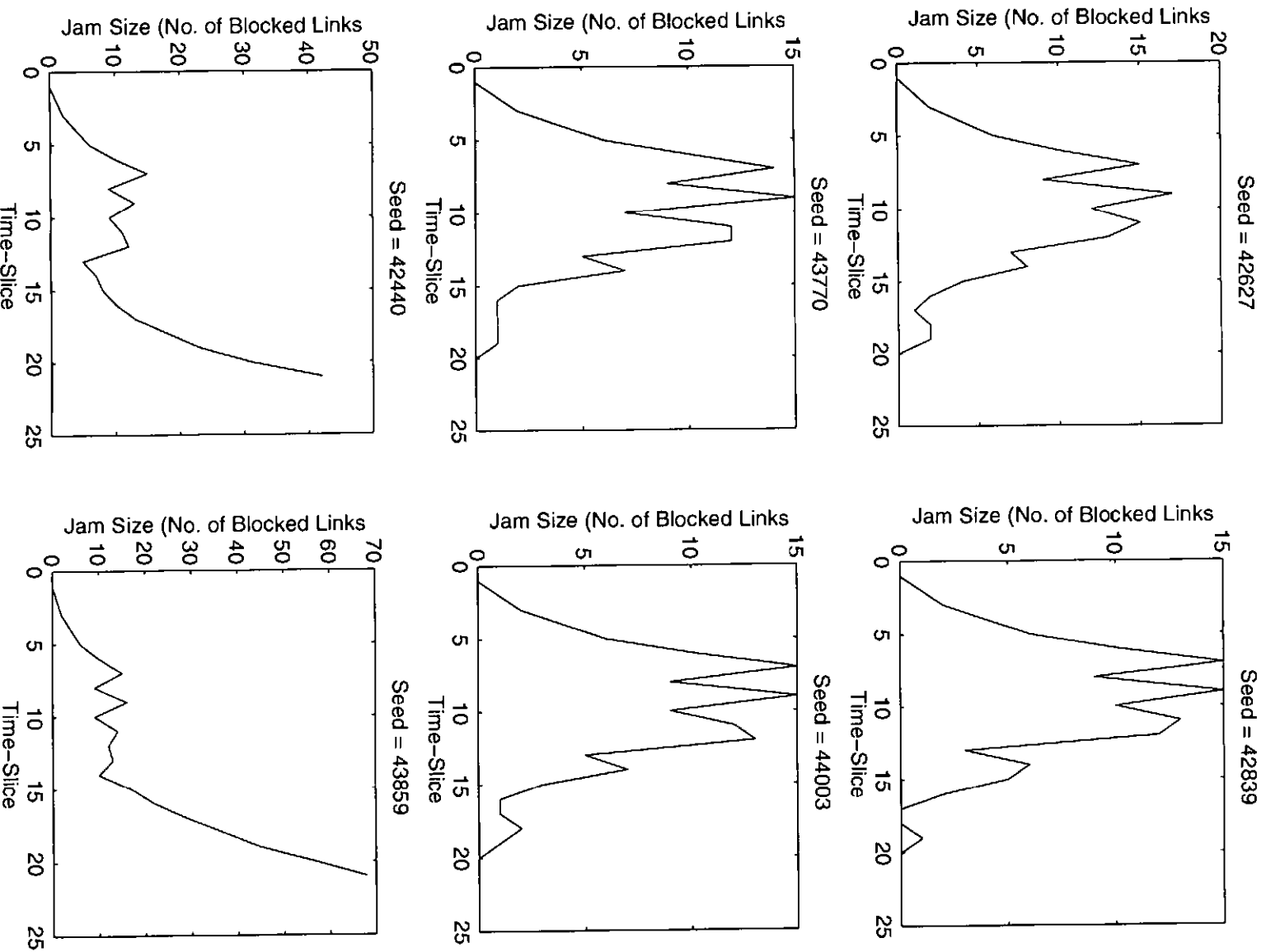


Figure B.2: *Dispersal Mechanism II: Traffic jam oscillates*

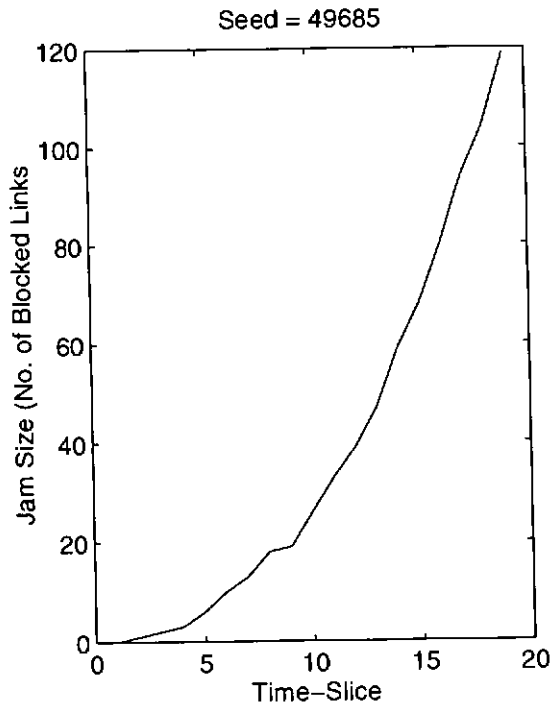
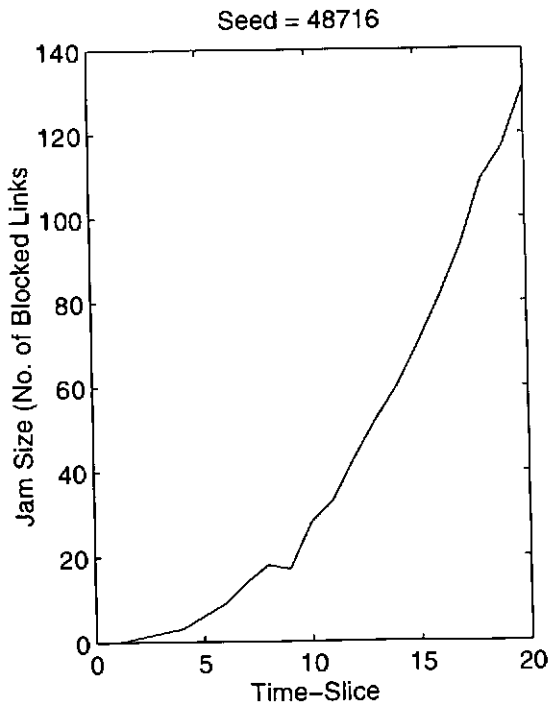
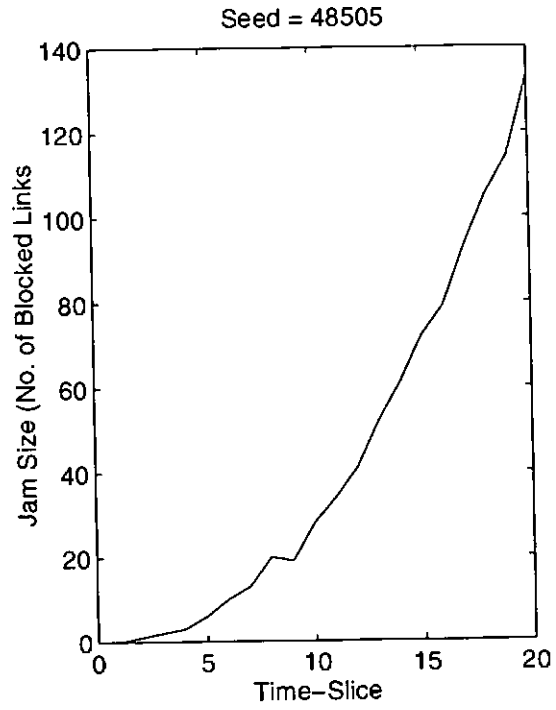
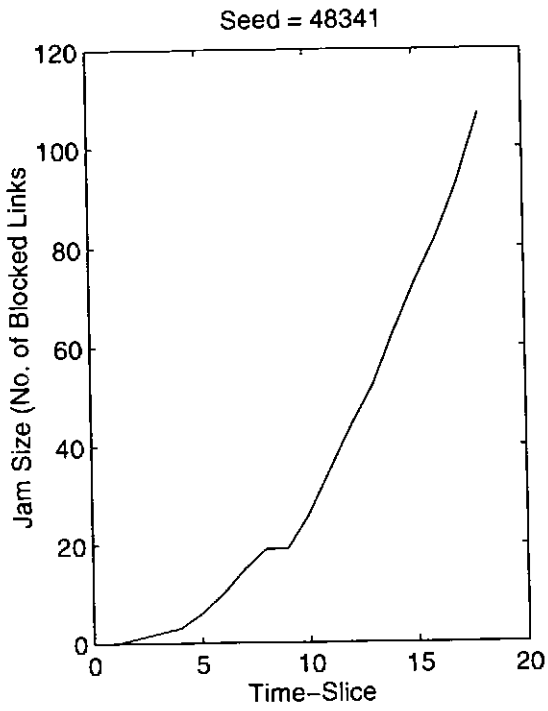


Figure B.3: *Dispersal Mechanism III: Traffic jam grows*

increases. Whilst the patterns of delay are not identical, a similar trend can be detected by comparing the respective columns of results. In addition, it appears that the time taken for the strategy to work does not change substantially with changes in the value of the seed. It follows that the mechanism of the B_3 strategy is robust with respect to increasing demand levels and does not appear to be affected by random fluctuations in the flow pattern.

Similar conclusions may be drawn concerning the B_3R_6 strategy. Again, one can observe that the strategy works only within the parameter set of $\mu = 18 \dots 27$ but fails as the level of demand increases further. By observing the columns in Table B.2 it can be seen that the patterns of the delay follow similar trends and so do the lengths of the respective dispersion periods. However, there does appear to be some variation in the total delay which becomes more noticeable under increasing demands. A possible reason for this might be the method of delay estimation. Both methods rely on the average number of vehicles in the system in the steady state (\bar{w}) and the standard deviation (s_w) of this figure. Both \bar{w} and s_w will be affected by the value of the random seed. The difference between the congested and steady state, when accumulated over a relatively long period (eg thirty or more time-slices) might contribute to the apparent fluctuations in the total delay. Nevertheless, the overall pattern of increasing demands implying increasing delay appears to apply regardless of the value of the random seed.

Note that with both the strategies (B_3, B_3R_6) there does appear to be quite a large standard deviation with respect to the delays incurred. This occurs as a result of the cumulative nature of the delay algorithms in which the delay is calculated relative to the number of vehicles in the system in the steady state. Following treatment, if the system returns to its steady state relatively quickly (eg $\mu = 18$), the variation between the levels of delay is not large. This is because the differences between the congested state and the steady state do not accumulate over too long a period. (The change in random seed induces a close yet different steady state for each trial). But, where

Demand	Seed = 41475	Seed = 42815	Seed = 43570	Seed = 5
18	2156.99 (16)	2695.55 (15)	2653.5 (14)	2581.2 (15)
20	4044.78 (16)	4836.14 (22)	4600.26 (17)	5767.2 (27)
23	18358.74 (35)	13904.81 (34)	17931.95 (32)	16455.1 (33)
25	FAIL	FAIL	FAIL	FAIL
27	FAIL	FAIL	FAIL	FAIL
29	FAIL	FAIL	FAIL	FAIL

Table B.1: *Total Delay incurred under increasing demands using strategy B_3 for various seeds*

the strategies are implemented over relatively longer periods, and the system takes longer to converge back to its original state (eg $\mu = 27$) the effects of this variability are more noticeable, and these accumulate from one time-slice to the next. These differences cannot be ignored, but, one can see that the time taken for the strategies to work is less sensitive to the value of the random seed.

Let us now consider the behaviour of the B_4R_6 strategy under a fixed level of demand $\mu = 25$ but with increasing turning proportions ($p = 0.1 \dots 0.5$) using different random seeds. This strategy has been selected over the other three alternatives because it was most effective.

In Section 5.4 of the thesis, we noted that the strategy performed less satisfactorily as the turning proportion increased. However, it was difficult to measure this effect because the behaviour of the strategy did not follow a systematic trend with respect to increases in p . (See the bottom right figure reproduced in Figure B.4). The oscillation effect in the total delay did not lessen even when the interval between successive values of p was reduced.

Demand	Seed = 41475	Seed = 42815	Seed = 43570	Seed = 5
18	2348.73 (7)	2228.59 (7)	3101.18 (9)	2950.2 (11)
20	3972.76 (9)	3763.48 (9)	5045.44 (10)	5409.6 (14)
23	11692.82 (29)	9906.53 (29)	11297.6 (27)	9376.6 (28)
25	23963.67 (45)	26339.45 (45)	19055.82 (36)	19415.0 (35)
27	44869.46 (84)	52486.67 (84)	41271.44 (88)	49450.5 (99)
29	FAIL	FAIL	FAIL	FAIL

Table B.2: *Total Delay incurred under increasing demands using strategy B_3R_6 for various seeds*

The series of experiments using different random seeds was performed in order to identify a possible cause for these apparently random fluctuations. The remaining graphs in Figure B.4 exhibit similar patterns of delay (the data for the graphs is provided in Table B.3). All four graphs are characterised by a drop in delay which is followed by a rise as p increases from 0.2 to 0.35. In all but the bottom right graph, the delay then drops again and is then followed by an upward trend as p approaches 0.5. By contrast, the delay in the bottom right graph fluctuates. The reason for this could be a random fluctuation in the demand which induced additional delay.

The above results suggest that the pattern of delay is fairly insensitive to the value of the random seed. This result is confirmed in Table B.3 by observing the comparative length of the dispersion periods.

Finally, let us consider the contribution of random seeds to the level of delay for various channelisation regimes and proportions of turning. For this experiment, we

Turning (p)	Seed = 42899	Seed = 54321	Seed = 49086	Seed = 5
0.1	20214.83 (47)	18647.13 (52)	17125.13 (50)	20064.0 (64)
0.15	11774.21 (28)	18893.04 (68)	10714.56 (19)	17250.0 (46)
0.2	11472.48 (21)	11683.88 (31)	10407.03 (29)	12465.0 (30)
0.25	23825.21 (47)	22056.76 (43)	18301.19 (43)	16560.0 (45)
0.3	18275.66 (42)	19465.14 (36)	17043.88 (30)	19602.0 (33)
0.35	35713.5 (98)	33251.16 (90)	25115.65 (64)	34020.0 (81)
0.4	17523.46 (50)	14669.44 (40)	19403.46 (60)	14658.0 (28)
0.45	14053.46 (31)	12891.44 (24)	13435.85 (25)	33056.0 (64)
0.5	14638.2 (34)	14650.65 (34)	15121.24 (40)	14742.0 (27)

Table B.3: *Total Delay incurred under increasing turning proportions using strategy B_4R_6 for various seeds*

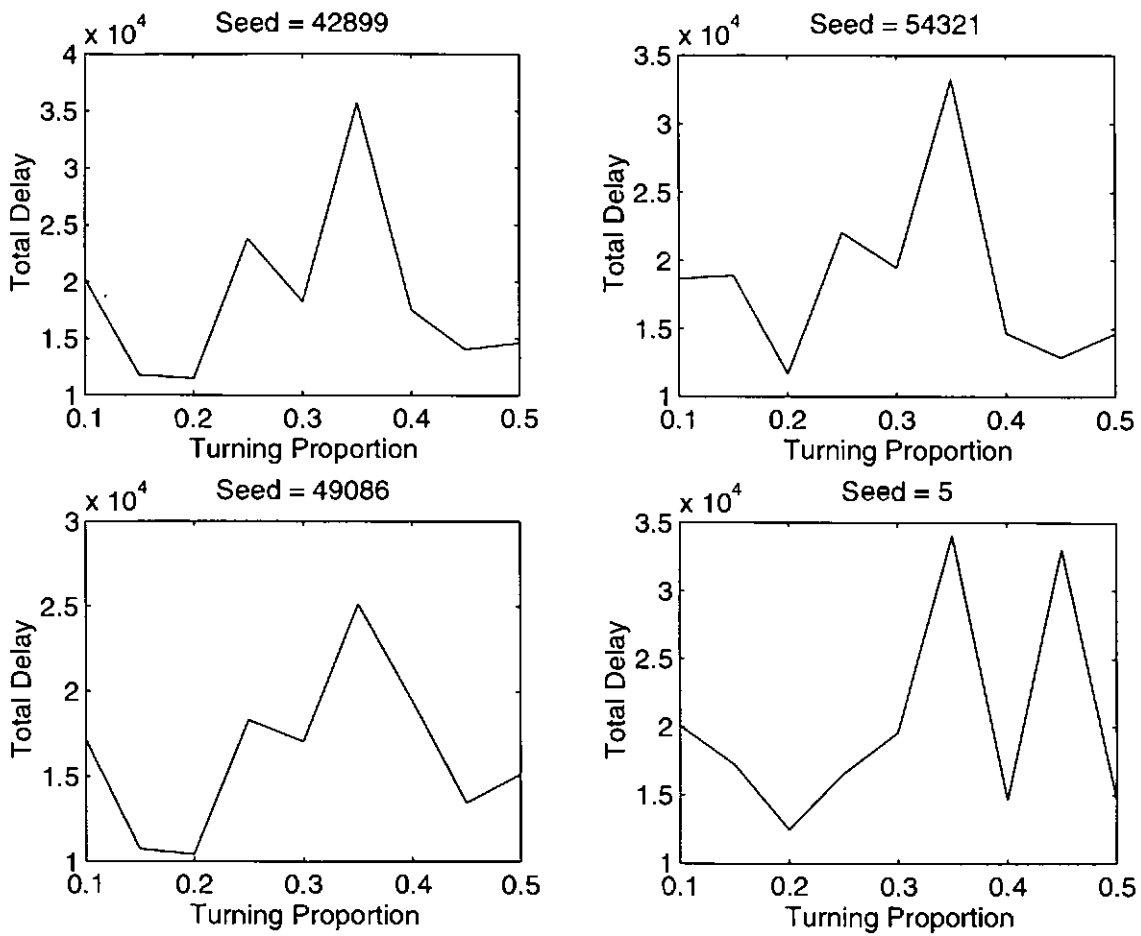


Figure B.4: The performance of control strategy B_4C_6 for increasing turning proportions and random seeds

Channelisation $\alpha = 0.5$				
Turning (p)	Seed = 42899	Seed = 42979	Seed = 45253	Seed = 5
0.1	41565.28 (60)	34179.01 (46)	41878.98 (59)	36525.0 (58)
0.15	25814.73 (47)	38343.13 (68)	27266.97 (55)	44148.0 (156)
0.2	22475.58 (38)	20729.71 (27)	23042.31 (31)	23264.0 (32)
0.25	47698.50 (60)	39133.95 (47)	43590.61 (58)	43004.0 (52)
0.3	35166.56 (42)	35438.76 (41)	36801.03 (58)	36850.0 (50)
0.35	54422.53 (99)	44405.76 (86)	47941.02 (91)	55826.0 (103)
0.4	28715.23 (43)	20743.42 (25)	23572.65 (25)	26864.0 (46)
0.45	23210.24 (31)	24393.57 (26)	23758.65 (26)	45152.0 (64)
0.5	26403.13 (32)	31528.86 (47)	29681.28 (27)	26610.0 (30)

Table B.4: *Total Delay incurred for increasing turning proportions using strategy B_4Q_6 with $\alpha = 0.50$ for various seeds*

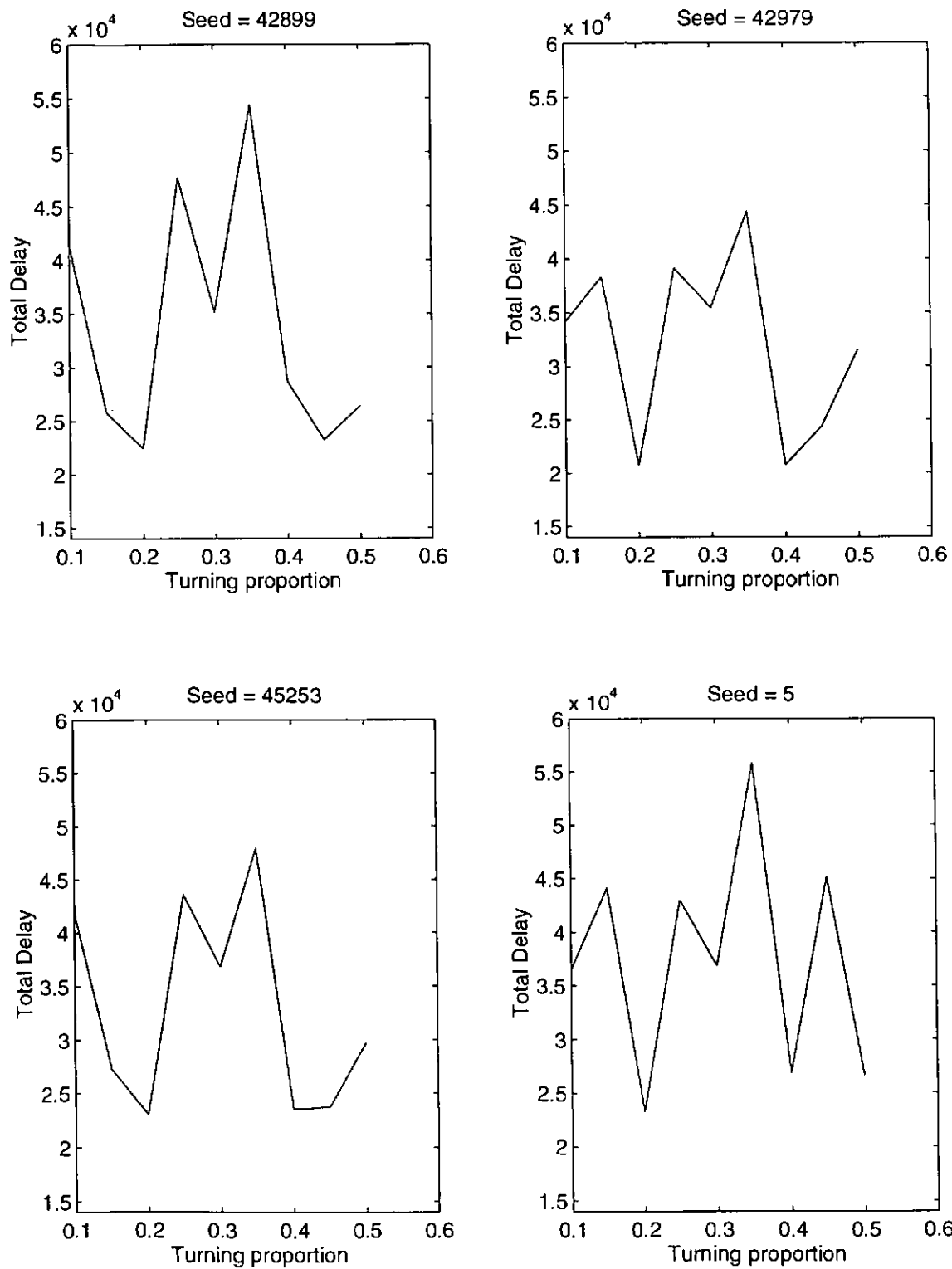


Figure B.5: The performance of control strategy B_4Q_6 for fixed α , increasing turning proportions and random seeds

have chosen to investigate the behaviour of the B_4Q_6 strategy under $\alpha = 0.50$ - the most common road layout. In the previous strategy, ahead bans are implemented around the jam boundary to channel vehicles *away* from the congested region. With B_4Q_6 , vehicles are queued at the periphery of the jam, but are not deflected away.

The results in Figure B.5 show the patterns of delay for different seed values. By comparing the bottom-right figure with the remaining graphs, it is possible to observe some similarities even though the overall trend is not identical. For example, all four graphs exhibit a maximum peak when $p = 0.35$ and a second, lower peak at $p = 0.25$. Similarly, the lowest point on all four graphs occurs at $p = 0.2$ and subsequent low points appear in the vicinity of $p = 0.4$. The data for the figures is provided in Table B.4 along with the lengths of the dispersion periods. Overall, these appear to be similar with respect to different seeds, although some exceptions have been observed (see for example $p = 0.1$). However, one would expect the dispersion periods to vary as stochastic fluctuations could give rise to increased demand at specific locations inducing extra delay. This would cause a longer dispersion period.

The delays incurred using the B_4Q_6 strategy follow a fluctuating pattern with respect to increasing turning proportions. This pattern has been observed for the other values of α . The results do not appear to be sensitive to the value of the random seed.

B.4 Summary

This appendix described two algorithms for calculating the total delay incurred as a result of a traffic jam. Following this, a comparison between the two methods was provided. A number of simulation experiments were then conducted. These evaluated the effect of random fluctuations in the arrival patterns on (a) the patterns of jam dispersal and (b) the levels of delay induced through the implementation of various combinations of control strategies. The overall conclusion was that the patterns of dispersal and delay were found to be relatively insensitive to changes in the random

seed value.

Appendix C

In this appendix, we describe a theoretical approach to traffic jam modelling. The material for this appendix has been adapted from Wright and Roberg (1997) and Abbess and Roberg (1995) and has been divided into two parts. The first part deals with spatial aspects of traffic jams on an idealised, one-way grid network. The latter part of this appendix is devoted to the fractal characteristics of traffic jams. The theoretical results are based on analytic approach to traffic jam modelling which sacrifices a degree of realism, for example, stochastic effects are not accounted for. Nevertheless, the results point to certain characteristics of traffic jam growth which would not be accessible in any other way. The fractal analysis has been conducted using the simulation tool which has been adapted by the author for this purpose. Both the theoretical work and the fractal analysis have been restricted to Type 1 traffic jams, since these are more likely to occur in practice.

C.1 Background

In this section, we derive some results concerning queue spillback and link spillback times and then show how certain features in the road layout can be manipulated to curb traffic jam growth and inhibit gridlock. We then describe the fractal nature of idealised traffic jams.

3.1.1 Queue spillback times

Any road link in the network is divided into three distinct queue storage areas. It is useful to derive a basic result that applies to each of them. Suppose that owing to congestion downstream, the discharge rate from the storage area in question is reduced to a value equal to a constant c times the demand, where $c < 1$, causing a queue to form. Eventually, the queue will grow until it fills the storage area.

Suppose the storage area is of length L' , the maximum number of vehicles that can be stored in it being N' . Let the journey time of vehicles through the storage area when vehicles are moving normally under unobstructed conditions be T' , and let the demand (ie, the input flow at the upstream end) be q' vehicles per unit time.

Define the origin as the moment when discharge is initially obstructed. There will be $q'T'$ vehicles in the area at this moment. Let the time taken for a queue to develop and fill the storage area be t' . Then the number of vehicles in the queue at the moment when the storage area fills will be equal to the number in the area at time zero plus the number entering during the interval t' , minus the number leaving during that interval. It follows that

$$N' = q'T' + q't' - cq't'$$

Rearranging, we get

$$t' = \frac{(N' - q'T')}{q'(1 - c)} \quad (B1)$$

The value of t' represents the time taken for a capacity restriction to propagate from one end of a queue storage area to the other.

C.1.2 Link propagation times - derivation

Referring to Figure 3.2 in the thesis, let N_A and N_{LR} respectively be the maximum number of vehicles that can be stored in the *ahead* and *turning* segregated areas. Also, let N_F denote the maximum number of vehicles that can be stored in the reservoir

proportion of the link. Then, N , the maximum number of vehicles that can be stored on a link, is equal to $N_A + N_{LR} + N_F$.

Now, let T_A be the journey time of vehicles through an *ahead* storage area, when there is no obstruction, to the flow, and let T_{LR} and T_F be the corresponding journey times through the turning segregated storage area and the upstream reservoir respectively. Since the ahead storage area has the same length as the turning storage area, we assume T_A and T_{LR} are equal. Also, T , the journey time of vehicles along a whole link when there is no obstruction to flow, is equal to $T_A + T_F$ or $T_{LR} + T_F$.

Using eq(B1), we can now evaluate each of the two propagation times t_A and t_{LR} as the sum of the relevant segregated storage area spillback time and the reservoir spillback time thus:

$$t_A = \frac{(N_A - q_A T_A)}{q_A(1-c)} + \frac{(N_F - q T_F)}{q(1-c)}$$

and

$$t_{LR} = \frac{(N_{LR} - q_{LR} T_{LR})}{q_{LR}(1-c)} + \frac{(N_F - q T_F)}{q(1-c)}$$

Now since $N_A = \alpha\sigma N$, $N_{LR} = (1-\alpha)\sigma N$, $N_F = (1-\sigma)N$, $q_A = (1-p)q$, $q_{LR} = pq$ and $T_A + T_F = T_{LR} + T_F = T$, it follows that

$$t_A = \frac{N}{q(1-c)} \left[\frac{\sigma(\alpha-1+p)}{(1-p)} + 1 - \frac{qT}{N} \right] \quad (B2)$$

and

$$t_{LR} = \frac{N}{q(1-c)} \left[\frac{\sigma(1-p-\alpha)}{p} + 1 - \frac{qT}{N} \right] \quad (B3)$$

A little algebra yields

$$t_A - t_{LR} = \frac{N}{q(1-c)} \frac{\sigma(\alpha-1+p)}{p(1-p)}$$

Note, that the above equation relies on the assumption that the journey time through the turning segregated storage area and the corresponding time through the ahead storage area are the same. However, even if they were different, the effect would only be to add a constant term to the equation. This does not alter the behaviour of the model fundamentally.

C.1.3 Minimising traffic jam growth

It is useful to examine the effect of varying some of the network parameters on the traffic jam's growth rate. These include the proportion of link area that is devoted to segregated storage space (σ), and the allocation of stopline widths between ahead and turning traffic (α). We shall consider Type I traffic jams only since these are most likely to occur in practice.

Now, Type I traffic jams have a diamond-shaped boundary with the original blockage located in the centre. The initial queue propagates upstream from link to link in the opposite direction to the ahead flow, and the tail of this queue defines one corner of this boundary. After an elapsed time τ , the number of links covered by this queue will be τ/t_A .

The diagonal of the square will be twice this distance, and the area, (denoted by A), of the traffic jam at time τ will be given by

$$A = \frac{2\tau^2}{t_A^2} \tag{B4}$$

blocks. Clearly, in order to minimise the rate of growth of the traffic jam we need to maximise the value of the ahead link propagation time.

Now, from eq(B2) it follows that this link propagation time is given by

$$t_A = \frac{N}{q(1-c)} \left[\frac{\sigma(\alpha-1+p)}{(1-p)} + 1 - \frac{qT}{N} \right]$$

which, when substituted into eq(B4) gives

$$A = \frac{q^2(1-c)^2}{(N-qT)^2} J \tau^2$$

where

$$J = \frac{2}{(1+\Gamma)^2} \quad \text{with} \quad \Gamma = \frac{N\sigma(\alpha-1+p)}{(N-qT)(1-p)}$$

and $0 < \alpha \leq 1-p$.

The quantity J is a constant, which we refer to as the 'jam area coefficient'. The larger the coefficient, the faster the jam grows.

Now, under appropriate conditions, α could lie anywhere within the closed interval $(0, 1)$. By considering all the possible combinations of α values that could occur together with the other parameters, one can show that the lowest possible value that Γ could assume is -1 .

The smallest possible value of J for Type I jams, and hence the smallest rate of jam growth, occurs precisely when $\Gamma = 0$. This is due to the constraints $\alpha < 1 - p$ and that Γ never assumes a value less than -1 . This condition will be met if either

- (a) $\alpha \rightarrow 1 - p$, ie, a 'balanced layout in which the widths allocated to the segregated approach queues are in exactly the same ratio as the demands, or
- (b) $\sigma = 0$, ie, the area and hence the length of road link over which the queues are segregated is equal to zero.

The two configurations (a) and (b) are alternative ways of achieving the same thing: they both avoid wasted space in the 'ahead' and 'turning' areas that cannot be filled because of queue interference between arriving vehicles. Consequently, they delay spillback to the upstream junction. This conclusion applies to any road link in any network, not just the idealised models described here.

So it seems desirable to aim in practise for

- a balanced layout insofar as it can be achieved
- the minimum length of segregated queue storage consistent with the need for lane discipline under normal operating conditions

C.1.4 Inhibiting gridlock

We now turn to the question of gridlock, where the situation is quite different. Consider a set of four links that surround a city block, and suppose that an obstruction occurs at the stopline on one of these four links. A queue will form and eventually

propagate to the the other three links in turn, so that the tail of the queue eventually arrives at the starting point of the jam.

Let us refer to the time required for the queue to tail back around the four sides of the block as the 'gridlock time'. Initially, the obstruction may affect only the turning traffic on the link under consideration, or the ahead traffic, or both simultaneously. We will assume that it affects the turning traffic: the other cases differ only in detail and the overall conclusions are the same. The gridlock time, denoted by t_g , is given by

$$t_g = 4t_{LR} \quad (B5)$$

But, from eq(B3) we have

$$t_{LR} = \frac{N}{q(1-c)} \left[\frac{\sigma(1-p-\alpha)}{p} + 1 - \frac{qT}{N} \right]$$

which when substituted into eq(B5) gives

$$t_g = \frac{4(N - qT)}{q(1-c)} G$$

where G , which we refer to as the 'gridlock time coefficient', is a function of the layout parameters, and is given by

$$G = 1 - \frac{\Gamma(1-p)}{p}$$

Other things being equal, we would wish to postpone gridlock for as long as possible, and hence maximise G . Now by definition the proportion of ahead vehicles $1-p$ satisfies the condition $0 < 1-p < 1$ so that $\frac{\Gamma(1-p)}{p} > 0$. It therefore follows that G is greatest when Γ is large and negative. This implies $\alpha < 1-p$. In other words proportionately more stopline width is allocated to the turning traffic so that $\alpha - 1 + p$ is strictly less than zero. We will have a Type 1 traffic jam, and it will be advantageous to set $\sigma = 1$, in other words, to segregate the ahead and turning queues along the whole length of each link.

Qualitatively, it is easy to see that to postpone gridlock, we reserve as much storage space as possible for turning vehicles reducing the rate at which spillback proceeds around the four sides of the block.

Such a strategy contrasts markedly with that required for minimising the jam growth rate, where we would aim for a balanced layout ($\alpha = 1 - p$, Type III jam) and minimum segregation. There is a direct conflict between the requirement to maximise the gridlock time, and the requirement to minimise the overall growth rate of the jam.

C.2 Fractal aspects of traffic jams

The fractal dimension of a particular shape can be estimated in a number of ways Mandelbrot (1983). One technique described by Morse et al (1985) is known as a box-counting method. In their paper, the authors apply their box-counting method to estimate the fractal dimension of certain forms of vegetation. The technique relies on superimposing a grid network over the object and counting the number of intersections of the vegetation with the grid at increasing levels of magnitude. In this section, we describe how to adapt this algorithm to assess the fractal dimension of traffic jams. The algorithm has subsequently been incorporated as part of the simulation tool so that it is possible to estimate the fractal dimension of a traffic jam automatically.

C.2.1 Box-counting technique

The technique can be summarised as follows:

- The simulation tool is used to generate a traffic jam image over a network comprising M^2 interconnecting nodes, arranged in the form of an $M \times M$ grid, where $M = 2^k$. (The nodes are connected by streets and in effect act as junctions.) The components of the traffic jam image are precisely those streets which have assumed a *blocked* status. A *blocked* street is one which is full and can accommodate no more vehicles.
- A square grid of side L is drawn to enclose the traffic jam image. The grid is superimposed upon the road network so as to avoid coincidence of the road network with the super-imposed grid. This eliminates possible ambiguities which

may arise during the counting procedure.

- The square grid comprises 2^{2k} square boxes of unit area. (In other words, $k = \log_2 L$.) The grid can be partitioned into $2^{2(k-i)}$ square boxes of side 2^i , where $i = 0 \dots k$.

The grid is then partitioned in a number of ways, this procedure being referred to as the *divisions process*. Initially, there will be L boxes of side 1, followed by $(L/2)^2$ boxes of side 2 in the second stage, up until the final stage when there will be a single box of side L . At this point the *divisions process* terminates.

- At each stage in the partitioning procedure, the number of boxes of side $r = 2^i$, denoted by $N(r)$, that intersect the components of the traffic jam image is recorded. A plot of $\ln(N(r))$ vs $\ln(1/r)$ is then obtained. Providing the resultant graph is approximately linear with gradient D , then D can be interpreted as the fractal dimension of the proposed image since

$$\ln(N(r)) = D\ln(1/r) + \ln(P)$$

where P is constant, whence

$$N(r)r^D = P$$

The algorithm presented above can be applied in principle to networks of any size. However, large networks require a large amount of computer time and resources, and so we have limited the size of the networks to 512×512 nodes. This, in practice, is very large compared with real road networks. The experiments have been conducted on a Digital Alpha AXP 3000 workstation.

C.2.2 Discussion

The fractal nature of the traffic jam boundary is currently only of theoretical interest. However, there may be some practical use for this result. Currently, it is difficult to

measure the extent and severity of a traffic jam that occurs in an irregular network since the traffic jam does not develop in a uniform manner. By estimating the fractal dimension of the traffic jam, by adapting the technique described above, it may be possible to quantify the density of congestion in an affected region. Thus, a fractal dimension close to 1 would indicate low severity whereas a dimension approaching 2 would point to severe congestion (See Chapter 4 for reasoning). In other words, the fractal dimension of a traffic jam could act as a congestion index or scale which would reliably classify the congested condition. The appropriate form of control could then be implemented accordingly.

Unfortunately, the fractal analysis in this research has been restricted to Type I traffic jams only. The practical constraints of the research programme have not made it possible to pursue this aspect further. However, if the resources had been available at the time, the research would have been conducted in the following way. First, the fractal dimension of Type II traffic jams would be estimated. Comparisons between these dimensions and the results for Type I traffic jams would then be made. The comparisons would focus on the relationship between traffic jam structure and the associated fractal dimension. This would lead to traffic jams on two-way networks. The structure of two-way traffic jams has been shown to be more dense than the one-way counterpart and this would be reflected in the fractal dimension. Eventually, with the development of the irregular network prototype, one could conduct further dimension estimation tests and develop an overall picture of the relationship between traffic jam structure and the associated dimension. This quantification of jam density would serve as a useful indicator when attempting to disperse the traffic jams since less dense structures are more likely to clear.

C.3 Summary

This appendix builds on some of the earlier analytic modelling work (described in Chapter 4) to provide a comprehensive analysis of traffic jam structures from a the-

oretical viewpoint. A number of practical recommendations related to reducing traffic jam growth and inhibiting gridlock are proposed. The algorithm which has been used to study the fractal characteristics of traffic jams is then described along with a discussion concerning the application of fractal research to the study of traffic jams.

Appendix D

Appendix D consists of the *USERGUIDE* for the simulation software. A floppy 3.5 inch disk can also be provided. This contains the PC version of the simulation tool and can be used for exploratory purposes. The simulation algorithms used in the model are described in a separate appendix, (Appendix E).

D.1 Introduction

D.1.1 Software objectives

The software portrays a screen simulation of the build up of traffic congestion on an idealised network. The software allows the user to explore the growth of congestion on the network by experimenting with the values of the key parameters, these being the level of demand, the proportion of turning vehicles and the geometric configuration of the road network. As well as these aspects, other related factors can also be adjusted culminating in a graphical description of a congested system. The software aims to gain a fundamental understanding of the nature of traffic jams, their causes and effects.

D.1.2 Userguide structure

The *USERGUIDE* is structured on a number of levels. Each level attempts to clarify problems at increasing levels of detail. At the lowest level the *USERGUIDE* is aimed

at helping the unfamiliar user cope with the software environment. This includes understanding the basic design of the software but mainly focuses on ways and means of getting the simulation to run and produce results. When running the simulation a number of options will be displayed. Section 5 provides a useful overview to the simulation in the form of a 'hands-on' tutorial.

Following this, the user should be directed to Sections 2 and 3 of the guide in order to interpret the screen images. In-depth descriptive algorithms are best omitted at this stage. The introductory approach should be supplemented with experimentation with the software. A study of Section 4 will help the user digest some of the textual output produced by the simulation. Following this, it would be useful to study the algorithms which constitute the simulation. In particular the user should refer to Section 3 and Appendix E which describes the main algorithms in detail.

D.1.3 Software versions

Currently, there are two versions of the software, one developed for the workstation environment and the other, for the PC environment. The USERGUIDE is intended to accompany the PC version which is attached on a 3.5" floppy disk at the end of this guide.

The workstation version is UNIX based and was developed on a Digital Alpha 3000 AXP workstation. It differs from the PC version in a number of respects, for example it incorporates a Graphical User Interface (GUI) to assist the research process as well as other facilities which have been described in Chapter 3 of this thesis.

D.2 The screen display

On running the program a series of white inter-connecting roads will be drawn on a light blue background. The roads run both horizontally and vertically across the grid. The size of the grid is set at 20×20 which implies that there will be 760 links

and 400 intersections. The position of each intersection can be described as follows:

The conventional x and y axes in a Cartesian system have been renamed East and North axes respectively. The origin is fixed at the bottom left corner of the screen. The intersections are ordered pairs of the form (i, j) where $1 \leq i, j \leq 20$.

Each intersection is marked by a square yellow filled box. This is supposed to emulate the yellow grid-like box painted at junctions of busy roads. This means that vehicles can only proceed if the destination road is clear. Otherwise they must wait at the particular road on which they are currently situated.

The roads are laid out in a staggered manner. This is to help identify the 'SOURCES' and 'SINKS' on the network. The sources are the extended roads around the network boundary. The sinks appear flush with the border of the grid. Traffic arrives at the edges directly into the sources and leaves via the sinks.

Once the skeleton network has been displayed, the distribution of vehicles appears on the screen. First, the traffic layout on the East/West road system is shown followed by the North/South counterpart.

D.2.1 Traffic directions on the network

The road network is a one-way system. At each intersection, vehicles must choose from one of two possible directions: ahead or turning. If one takes a glance at the screen display, a quick method to interpret the system of directions is to identify a road, and observe the position of the source/sink. If the source is situated on the left of the screen then vehicles travelling on this road will be progressing eastwards. Similarly if the source is on the right of the screen, then the traffic will be moving westwards. An identical approach can be adopted for determining Northbound/Southbound traffic movements.

D.2.2 Road link layout

A road is divided into two distinct zones: a downstream queue storage area where vehicles are organised into separate turning movements and an upstream reservoir where the turning movements are mixed.

The queueing areas are coloured according to the following scheme. In the segregated lanes the turning traffic will either be represented by a green rectangle (right-turning) or a rectangle (left-turning). The ahead traffic will be coloured in black. Traffic in the reservoir will be represented by a dark-blue rectangle.

A blocked road link can take on two forms, which are essentially permutations of one another. In both cases, the reservoir area is full and is drawn as a solid blue rectangle extending the full width of the road. If the 'ahead' lane is blocked, then no 'turning' traffic will be allowed to leave the reservoir due to the vehicles wishing to move ahead which are blocking their exit. As a result the turning queue becomes empty (due to the lack of new vehicles entering). A similar pattern of events occurs if the turning lane becomes blocked.

D.3 Movement of vehicles

As stated above, if a vehicle (c) wishes to move from one road to another, in other words from A to B , then c will progress from A to B providing there is space in B to accommodate c . Otherwise, c will wait in A . For each intersection, the program therefore, must calculate the capacities of both the link directly ahead as well as the link into which turning cars will move.

D.3.1 Calculation of spaces

Vehicles turning left or right

The calculations in this subsection rely on the assumption that a vehicle turning (L/R) will not turn (L/R) in the immediately following move. In other words, a vehicle will not reverse on itself.

To determine a movement, one uses the state of the reservoir in the link into which a vehicle will turn. If the reservoir is empty, then vehicles turning from link A, can fill up any empty spaces in the *on* queue (see assumption) and the rest can enter the reservoir. If, however, the reservoir is not empty, this implies that the segregated *on* queue is full and as a result the number of spaces available in the reservoir is given by (total capacity of reservoir) - (total number of spaces already occupied).

Vehicles moving ahead

Again, the state of the reservoir is used as a test. However, the vehicles moving from one link to the link directly ahead, will not all wish to continue their journey in this direction, and may wish to turn after moving ahead. If the reservoir is not empty, then vehicles will gain entry providing there is space. If the reservoir is empty, then the segregated lanes must be filled, but vehicles must be placed in the correct segregated lane. This is why the calculation in this case involves sorting the traffic into those who still wish to move ahead and those who will want to turn. To do this one first calculates which of the segregated lanes has the least available spaces. This lane (say L1) is then chosen and vehicles wishing to go in that direction will enter L1. At the same time, the other lane (say L2) is also being filled but L1 will reach capacity faster than L2 resulting in the reservoir being filled. This in turn, will ensure that no more cars will be able to enter L2. This implies that some spaces remain available in L2.

To summarise, let

n = no. of cars wishing to move into the link ahead

u = no. of spaces in turning lane (in link ahead)

v = no. of spaces in ahead lane (in link ahead)

p = probability that a vehicle turns

Then, $u = pn \Rightarrow n = u/p$ and $v = (1 - p)n \Rightarrow n = v/(1 - p)$. The chosen lane is then $\min(u/p, v/(1 - p))$.

D.3.2 Calculating spillback lanes

The simulation involves the movement of traffic around the network. The progression of a vehicle from one link to another consists of two parts. Both parts, together constitute an operation known as the *release* mechanism which is performed using a 'bottom-up' approach. This mechanism works as follows. Consider a particular link, say A . Vehicles moving out of A can either move forward into a link which will be referred to as BF (B forward) or turn into the adjacent link which will be referred to as BT (B turning). To complete a successful *release* operation, vehicles from the upstream reservoir in A must be moved to the downstream queue storage area of A , and vehicles from the downstream queue storage area of A must be moved into their appropriate positions in BT and BF . The queue lengths in A , BT and BF must then be adjusted accordingly.

It is important to note that if the upstream reservoir in a link is empty then no spillback lanes will be generated during the release mechanism. However, if vehicles are present in the upstream reservoir, then the movement of vehicles from the downstream queue storage area will in turn create space into which the reservoir vehicles can subsequently move. If there is insufficient space for all these vehicles, one lane will spillback into the upstream area of the link.

The amount of available space in BT and BF was calculated in Section 3.1 of this

appendix. These results are utilised in this section when calculating spillback lanes.

Upstream reservoir movements

This section deals with spillback definitions within A .

In the downstream queue storage area, the lane with the least available spaces is defined as the spillback lane.

It so happens that the total number of vehicles that will move out of A will be the same as the total that can be released from the reservoir. This number consists of two parts. The first is *dres* defined by the number of vehicles leaving the reservoir and wishing to travel in the spillback lane direction. The second, *otheres*, is similarly defined yet the vehicles will move in the other direction. If all the vehicles moving into the none spillback lane can be accommodated then the spillback definition remains unchanged. If this is not the case, then a situation referred to as spillback crossover arises. This involves re-defining the spillback lane and repeating the procedure described above. The queue totals in the respective areas in the link are then adjusted.

Downstream reservoir movements

This section deals with spillback lane definitions for BF and BT .

Vehicles moving into BT:

The turning case is straightforward due to the assumption that a vehicle will not reverse on itself. If an upstream reservoir exists, then vehicles will be appended to the reservoir, their destination being ahead. If no upstream reservoir exists in BT , then vehicles will be allocated positions in the downstream *ahead* segregated lane. If there is not enough room for all the vehicles then the *ahead* lane assumes a spillback status, whence the remaining vehicles will begin to spillback into a newly formed upstream reservoir of BT .

Vehicles moving into BF:

The vehicles moving ahead must be divided in the correct proportions into vehicles that wish to continue in the *ahead* direction and those that wish to turn. As in the previous case, if an upstream reservoir already exists, then vehicles will be appended to this reservoir. However, if no reservoir exists, there are two possibilities. Either there will be enough space in the downstream queue storage area for all the vehicles to be accommodated generating no spillback effect, or there will not be sufficient space. If there is not enough space, then there exists a critical point such that adding an extra vehicle will trigger a spillback. Thus, one can only allow a number of vehicles into this segregated lane such that a spillback is not generated. A corresponding proportion may be released into the none spillback lane. Any remaining vehicles that could not be accommodated in the segregated lanes will then be allocated appropriate positions in the upstream reservoir. The queue lengths in *BT* and *BF* can then be updated. The storage capacities of *BT* and *BF* must then be adjusted accordingly.

D.3.3 Capacities and obstructions

Capacity initialisation on the network

The maximum number of vehicles able to cross the lights per lane is described by the parameter *ocap*. This value can be seen as an upper bound limiting the capacity of each junction. Associated with each intersection are two capacity variables, known as *ecap* and *ncap*, one for the east flow direction, and one for the north flow direction. These variables must be set for each intersection as follows.

The flow of traffic is governed by a set of crude traffic signals. Two possibilities exist: either traffic will flow in the E/W direction or in the N/S direction. The traffic lights alternate between red and green allowing one system of roads to be processed at a time. Thus if traffic is being processed in the E/W direction all the signals in the N/S direction will be red and vice versa. During a cycle or time-slice, there will be a percentage time for which the traffic lights will be green (*pcgreen*) as well as a percentage time for which the lights will have changed. Furthermore, the maximum

number of intersections which a vehicle can cross in the ahead direction during a single time slice is five. The number of vehicles allowed through at each intersection will be described by two values, (one for upstream capacity, the other for turning capacity). The turning capacity tc is represented as a fraction of the upstream capacity.

Thus

$$ecap(on), ncap(on) = ocap * cyclet * pcgreen$$

$$ecap(tn), ncap(tn) = tc * ecap(on)$$

Obstructions - a restriction on capacity

As the capacity of a destination link governs whether a vehicle shall proceed upstream, it follows that an obstruction on the network can be represented as a reduction in capacity at a particular intersection. The degree of severity of an obstruction can vary between 0...100 percent. If the percentage capacity is zero, no vehicles can pass through the junction, creating a total blockage. Other levels of blockage can be achieved by varying the percentage capacities.

D.4 Filenames in TSN

D.4.1 TSN.PAS

TSN is an acronym for **T**raffic **S**imulation **N**etwork. The Pascal source code for the simulated system of roads can be found in the filename *TSN.PAS*. Accompanying *TSN.PAS* will be an object code file called *TSN.OBJ* and an executable file which runs on any PC called *TSN.EXE*. The PC version was developed using a version of Pascal known as Prospero Pascal.

	TURN	AHEAD
EWS	1	12
EWR	2	15
NWS	1	12
NWR	2	9

Table D.1: *Interpretation of TSN.PAR data*

D.4.2 TSN.PAR

As part of the simulation exercise, TSN utilises an input file called *TSN.PAR*. On viewing the document a series of numbers will be displayed. The first line describes the simulation parameters, the second lists some of the simulation constants calculated during the stabilisation period. The remainder of the file constitutes some 6400 numbers displayed with no annotation. To interpret these numbers, one must first divide the page into sections consisting of two columns and then into sections consisting of four rows. As a result one has twenty horizontal blocks of eight numbers, and a corresponding number of vertical blocks. Each block represents an intersection. Taking an example block such as the one shown in Table D.1: one observes that each number corresponds to the number of vehicles in a particular queue. The left column in the block describes turning vehicles whilst the right column describes ahead vehicles. The top two pairs of numbers represents the number of vehicles in the E/W road links of the intersection, whilst the bottom two represent the corresponding number in the N/S road links of the intersection. The first and third pairs characterise vehicles in the segregated queue storage area of the link whilst the second and fourth pairs characterise vehicles in the upstream reservoirs of the road link.

D.4.3 TSN.DAT

Besides the graphical output, TSN produces some textual output in a file called *TSN.DAT*. The file records the simulation constants as well as variables calculated

during the stabilisation period. Following this a summary of the simulation is tabulated including the position of the obstruction, the timing of its removal as well as the timing of ban installation and removal. At the termination of each time slice, the number of blocked links in the network is recorded as well as the number of waiting vehicles. These totals are decomposed into E/W totals and N/S totals. An example of *TSN.DAT* follows:

tsn JAMS LISTING

Parameter Values

Runup Parameter values

steady	wbar	ewbar	nwbar	
80	7132.77	5195.90	1936.88	
sw	esw	nsw	KQE	KQN
53.06	34.99	20.37	133.90	73.59

SIMULATION PARAMETERS

nmax	runup	linkm	ocap	flush	cyclet	zone	demand	probtn
200	80	50	50	2	1.000	10	24	0.500

Obstruction Locations on the Network

Cycle	E or N	Position	Percentage
1	E	(7,10)	0.0

Current Link Information:

Cycle	Full Links			Waiting Vehicles		
	E/W	N/S	Total	E/W	N/S	Total
1	0	7193	991	989		
1	0	7248	1029	974		
2	1	7244	969	973		
3	2	7208	933	969		
4	3	7252	1021	977		
5	6	7224	940	968		
6	10	7205	926	946		
7	13	7176	921	949		
8	20	7172	931	935		

D.4.4 TSN.LIS

In addition to the simulation data, TSN records the calculations for the delay incurred as a result of a traffic jam using two methods (see Appendix B for details). Sample output from a typical delay calculation exercise is provided below. In this example, the obstruction was removed in the tenth time-slice and the simulation required a further twenty time slices to re-achieve its steady state.

Cycle	WAIT T	Cycle	QUEUE T
11	354.75	11	1469.70
12	770.50	12	3297.06
13	1334.25	13	5490.44
14	2105.00	14	7725.32
15	3113.75	15	10229.85
16	4348.50	16	13045.37
17	5844.25	17	15387.80
18	7539.00	18	17578.05

19	9330.75	19	19641.95
20	11171.50	20	21616.68
21	12986.25	21	23441.57
22	14719.00	22	25149.66
23	16260.75	23	26687.69
24	17647.50	24	28084.35
25	18875.25	25	29292.98
26	19999.00	26	30386.53
27	20968.75	27	31328.57
28	21880.50	28	32164.23
29	22636.25	29	32874.85
30	23248.00	30	33454.37
31	23711.75	31	33909.77
32	24104.50	32	34278.86
33	24368.25	33	34536.59
34	24504.00	34	34699.04
35	24571.75	35	34802.04
36	24590.50	36	34854.74
37	24526.25	37	34836.32

Useful Statistics:

mu	no of cycles	entry points	
25	27	40	
Average Delay using WAIT method is			0.908
Average Delay using QUEUE method is			1.290

D.4.5 SCR**.SSC

During the simulation, the screen display is updated and changed. Implicitly, the screen display is saved to a file which can be retrieved at some later date and printed

if required. The frequency of saving screens is set by utilising a variable called `period`. If `period` divides the cycle number exactly then the screen displayed at the end of the cycle is saved to a file of the form `SCR**.SSC` where the `**` represents the cycle number.

D.5 Running the simulation

D.5.1 Basic instructions

The simulation model *TSN*, examines the propagation of queues over a fixed time period. This time period is divided into equal slices, each slice representing a simulation cycle. During each cycle, all the E/W roads are processed in turn, starting with the link at the downstream end, and moving progressively upstream from link to link. Then, all the N/S roads are dealt with in a similar way. Initially this cycle is carried out several times without any obstruction being present, to allow the system to stabilise. (This period is known as the *runup* period. The number of cycles in the runup period is given by the variable `runup`.) An obstruction is then placed on a link and the jam evolves from the obstruction.

To run the simulation type *TSN* at the prompt. The following information will appear on the screen.

```
Variable parameters: Codes and Values
1:nmax=200 2:runup=20 3:linkm=50 4:ocap=50 5:flush=2 6:cyclet=1.00
7:zone=10 8:demand=31 9:probtn=0.40 0:period=10
Change defaults? (Y/N):
```

```
Simulation constants
Number of Cycles till steady      80
Mean values: wbar = 7994.58 ewbar = 5376.05 nwbar = 2618.52
Standard devs  s(w) 128.28 es(w) 96.16 ns(w) = 34.51
```

Default obstruction? Y/N:

D.5.2 Change Defaults ? No

This response eliminates the need for the runup phase since the program retrieves the parameters of the simulation from the previous run and uses the statistics which have already been calculated and remain unchanged. This improves the efficiency of the program.

D.5.3 Default Obstruction ? Yes

A single obstruction will be introduced on the network. The position of a default obstruction is seven blocks in the East direction and ten blocks in the North direction (7,10). This obstruction is placed on the E/W road to establish a blockage there. The blockage is total: no vehicle can pass through. This causes a traffic jam to evolve.

D.5.4 Change Defaults? Yes

This response invokes a second-level menu with the command:

Enter reference (0..9), C=Confirm, F=Finish:

Each variable parameter is associated with an integer value in the range {0...9}. To change any parameter, it is necessary to specify the associated numerical value which will in turn enable the change.

Example

Changing the storage capacity of a link:

Enter reference(0..9),C=confirm,F=finish: 3

linkm=(max. no. vehicles in link): 60

Enter reference(0..9),C=confirm,F=finish:

Typing **C** at this stage will cause all the parameters to be displayed including any updated values:

```
1:nmax=200 2:runup=20 3:linkm=60 4:ocap=50 5:flush=2 6:cyclet=1.00
7:zone=10 8:demand=31 9:probtn=0.40 0:period=10
```

This will be followed by the question **Enter reference(0..9),C=confirm,F=finish:**. When all necessary changes have been made the user will respond with **F** which will cause the program to initiate the *runup* stage of the simulation and re-calculate the associated statistics.

D.5.5 Default Obstruction ? No

Once the shape of a jam evolving from the default obstruction has been recognised, the user may wish to explore further possibilities. By requesting a non-default obstruction the user will be prompted with **Enter coords. (East, North) and % Capacity:**.

To satisfy this question, the user must specify the relative position of the obstruction (x, y) along with a percentage value. The percentage may vary between 0...100 percent. A total blockage is represented by zero percent, whereas 100 percent signifies the removal of the obstruction. A valid answer to the above question would be $x\ y\ p$ where $1 \leq x, y \leq 20$ and $0 \leq p \leq 100$. Any value not satisfying this range will not be accepted and the program will wait until a satisfactory answer is supplied.

Having specified the position of the obstruction, the next command will be **Enter Direction, E=E/W, N=N/S:**. Since traffic can travel in the E/W network as well as in the N/S network, it is necessary to outline the position of the obstruction relative to the direction of flow. If **E** is typed, then the blockage will be installed in the E/W network, allowing the N/S vehicles to proceed unobstructed).

Multiple obstructions may be installed by providing a positive response to the next question **More Obstructions ? (Y/N)**. This will invoke the Obstructions menu once again.

D.5.6 Turn Bans

During the simulation period, the user may experiment with the installation of various patterns of bans. These include single bans at particular locations, 'core strategies' which target gridlock cycles and 'diamond strategies' which protect the developing jam from excessive demand during the control period. The *Turn ban* menu may be invoked by pressing the ENTER key during the simulation. The user will then be prompted as follows:

```
Would you like to some (more) bans ? (Y/N): y
Group bans ? (Y/N) Y
Select Strategy
(1: Core 2: Remove Core 3: Diamond 4: Remove Diamond:)
```

Generally, one would implement the grouped bans since these are designed for the traffic jam structures. However, one can explore different patterns of ban arrangement using non-grouped bans.

Core Strategy

To specify the location of a *core strategy* one requires the coordinates of the bottom-left intersection which comprises the gridlock cycle, for example, with the default obstruction at 7,10 one would specify 7,10. This would then invoke the second-level option as to which type of ban: 1: ahead only 2: turn only. The user should specify option 1. Multiple core strategies may be implemented in a self-explanatory way. The removal of core strategies is achieved by selecting option 2 in response to the question 1: Core 2: Remove Core 3: Diamond 4: Remove Diamond:. The user needs to specify the bottom-left coordinate of the gridlock cycle whereupon the restrictions will be lifted.

Diamond Strategy

With this arrangement, the user encounters the following options:

Enter reference coordinates (E,N):
Enter reference direction E=E/W N=N/S
Enter size parameter (2,4,6,8)
Enter (old) ban type 1: ahead only 2: turn only
Would you like some (more) bans ?

For example, to install a diamond strategy centered on the gridlock cycle whose reference coordinates are (7,10), enter these coordinates in response to the first question, then specify E for the direction, then choose a size parameter which would enclose the jam, eg 6 and finally set the ban type to option 2. This would cause an envelope of ahead bans to be set around the traffic jam, centred on the original gridlock cycle and extending six blocks in each direction. This envelope would channel vehicles away from the congested region. The restrictions can later be removed by selecting option 4 in response to the `Select strategy` question.

The user is advised to practice installing various patterns of bans to become familiar with the sequence of events.

D.6 Quitting, suspending and re-starting the simulation

To halt the simulation during any cycle, press the `ENTER` key. At the end of the cycle, a message will appear in the form `ESC = stop`. If the `ESC` key is then pressed the simulation will be stopped. Alternatively, to resume the simulation, the `ENTER` key should be re-pressed.

D.7 Summary

This appendix provides some of the technical detail associated with the simulation tool in the form of a `USERGUIDE`. In addition, some practical examples of the

application of the model are described. The next appendix summarises some of the key algorithms which have been developed as part of the tool.

Appendix E

This appendix describes the main algorithms used in the simulation program. These have been condensed in the form of truth tables to assist the reader in gaining an overview of the methodology which has been employed. The following algorithms are provided:

- Simulation of demand
- Vehicle progression mechanism
- Traffic directions on the network
- Implementation of ban patterns

E.1 Simulation of Demand

Input: μ, σ

Output: Array of 1...40 values corresponding to the demand at each entry point divided into ahead and turning vehicles

Procedure

- Simulate the distribution of numbers with mean given by μ (specified as the level of demand at the start of the simulation) and standard deviation $\sigma = \sqrt{\mu}$ using the approximation

$$RD = \sigma \left(\sum_{i=1}^{12} r_i - 6 \right) + \mu$$

where r_i is a random number in the range $0 \dots 1$.

- Each number represents the demand at a single entry point. The demand is split into demand for ahead movement and demand for turning movement. The results are then stored.
- A record is made of the total input to the network for each cycle. This statistic is printed in the *TSN.DAT* file and used amongst others, when calculating the total delay incurred.

The above procedure is carried out at the start of the simulation and repeated for each simulation cycle. The seed for the random number generator is set at random at the start of the simulation.

E.2 Vehicle progression

The movement of vehicles through the road network is described in detail in the USERGUIDE for the software (Appendix D, Section D.3 as well as in Chapter 3 of this thesis. In this section we describe the *release* procedure via a series of tables. This greatly simplifies the problem since the source code for this procedure is complicated. The release mechanism is central to the simulation program.

Using the notation adopted in Section D.3 the release operation consists of progressing vehicles from link A to links BT and BF respectively. In order to do so, a number of calculations are required. These include the calculation of space availability, denoted by Q , the saturation capacity for the intersection, denoted by cap , the actual space available in practice, denoted by R and the actual number of vehicles that wish to proceed, denoted by S .

Input: Queues in links A, BT, BF, cap

Output: Adjusted queues in links A, BT, BF

Procedure:

CASE	$Q > 0$	$cap > 0$	ACTION
1	T	T	IF $Q > cap$ THEN $R = cap$ ELSE $R = Q$
2	T	F	$R = 0$
3	F	T	$R = 0$
4	F	F	$R = 0$

Table E.1: Calculation of space availability in links *BT* and *BF*

1. Calculate the spare space available in *BT* and *BF* respectively for each individual queue (turning and ahead) and compare with corresponding saturation capacities (Table E.1). During this stage, one needs to be aware of the ban status on links *A*, *BT* and *BF* since this will affect the amount of space available.
2. Calculate the respective totals of vehicles that wish to leave link *A* in principle and compare with amount of space available in practice (calculated in (1)), Table E.2. During this stage, the effect of releasing the vehicles on the vehicles left behind is calculated. This includes working out the direction of spillback and whether a spillback crossover occurs. This arises when for example, initially, the ahead segregated area spills back to the reservoir, but as a result of a release operation, these vehicles clear whereas the turning queue does not. This leads to a change in the spillback direction.
3. Having considered the effect of the release operation on link *A*, the procedure then deals with the re-allocation of vehicles in links *BT* and *BF* respectively. Here, one needs to add the newly arrived vehicles (maintained in the *S* variable) in their appropriate positions and consider the effects of spillback where appropriate. Once all the queues have been adjusted, the release operation for a single link is complete. The entire procedure then repeats itself for the adjacent link.

CASE	$R > 0$	$S > 0$	ACTION
1	T	T	IF $S > R$ THEN $S = R$
2	T	F	$S = 0$
3	F	T	$S = 0$
4	F	F	$S = 0$

Table E.2: *Actual number of vehicles that are released from link A*

E.3 Traffic directions on the network

The vehicle progression mechanism described above is invoked each time vehicles move from one link to another. In the program, there are two procedures which deal with the movement of vehicles through the street network. These are called *EastWest* and *NorthSouth*. Procedure *EastWest* deals with vehicles travelling eastwards through the network and westwards through the network. Procedure *NorthSouth* deals with northbound and southbound vehicles. The operation of each of procedures is described next.

Initially, the capacities of all intersections in the network are reviewed. Then, the demand at each entry point is refreshed using the algorithm described in Appendix E, Section E.1. The program then considers the east-west traffic phase and when complete, the north-south phase.

Introduction:

A particular intersection in the network can be referred to as an ordered pair (i, j) . The first coordinate represents the intersection in the EAST direction and the second coordinate represents the NORTH direction. Associate with each intersection pair (i, j) there will be queues corresponding to the EAST-WEST traffic phase and queues relating to the NORTH-SOUTH traffic phase. Thus, $E(i, j)$ represents the east-west bound queues waiting at intersection (i, j) and $N(i, j)$ describes the north-south bound queues waiting at intersection (i, j) . One can determine the direction of traffic using the following rule: For $E(i, j)$, if j is odd then traffic is moving to the east and

if j is even then it is going to the west. Similarly, for $N(i, j)$ if i is odd then traffic will be travelling northwards, whereas when i is even, traffic will be moving southwards.

Input:

- $E(i, j), N(i, j)$ - east-west and north-south queues on the network
- $cap(i, j)$ - capacities of each intersection
- $ban(i, j)$ - ban status of each link

Output:

- $E(i, j), N(i, j)$ - adjusted queues
- input statistics - total input during time slice
- output statistics - total leaving network during time slice

The *input* and *output* processes are identical for both procedures. However, the differences between the procedures arises in the direction of traffic movement which is explained next. We first describe the east-west movements and then the north-south ones.

Eastbound traffic movements:

For traffic travelling from west to east, the program processes traffic beginning at link $E(20, j)$ and proceeding towards link $E(1, j)$ where $j \in \{1 \dots 20\}$. There are a number of special conditions which may arise. These are

- (A) $j = 20$ and $odd(i)$ or $j = 1$ and $even(i)$
- (B) $i = 20$
- (C) IF $odd(i)$ THEN $inc = 1$ ELSE $inc = -1$

These conditions relate to traffic reaching the edge of the network. As a result, vehicles are sent *out of the grid*. This is denoted by o.o.g. The combined affect of these conditions is summarised in Table E.3.

$E(i, j)$	A	not A
B	ahead: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow \text{o.o.g}$
	turn: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow N(i, j + inc)$
not B	ahead: $E(i, j) \Rightarrow E(i + 1, j)$	$E(i, j) \Rightarrow E(i + 1, j)$
	turn: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow N(i, j + inc)$

Table E.3: *Eastbound traffic movements*

$E(i, j)$	A	not A
B	ahead: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow \text{o.o.g}$
	turn: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow N(i, j + inc)$
not B	ahead: $E(i, j) \Rightarrow E(i - 1, j)$	$E(i, j) \Rightarrow E(i - 1, j)$
	turn: $E(i, j) \Rightarrow \text{o.o.g}$	$E(i, j) \Rightarrow N(i, j + inc)$

Table E.4: *Westbound traffic movements*

Westbound traffic movements:

For traffic travelling westbound, the program processes traffic from link $E(1, j)$ up to link $E(20, j)$ where $j \in \{1 \dots 20\}$. Due to the network configuration, the conditions that may arise are slightly simplified and can be summarised as follows: may arise. These are

- (A) $j = 20$ and $odd(i)$
- (B) $i = 1$
- (C) IF $odd(i)$ THEN $inc = 1$ ELSE $inc = -1$

The combined effect of these conditions is summarised in Table E.4.

Northbound traffic movements:

For traffic travelling from south to north, the program processes traffic beginning at link $N(i, 20)$ and proceeding towards link $N(i, 1)$ where $i \in \{1 \dots 20\}$. As in the previous procedure a number of special conditions arise. These are

$N(i, j)$	A	not A
B	ahead: $N(i, j) \Rightarrow \text{o.o.g}$	$N(i, j) \Rightarrow \text{o.o.g}$
	turn: $N(i, j) \Rightarrow \text{o.o.g}$	$N(i, j) \Rightarrow E(i + inc, j)$
not B	ahead: $N(i, j) \Rightarrow N(i, j + 1)$	$N(i, j) \Rightarrow N(i, j + 1)$
	turn: $N(i, j) \Rightarrow \text{o.o.g}$	$N(i, j) \Rightarrow E(i + inc, j)$

Table E.5: *Northbound traffic movements*

(A) $i = 20$ and $odd(j)$ or $i = 1$ and $even(j)$

(B) $j = 20$

(C) IF $odd(j)$ THEN $inc = 1$ ELSE $inc = -1$

The combined affect of these conditions on the movement of northbound traffic is summarised in Table E.5.

Southbound traffic movements: For traffic travelling from north to south, the program processes traffic beginning at link $N(i, 1)$ and proceeding towards link $N(i, 20)$ where $i \in \{1 \dots 20\}$. As in the previous procedure a number of special conditions arise.

These are

(A) $i = 20$ and $odd(j)$

(B) $j = 1$

(C) IF $odd(j)$ THEN $inc = 1$ ELSE $inc = -1$

The combined affect of these conditions on the movement of southbound traffic is summarised in Table E.6.

E.4 Turn Bans

The installation of turn bans takes place as follows. Each road link is referenced by a pair of coordinates (i, j) . For example, consider the road link on the East/West

$N(i, j)$	A	not A
B	ahead: $N(i, j) \Rightarrow \text{o.o.g}$ turn: $N(i, j) \Rightarrow \text{o.o.g}$	$N(i, j) \Rightarrow \text{o.o.g}$ $N(i, j) \Rightarrow E(i + inc, j)$
not B	ahead: $N(i, j) \Rightarrow N(i, j - 1)$ turn: $N(i, j) \Rightarrow \text{o.o.g}$	$N(i, j) \Rightarrow N(i, j - 1)$ $N(i, j) \Rightarrow E(i + inc, j)$

Table E.6: *Southbound traffic movements*

New Ban Status	Old Ban Status		
	0	1	2
0	No Change	Ahead Only	Turn Only
1	Turn ban removed	No Change	Ahead only to Turn Only
2	Ahead ban removed	Turn only to Ahead Only	No Change

Table E.7: *Implementation of bans*

network whose coordinates are given as (7,10). The data structure associated with each link can be divided as follows. There are numbers which refer to the respective sizes of queues in the *turning* and *ahead* lanes as well as a number which describes the status of the ban on this link. Initially, the status of all bans is set to zero representing the fact that no restrictions exist on the network. However, the status of the ban can be changed during the simulation. There are a number of cases to consider. These have been summarised in Table E.7. The change in ban status invokes the program to re-organise the vehicles into appropriate positions. For example, when an *ahead only* ban is installed, all vehicles are stored in the ahead segregated area since no turning is allowed. Similarly, when a *turn only* ban is installed all vehicles are accommodated in the turn segregated area. When a ban is removed, the vehicles must first be divided according to fixed proportions into *ahead* and *turning* vehicles. These vehicles are then accommodated in the segregated zones. However, if these

totals exceed the capacity of the segregated areas, then a spillback will occur and the remaining vehicles are put in the reservoir area.

The *TURNBAN* procedure uses Table E.7 to install bans both at specific intersections as well as in grouped patterns. Two additional procedures have been developed to assist the user. These are the *CORE* procedure which targets gridlock cycles with turn bans and the *DIAMOND* procedure which produces a cordon of bans around the jam originating at the reference position specified by the user. The size of the cordon can be altered relative to the reference position.

E.4.1 The CORE Algorithm

The aim of the *CORE* procedure is to assist the user in specifying ban locations in particular patterns. By specifying the 'bottom-left' junction in a gridlock cycle, along with the phase direction (East or North) and the ban status (typically *ahead only*), the program then proceeds to install the bans for the entire cycle. The procedure can be summarised as follows:

Input: (i, j) , ban_type $\{0, 1, 2\}$, direction $\{E, N\}$

Output: Junctions comprising the gridlock cycle are banned

Procedure: The program considers the junction specification given by the ordered pair (i, j) . The procedure first tests the legality of the location and then proceeds to install the bans in the appropriate locations. Table E.8 summarises the results of these tests.

For example, the specification $(7, 10)$, East, ban = 1 will install turn bans in junctions $E(7,10)$, $N(8,10)$, $E(8,11)$, $N(7,11)$ as described in Table E.8.

E.4.2 The DIAMOND Algorithm

A second arrangement of bans which appears to be particularly effective is the installation of a cordon of ahead bans around the developing traffic jam. Because the

	<i>odd(i)</i>	<i>even(i)</i>
<i>odd(j)</i>	Illegal	Clockwise N(i,j), E(i,j+1) N(i+1,j+1), E(i+1, j)
<i>even(j)</i>	Anti-Clockwise E(i,j), N(i+1,j) E(i+1,j+1), N(i,j+1)	Illegal

Table E.8: *Implementation of CORE Algorithm*

East orientation	<i>odd(i+j)</i>	<i>even(i+j)</i>
<i>odd(j)</i>	1	2
<i>even(j)</i>	3	4

Table E.9: *East oriented diamond cordon situations*

shape of the emerging jam is predicted to be diamond-shaped, it is possible to deduce the locations of the bans at particular distances from the developing jam. To aid the user, the program automatically calculates these positions and installs the bans accordingly. An almost identical process is invoked in order to restore the network to its unrestricted conditions.

Input: (i, j) , size_parameter $\{2, 4, 6, 8\}$, direction E/N, ban_type $\{0, 1, 2\}$

Output: Diamond shaped cordon of East and North bans whose reference is given by (i, j) and whose extent is specified by the size_parameter.

Procedure:

There are a number of cases to consider. These are summarised in Table E.9 and E.10 respectively.

For each situation, we need to specify a number of parameters. These are (x, y) which represent the start position of the cordon as well as a number of matrices which are used to calculate the remaining locations on the cordon. These are known as *sstep* for the offsets and *xyincr* for the increments. These pairs of matrices are different

North orientation	$odd(i+j)$	$even(i+j)$
$odd(j)$	5	6
$even(j)$	7	8

Table E.10: *North oriented diamond cordon situations*

DIR	CASE	X	Y
$W \Rightarrow E$	1,2	$i-(1+size_p)$	j
$E \Rightarrow W$	3,4	$i+(1+size_p)$	j
$N \Rightarrow S$	5,6	i	$j-(1+size_p)$
$S \Rightarrow N$	7,8	i	$j+(1+size_p)$

Table E.11: *Starting point for cordon*

for each case. Table E.11 summarises the start positions for the cordons for each of the eight cases. Tables E.12 and E.13 provide the matrices required for each of the eight cases. The matrices are calculated on the basis of a dot-product. The results are provided to enable the user to understand how each of the cordons is traced. Following the presentation of these two tables, we describe a condensed form of the algorithm. The section concludes with an example of the application of the algorithm.

CASE	SSTEP	XYINCR
1	0 -1	1 -1
	1 0	1 1
	0 1	-1 1
	-1 0	-1 -1
2	0 1	1 1
	1 0	1 -1
	0 -1	-1 -1
	-1 0	-1 1
3	0 1	-1 1
	-1 0	-1 -1
	0 -1	1 -1
	1 0	1 1
4	0 -1	-1 -1
	-1 0	-1 1
	0 1	1 1
	1 0	1 -1

Table E.12: *Matrices for cases 1-4 required for offsets and increments*

CASE	SSTEP	XYINCR		
5	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix}$		
	6	$\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 \\ 1 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix}$	
		7	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -1 \\ -1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}$
			8	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}$

Table E.13: *Matrices for cases 5-8 required for offsets and increments*

PROGRAM DIAMOND

PROCEDURE SETUP

{ Set up starting line of cordon}

BEGIN

id = id mod size_p

jd = 1 + id mod size_p

x = x + SSTEP(jd,1)

y = y + SSTEP(jd,2)

dr = other(dr) % Changes direction drom E to N and vice versa

xinc = XYINCR(jd,1)

yinc = XYINCR(jd,2)

END

PROCEDURE INCREMENT

{ Calculate next junction on cordon line}

BEGIN

x = x + xinc

y = y + yinc

dr = other(dr) % Changes direction drom E to N and vice versa

END

BEGIN{DIAMOND}

id = 1

WHILE (id <= 4*size_p) DO

BEGIN

if (id mod size_p = 1) SETUP

CHECK EXISTENCE OF JUNCTION

SETBAN(x,y,bantype,dir)


```

dir=other(dir)
if (id mod size_p <> 0) INCREMENT
END
END

```

Example

Consider the diamond cordon of size four blocks, centered on the initial obstruction (7,10) with East orientation. This corresponds to case 3. Referring to Table E.11 the cordon start position can be calculated as

$$x = 12, y = 10$$

The application of PROGRAM DIAMOND will produce ahead bans at the following junctions:

```

id = 1
x = i, y = j + 1, xinc = -1, yinc = 1
N(12, 11)
E(11, 12)
N(10,13)
E(9,14) id mod 4 = 1

x = i - 1, y = j, xinc = -1, yinc = -1
N(8,14)
E(7,13)
N(6,12)
E(5,11) id mod 4 = 1

x = i, y = j - 1, xinc = 1, yinc = -1
N(5,10)
E(6,9)

```

N(7,8)

E(8,7) id mod 4 = 1

x = i + 1, y = j, xinc = 1, yinc = 1

N(9,7)

E(10,8)

N(11,9)

E(12,10)

In this example, bans will subsequently be installed at the junctions listed above. Similar arrangements of bans can be implemented using different combinations of the size parameter and the start position. The pattern described in Figure 5.5 of the thesis shows a scheme of ahead bans where the start position is given by E(7,10) and the size parameter is set at six blocks. This is known as the integrated control scheme and has been tested extensively in the thesis.

E.5 Summary

In this appendix, we described the main algorithms used in the simulation program. The following algorithms were considered: simulation of demand, the vehicle progression mechanism, the traffic directions on the network and the banning techniques. These help provide a flavour for some of the intricate modelling work which plays a central role in the simulation tool.