

# **Highway Asset Management**

By

Chao Yang

**A Doctoral Thesis**

Submitted in partial fulfilment of the requirements for the award of  
Doctor of Philosophy of the University of Nottingham

## **Abstract**

The aim of this thesis is to provide a framework for a decision making system to operate a highway network, to evaluate the impacts of maintenance activities, and to allocate limited budgets and resources in the highway network. This integrated model is composed of a network level traffic flow model (NTFM), a pavement deterioration model, and an optimisation framework.

NTFM is applicable for both motorway and urban road networks. It forecasts the traffic flow rates during the day, queue propagation at junctions, and travel delays throughout the network. It uses sub-models associated with different road and junction types which typically comprise the highway. To cope with the two-way traffic flow in the network, an iterative algorithm is utilised to generate the evolution of dependent traffic flows and queues. By introducing a reduced flow rate on links of the network, the effects of strategies employed to carry out roadworks can be mimicked. In addition, a traffic rerouting strategy is proposed to model the driver behaviour, i.e. adjusting original journey plans to reduce journey time when traffic congestion occurs in the road network. A pavement age gain model was chosen as the pavement deterioration model, which is used to evaluate the current pavement condition and predict the rate of pavement deterioration during the planning period. It deploys pavement age gain as the pavement improvement indicator which is simple and easy to apply. Moreover, the deterministic pavement age gain model can be transformed to a probabilistic one, using the normal distribution to describe the stochastic nature of pavement deterioration.

A multi-objective and multi-constraint optimisation model was constructed to achieve the best pavement maintenance and rehabilitation (M&R) strategy at the network level. The improved non-dominated sorting genetic algorithm (NSGA-II) is applied to perform system optimisation. Furthermore, the traffic operations on worksites, i.e. lane

closure options, start time of the maintenance, and traffic controls, are investigated so as to prevent, or at least to reduce, the congestion that resulted from maintenance and reconstruction works.

The case studies indicated that NTFM is capable of identifying the relationship between traffic flows in the network and capturing traffic phenomenon such as queue dynamics. The maintenance cost is reduced significantly using the developed optimisation framework. Also, the cost to the road users is minimised by varying the worksite arrangements. Consequently, the integrated decision making system provides highways agencies with the capability to better manage traffic and pavements in a highway network.

**Keywords:** NTFM, iterative algorithm, traffic rerouting strategy, optimisation, genetic algorithms, M&R strategy, NSGA-II, worksite arrangements

## **Acknowledgements**

First and foremost, I would like to express my sincere appreciation and gratitude to my supervisors, Prof. John Andrews, Dr. Rasa Remenyte-Prescott, and Dr. Matthew Byrne, for their inspiring guidance and encouragement throughout the research as well as for always being friendly and helpful.

I would like to extend my thanks to other members of Infrastructure Asset Management Group and the people of the Nottingham Transportation Engineering Centre for their assistance, encouragement and friendship.

Finally, I wish to thank my parents and sister for their endless inspiration, constant support and understanding during the research.

## Abbreviations

AADT	Annual Average Daily Traffic
ARRB	Australian Road Research Board
BGA	Binary-coded Genetic Algorithm
CTM	Cell Transmission Model
D	Diverge junction
D2AP	Dual 2 lane Carriageway
D3M	Dual 3 lane Motorway
GAs	Genetic Algorithms
HDM-4	Highway Development and Management model
LWR	Lighthill, Whitham and Richards
MG	Merge junction
MOGA	Multi-Objective Genetic Algorithm
M&R	Maintenance and Rehabilitation
NSGA	Non-dominated Sorting Genetic Algorithm
NTFM	Network level Traffic Flow Model
O-D	Origin-Destination
Ov	Overlay
PCI	Pavement Condition Index
PCU	Passenger Car Unit
PMS	Pavement Management System
PSI	Pavement Serviceability Index
R	Roundabout
Re	Resurfacing
RN	Roadwork Node

S	On-ramps and Off-ramps
S2	Single Carriageway
SD	Surface Dressing
SI	Signalized Intersection
SR	Signalized Roundabout
ST	Signalized T-junction
VEGA	Vector Evaluated Genetic algorithm

# Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Background.....	2
1.1.1	Highway Networks.....	2
1.1.2	Road Classes .....	2
1.1.3	Traffic Related Organisations .....	3
1.2	Objectives .....	3
1.3	Contributions of this Thesis.....	4
<b>2</b>	<b>Literature Review on Traffic Modelling.....</b>	<b>6</b>
2.1	Introduction .....	6
2.2	Traffic Models .....	6
2.2.1	Macroscopic Traffic Flow Models.....	7
2.2.2	Priority Junction Models .....	9
2.3	Traffic Rerouting Control Approaches.....	12
2.4	Summary.....	14
<b>3</b>	<b>Network Level Traffic Flow Model .....</b>	<b>16</b>
3.1	Introduction .....	16
3.1.1	Network Representation.....	17
3.1.2	Model Data.....	18
3.1.3	Main Principle of NTFM.....	19
3.2	Junction Sub-models .....	23
3.2.1	Signalized Intersection Model.....	24
3.2.2	T-junction Model .....	27
3.2.3	Signalized T-junction Model.....	31
3.2.4	Roundabout Model.....	32
3.2.5	Signalized Roundabout Model.....	43
3.2.6	On-ramp and Off-ramp Models.....	44
3.2.7	Merge and Diverge Models.....	46
3.2.8	Roadwork Node Model .....	48
3.3	Link Model .....	50
3.3.1	Calculation of Travel Duration .....	51
3.3.2	Model Formulation.....	51

3.3.3	Extension for Multi-lane Link.....	54
3.3.4	Extension for Roadwork .....	57
3.4	Highway Network Performance Metrics .....	58
3.5	Network Solution Routine .....	60
3.5.1	Convergence of In Profiles and Out Profiles.....	60
3.5.2	Forecast of Traffic Condition.....	61
3.6	Case Study .....	63
3.6.1	Network Simulation .....	64
3.6.2	Evaluation of Traffic Condition .....	64
3.6.3	Discussion .....	68
3.7	Traffic Rerouting Strategy .....	69
3.7.1	Introduction .....	69
3.7.2	Methodology .....	69
3.7.3	Case Study.....	75
3.8	Summary.....	82
<b>4</b>	<b>Literature Review on Highway Asset Modelling.....</b>	<b>83</b>
4.1	Introduction .....	83
4.2	M&R Operations for Pavement.....	84
4.2.1	Minor Maintenance .....	85
4.2.2	Surface Treatments.....	85
4.2.3	Major Maintenance .....	86
4.3	Decision Making Strategies for M&R.....	87
4.3.1	M&R Prioritisation Model .....	88
4.3.2	M&R Optimisation Model .....	90
4.4	Genetic Algorithms .....	94
4.4.1	Introduction .....	94
4.4.2	Simple Genetic Algorithms.....	95
4.4.3	Multi-objective Genetic Algorithms .....	101
4.5	Summary.....	112
<b>5</b>	<b>Scheduling of Pavement Maintenance .....</b>	<b>114</b>
5.1	Introduction .....	114
5.2	Scheduling of Pavement Maintenance using NSGA-II at the Network Level .....	115
5.2.1	Age Gain Maximisation Model.....	115
5.2.2	Classification of Road Sections.....	117



5.2.3	Model Formulation.....	118
5.2.4	Case Study.....	120
5.2.5	Probabilistic Pavement Deterioration Model .....	133
5.3	Scheduling of Pavement Maintenance using the NTFM at the Project Level.....	141
5.3.1	Methodology .....	142
5.3.2	Case Study.....	148
5.4	Maintenance Planning using the NTFM in the Long-term.....	156
5.5	Maintenance Planning by Balancing between Maintenance Costs and Road User Costs .....	159
5.6	Summary.....	162
<b>6</b>	<b>Model Application to the Loughborough-Nottingham Network.....</b>	<b>164</b>
6.1	Loughborough-Nottingham Highway Network .....	164
6.2	Model Calibration.....	168
6.2.1	Network Calibration.....	168
6.2.2	Available Traffic Flow Data .....	168
6.2.3	Movement of Traffic Flow.....	169
6.3	The Evaluation of Traffic Condition .....	170
6.3.1	Highway Network Performance under Normal Conditions .....	170
6.3.2	The Implementation of Roadworks.....	182
6.3.3	Discussion .....	188
6.4	Maintenance Planning for Road Sections in the Long-term.....	189
6.4.1	Dual Carriageway (MG2-R4).....	189
6.4.2	Single Carriageway (SR2-R4).....	194
6.4.3	Motorway (S4-S7).....	196
6.4.4	Dual Carriageway (SI1-ST2) .....	198
6.4.5	Motorway (S2-S3).....	200
6.4.6	Discussion .....	201
6.5	Maintenance Planning for the Road Sections in the Short-term.....	202
6.5.1	Possible Maintenance Arrangements on Dual Carriageway (MG2-R4) .....	202
6.5.2	Optimisation of Maintenance Arrangements on Dual Carriageway (MG2-R4) .	204
6.5.3	Optimisation of Maintenance Arrangements on Single Carriageway (SR2-R4)	207
6.5.4	Discussion .....	209
6.6	Summary.....	209
<b>7</b>	<b>Conclusions and Future Work.....</b>	<b>211</b>

7.1	Summary.....	211
7.2	Sensitivity of Main Assumptions .....	213
7.2.1	Junction model .....	213
7.2.2	Queue model .....	215
7.2.3	Pavement deterioration model.....	216
7.2.4	Road user cost .....	216
7.2.5	Maintenance implementation .....	217
7.3	Conclusions .....	217
7.4	Future Model Validation .....	218
7.5	Future Work.....	219
	<b>References .....</b>	<b>221</b>
	<b>Appendix A: Flow Diagram of NTFM .....</b>	<b>230</b>
	<b>Appendix B: Practical Implementation and Computational Time of the Decision Making System .....</b>	<b>233</b>

# 1 Introduction

Since the late-1980s, more than 90% of motorised passenger travel and around 65% of domestic freight in the UK have been delivered on the highway network. The total road length of highway networks in the UK was estimated to be 245,000 miles in 2010 [1]. Also, the total road length in the UK has increased by about 2,500 miles in the decade since 2000, approximately 1% per year. It demonstrates that the highways authorities have shifted its attention from construction of new roads to the maintenance and rehabilitation of existing ones. According to the Annual Local Authority Road Maintenance Survey [2], in England, the government has spent heavily on road maintenance in recent years. £2,240m was spent in 2006, £937m in 2007 and another £861m in 2008, an aggregate total of £6,867m from 2002 to 2008, in order to maintain the serviceability level of pavements. The survey also reports that a further £10.65b is currently required to bring the UK's roads to the desired standard. In addition to the expenditure on carrying out the work, the travel delay cost to the road users caused by maintenance is significant and expected to substantially exceed the corresponding cost of maintenance [3].

In this thesis a framework of decision making system is developed. Its purpose is to assist the highways agencies in operating the highway network, predicting traffic characteristics in the network, preventing, or at least reducing, traffic congestion, maintaining the pavement condition at a serviceable level, identifying the maintenance requirements, and minimising both maintenance and road user costs.

## 1.1 Background

### 1.1.1 Highway Networks

The highway network in the UK is comprised of motorways, 'A' roads, rural minor roads and urban minor roads. Motorways are usually used by long-distance traffic, while 'A' roads are used by both medium-distance and long-distance traffic. As for the short-distance traffic, urban minor roads and rural minors are employed. Motorways and 'A' roads occupied 1% and 12% respectively of the total road length of the highway network in the UK, while rural minor roads and urban roads accounted for the remaining 54 % and 33%. Nonetheless, 19.8% of the total transportation volume was conveyed on motorways and 44.3% on 'A' roads.

### 1.1.2 Road Classes

To distinguish the purposes and functions of different kind of roads, the road classes adopted in UK are listed in Table 1-1 [4]:

Table 1-1: COBA Road classes

Road class	Description	Speed limit (mph)
1	Rural single carriageway	70
2	Rural all-purpose dual 2 lane carriageway	70
3	Rural all-purpose dual 3 or more lane carriageway	70
4	Motorway (urban or rural), dual 2 lanes	70
5	Motorway (urban or rural), dual 3 lanes	70
6	Motorway (urban or rural), dual 4 or more lanes	70
7	Urban road, Non-central, single or dual carriageway	30
8	Urban road, Central, single or dual carriageway	30
9	Small town road, single or dual carriageway	30 or 40
10	Suburban main road, single carriageway	40
11	Suburban main road, dual carriageway	40

The COBA (COst Benefit Analysis) software can be used to compare the costs of road projects with benefits derived by road user costs. Classes 1 to 6 are recognised as all-purpose roads ('A' roads) and motorways that are generally restricted by the maximum speed, 70 mph. The traffic flow model developed in this thesis works with average speeds which are not allowed to exceed the specified speed limits. Further, the flow capacity for each road class as stated in Table 1-1 is depicted in Table 1-2 [5]:

Table 1-2: Default road capacity

Road class	Capacity (pcu/hr/standard lane)
1	1400
2 and 3	1800
4, 5 and 6	2000
7, 8, 9 and 10	1400
11	1800

1. "pcu" stands for passenger car unit, which is used to express highway capacity. For example, one car is recognised as a single unit, motorcycle is considered as half unit, heavy vehicles are considered as more than 3 units.

### 1.1.3 Traffic Related Organisations

In England, the Highways Agency (HA) is responsible for managing motorways and trunk roads, whereas other roads are run by the local authorities, e.g. Nottingham County Council. While in Scotland and Wales, roads are operated by Transport Scotland and the Welsh Assembly Government, respectively. As for Northern Ireland, roads are the responsibility of Roads Service Northern Ireland.

## 1.2 Objectives

The aim of this thesis is to develop an integrated decision making system to forecast and operate traffic flows, evaluate the impacts of different maintenance activities, and allocate limited budgets and resources in a highway network. This system will be

utilized to facilitate traffic on the network and to maintain the pavement service level to ensure optimal performance. To fulfil this aim, the following specific objectives are considered:

- To construct a macroscopic traffic flow model for identifying the traffic characteristics in a highway network.
- To develop a traffic rerouting strategy based on the proposed traffic model for modeling driver behavior when traffic congestion takes place in the highway network, i.e. divert traffic heading to the congested areas to take alternative routes and avoid traffic delays.
- To select an appropriate pavement deterioration model to evaluate pavement condition during the planning period.
- To derive all combinations of maintenance and rehabilitation actions that can be applied to the network and model their effects.
- To develop a genetic algorithm-based optimization technique for pavement maintenance management. The optimization technique will be used to achieve the optimal pavement maintenance and rehabilitation strategy with the purpose of minimising maintenance cost and user cost and maximising pavement condition.

In summary, this thesis provides a framework and the sub-models to improve network performance, enhance pavement serviceability, and minimise maintenance cost and road user cost.

### **1.3 Contributions of this Thesis**

The most significant contributions of the research studied in this thesis are:

- A novel network level traffic flow model (NTFM) is built and the software developed that is applicable for an integrated motorway and urban road network, in which two-way traffic flow is predicted using an iterative simulation method. More junction types are taken into account by NTFM than the existing integrated traffic flow models.
- A roadwork node sub-model is introduced to NTFM, which is used to evaluate the traffic conditions in a highway network under normal conditions and different maintenance scenarios.
- A traffic rerouting strategy is developed to model driver behavior when traffic congestion takes place in the highway network.
- An optimisation technique is built that allows for the selection and scheduling of maintenance and rehabilitation operations for road sections on a highway network.
- NTFM is utilised to optimise maintenance cost and road user cost together with pavement deterioration models in both long-term and short-term, where Genetic Algorithms are employed to perform optimisation.
- Overall, the capability of the model is to help highway engineers to make more effective decisions in terms of the limited annual budgets as well as the frequencies of maintenance activities.

## **2 Literature Review on Traffic Modelling**

### **2.1 Introduction**

The modelling of traffic flow in highway networks is of vital importance for pavement management, providing information for traffic control, traffic flow prediction, traffic diversion, and the implementation of roadwork. In this chapter, the literature on the traffic flow models will be discussed first. Since traffic rerouting techniques are for reducing excessive congestion, and, making more efficient use of the existing roads, the second part of the literature review will be devoted to that area.

### **2.2 Traffic Models**

Typically, traffic flow models are categorised into two main groups: macroscopic models and microscopic models. Macroscopic traffic models are used to identify the aggregate behaviour of sets of vehicles, easy to validate and ensure a good real-time quality, such as the fluid-dynamic traffic models [6, 7]. Microscopic models are applied to model the travel behaviour of an individual vehicle which is recognised as a function of the traffic conditions in its environment [8, 9]. As drivers' behaviour in real traffic is difficult to observe and measure, microscopic models are difficult to validate accurately [10]. In addition, the computational effort required by microscopic models is significantly higher than that for macroscopic models and the data required by microscopic models are harder to get. As for macroscopic models, they do not distinguish their components flows by origins and destinations; therefore when the traffic stream arrives at junctions, macroscopic models usually assign fixed turning ratios for the traffic stream [10]. In this thesis, focus is on macroscopic models.



### 2.2.1 Macroscopic Traffic Flow Models

During the past decades, a number of macroscopic traffic models have been constructed to identify traffic behaviour on motorways and urban roads. Lighthill and Whitham [6] and Richards [7] provided a pioneering flow-dynamic model based on first-order differential equations, termed the LWR model, which was the first model used to describe unidirectional traffic flow on highway networks.

As an extension to LWR model, Payne [11] developed a second-order model [12] based on the car-following model [13] that considered the driver's reaction time. This results in a dynamic mean speed equation rather than the static one applied in first-order models, in which the dynamic flow phenomena, i.e. emergent traffic congestion and stop-and-go traffic, are modelled. Because first-order models and second-order models are mainly focused on the traffic characteristics on road links, they would only be sufficient for signalized networks where the traffic interaction among flows from competing arms is eliminated.

Daganzo [10] developed a cell transmission model (CTM) that adopted a convergent approximation to the LWR model to evaluate the traffic on a highway network with a single entrance and exit. Afterwards, Daganzo [14] applied merge cell and diverge cell in a CTM to generalise the junctions within a highway network. CTM can be used to reproduce kinematic waves, the formation and dissipation of a queue in an explicit manner, in both congested and uncongested regimes, and to model the traffic movement on a simple highway network with three-legged junctions. Lo [15] introduced signal control to CTM, and Lo et al. [16] further developed signalized merge and diverge cells in CTM to broaden its applicability. However, priority junctions, i.e. T-junction and roundabout, are not considered, so the conflicting requirements of the traffic flows from

competing directions cannot be captured. Thus it is only suitable for signalized networks.

Messmer and Papageorgiou [17] considered the situation of a motorway network based on the second-order traffic flow model, METANET model. The approach has been widely applied to simulate traffic flow phenomena on motorway networks of arbitrary characteristics, including motorway links, on-ramps and off-ramps (slip roads). Other research has been performed to improve METANET, the adoption of variable speed limits on motorways is described in Breton, et al. [18] and Hegyi, et al. [19], and the application of route guidance is depicted in Deflorio [20] and Karimi et al. [21]. Subsequently, Van den Berg et al. [22] developed an integrated traffic control capability for mixed urban and motorway networks, as motorway traffic is heavily influenced by the traffic flows on the connected urban roads, and vice versa. This model is composed of the METANET model that was used to evaluate motorway traffic and a queue length model based on the Kashani model [23] for urban traffic, coupled via on-ramps and off-ramps. However, these studies referred to METANET do not identify the traffic interaction among competing flows, since the inflow and outflow for each node in METANET is only characterised by turning ratios of traffic flows from each link that is connected to the node. In Van den Berg et al. [22], the traffic in the major road and the slip road for an on-ramp are treated separately. They need to be addressed dependently since traffic from slip road is also restricted by the traffic on the major road. Also, the updated urban traffic model assigns sub-queues for each turning direction on road link; however, shared lanes are not taken into account where traffic heading to different directions might be mixed together.

## 2.2.2 Priority Junction Models

Many different approaches for evaluation of traffic at unsignalized intersections have been presented and investigated, including gap acceptance theory, and queuing theory.

### 2.2.2.1 Gap Acceptance Theory

In traditional gap acceptance models, it has been assumed that the vehicles in the major stream have absolute priority over the vehicles in the minor stream. It means that the major stream is unaffected by the minor stream vehicles. Within unsignalized intersection theory, it is also assumed that drivers are both consistent and homogenous [24]. A consistent driver is assumed to act the same way every time in all similar situations. For a homogeneous traffic stream, all drivers are supposed to behave in the same way.

In the gap acceptance theory two major parameters are investigated, critical gap and follow-up time. The critical gap is the minimum gap for drivers in the minor stream to accept and enter the intersection, any shorter gap will be rejected and any longer gap accepted. The follow-up time is the headway between two departing minor stream vehicles in very long gaps. Many techniques have been applied to estimate the critical gaps at unsignalized intersections. Some of the important methods are examined in [24] and [25], and a set of quality criteria has been formulated by which the usefulness of each method is appraised [25]. Furthermore, Brilon et al. [25] found that the maximum likelihood method [26] and Hewitt's method [27] are superior to other methods for estimating the critical gap.

In addition, the capacity at the unsignalized intersection is very important. Capacity on the minor road is determined as the maximum number of minor stream vehicles that can cross the intersection during a unit time under predefined conditions. In 1973, Siegloch [28] developed the fundamental capacity equation of the minor road. Based on this

equation, numerous capacity formulas are computed using different headway models [24]. The distribution of headway is essential for the calculation of capacity at unsignalized intersections, including negative exponential distribution (M1), shifted exponential distribution (M2) and dichotomized distribution (M3) [29]. The M3 model allows a portion of the vehicles to be bunched and the remaining vehicles are free vehicles that move without interacting with the front vehicles, where the latter is characterised by a shifted exponential distribution. A more practical M3 model is proposed by Cowan [29], which does not attempt to model the bunched vehicles as their headways are not accepted by minor stream vehicles but rather model the larger gaps between the free vehicles. Based on Cowan's M3 headway model, Plank and Catchpole [30] derived the corresponding capacity formula using Siegloch's capacity equation [28].

Kimber [31] found that the vehicles in the minor stream could also affect the vehicles in the major stream, especially under high flow circumstances. Thus a new gap acceptance model based on limited priority for the major stream was proposed by Troutbeck and Kako [32], in which the major stream vehicles are expected to be slightly delayed so as to accommodate the merging minor stream vehicles. The limited priority gap acceptance model has been applied to evaluate the merging behaviours at roundabouts [32] and freeway on-ramps [33].

#### 2.2.2.2 Queuing Theory

Starting from the capacity at unsignalized intersections, further traffic parameters, which represent the quality of traffic operations, can be evaluated. Queuing theory is generally used to evaluate situations which involve average delays, average queue lengths, distribution of delays and distribution of queue lengths. In 1962, Tanner [34] established the equations for the average delays for minor stream vehicles at

unsignalized intersections with only one major stream and one minor stream. This kind of intersection belongs to the M/G2/1 queuing system [35], M represents the traffic flow arrival pattern of the minor stream, i.e. exponentially distributed headway; G is service time, i.e. the time spent in the first position of the queue. Two types of service times are involved, one is the service time for vehicles entering the empty system and the other is the service time for vehicles joining the queue when other vehicles are already queuing. Finally “1” stands for one service facility, i.e. one service lane for the minor stream. Comparable solutions to the M/G2/1 queuing system have been introduced by Yeo and Weesakul [36] and Kremser [37, 38], who [37] proposed the formulae for the expectations of the two service times. Afterwards, Daganzo [39] and Poeschl [40] have derived new formulae to improve Kremser’s approach. Based on queuing theory, Ning [41] proposed a universal procedure for calculating the capacity for the M/G2/1 queuing system under different predefined conditions, extending the method to cope with the queuing system with more than one major stream.

#### 2.2.2.3 Discussion

One disadvantage for the gap acceptance models is that they have failed to capture conflicts among the major streams. For instance, the right-turning vehicles (for left-side driving, i.e. in the UK) in the major stream have to give way to the vehicles going ahead from the opposing direction, resulting in a queue forming on the major road. It also led to the variation of the headway distributions on the major road so that the original gap acceptance criteria no longer applies [42]. As a result, gap acceptance models are only sufficient for unsignalized intersections under simpler conditions.

In a mixed road network, adjacent signalized intersections can have a significant impact on capacity and performance of unsignalized intersections [43]. On one hand, signalized intersections can group the vehicles into a queue during the red phases, and then result

in a cyclic recurrence of a set of long headways (between vehicles departing freely during the green phase) and a set of short headways (between vehicles departing from the queue), thus it is impractical to model the recurrence of headways with gap acceptance theory [43]. On the other hand, signalized intersections can cause the vehicles to arrive at the downstream intersections in platoons, while the gap acceptance models can be applied only when the platoon does not exist [44]. This is because the headways among a platoon are supposed to be shorter than the critical gap that led to no entry for minor stream vehicles. Consequently, the traditional gap acceptance models are not readily applied in network level analysis.

Moreover, the critical gap is difficult to determine and implement for some special conditions, including a two-stage gap-acceptance process, downstream queue spill back, and driver lane preference [26]; and gap acceptance models are not accurate for modelling directional flow [45].

Compared to gap acceptance theory, queuing theory is a more abstract technique for describing driver departure patterns, and it can be more easily applied to measure the delays at more complicated unsignalized intersections [41]. However, it is also constructed based on headway distribution models, thus it suffers the same drawbacks of the gap acceptance theory that resulted from the variation of headway distributions. As gap acceptance theory and queuing theory mainly focused on investigating the traffic at a single intersection, and not accurate for modelling directional flow [45], they are not capable of identifying traffic characteristic at the network level.

### **2.3 Traffic Rerouting Control Approaches**

A number of techniques have been developed previously for routing control in transportation networks: so as to alleviate traffic congestion, which involve user equilibrium assignments and control strategies that are constructed by combining

optimal control theory and macroscopic traffic flow models. Messmer and Papageorgiou [46] applied a nonlinear optimisation approach based on the METANET model [17] to handle the route guidance problem in motorway networks, a parallel solution derived in terms of user equilibrium principles [47] has been proposed by Wie et al. [48]. Another method to this problem of integrated control is the application of a linear programming approach as described in Papageorgiou [49] where both motorways and signal-controlled urban roads are considered. Iftar [50], [51] also presented a linear optimisation approach based on a decentralized routing controller to reduce queue build-up for congested highways, the deployed routing controller is decentralized in the manner that the computations for each node on the network are done locally without any information transfer from any other nodes. Van den Berg et al. [22] formulated a model predictive control approach for mixed motorway and urban networks, based on the METANET model [17], Kashani urban network model [23] and improved on-ramp and off-ramp models. A practically integrated control model when both a motorway and a parallel arterial are included in the network was proposed by Chang et al. [52], [53]. On the basis of the previous studies [52-55], Yue et al. [56] proposed an integrated control approach for mixed road networks, in which traffic flow evolution on on-ramps, off-ramps and surface streets, as well as queue propagation, have been explicitly identified. However, this integrated control is only applicable for motorway and signal-controlled urban network, as the effect of conflicting flow from competing routes on the traffic demand of road link is not taken into account.

In addition, some of the existing traffic control strategies are implemented using origin-destination (O-D) trip matrix, which are exposed to the difficulty of collecting accurate O-D information. The reliability of the obtained O-D trip matrix and computational complexity are the two main disadvantages hindering such strategies from being applied

broadly. Furthermore, the traffic flow models themselves are very complicated, which makes these control strategies hard to reach the global optimality [57].

## **2.4 Summary**

The literature review showed that there are numerous different techniques for predicting the traffic condition in the network, but that the following deficiencies exist:

- a. A review of the existing traffic flow models indicated that CTM and METANET are widely adopted in traffic management. However, both the models failed to capture the traffic movement taking place at the priority junction.
- b. Only limited junction types are taken into account by CTM and METANET, it means that the traffic interaction experienced at some junction types, i.e. signalized T-junction and roundabout, cannot be identified, especially not suitable for the integrated motorway and urban road network that composed of different kinds of junctions.
- c. As for priority junctions, the most common methods are gap acceptance theory and queuing theory. They have been applied to investigate the traffic at different kinds of priority junctions. Nevertheless, they become inefficient when modelling directional flow and identifying traffic behaviours at the network level.
- d. Based on the proposed traffic models, numerous traffic rerouting control measures have been developed. Most of the control measures take advantage of O-D matrix to describe the modification of journey plans, which require abundant information on individual journey plans.

Considering the traffic interaction at both signalized and priority junctions, this thesis describes a macroscopic traffic flow model the purpose of which is to provide a method for predicting the traffic flow and travel delay for each junction in the network. One



feature of this model is that both motorway junctions and urban junctions are evaluated based on the principle of a maximum capacity flow rate at these junctions where flows compete. Another feature is that this model works with the observed junction turning ratios rather than O-D matrix. In addition, queue propagation is used to model traffic congestion through the network.

## **3 Network Level Traffic Flow Model**

### **3.1 Introduction**

This chapter develops a network level traffic flow model (NTFM) which is applicable for an integrated motorway and urban road network. It forecasts the traffic flow rates, queue propagation at the junctions and travel delays through the network. NTFM uses sub-models associated with all road and junction types which comprise the highway. The principles involved in the modelling methodology are explained and a detailed description given for the signalized intersection, T-junction and roundabout sub-models provided. These are typical of the unit models developed and demonstrate how the traffic flow and queuing is calculated at the junctions. Where the flow to a junction exceeds the capacity of the network then a queue forms and the propagation of this queue back through the network will impact upon the flow achieved at other junctions. The flow at any one part of the network is obviously therefore very dependent upon the flows at all other parts of the network. To predict the two-way traffic flow in NTFM, an iterative simulation method is executed to generate the evolution of dependent traffic flows and queues. Moreover, instead of applying an O-D trip matrix this model takes advantage of the observed traffic flow turning ratios at junctions. It should be noted that the accuracy of the results of this model will depend critically on the validity of the monitored turning ratios. But it is considered that this information is more practically obtained than individual journey plans.

To demonstrate the capability of the model it is applied to a case study network. The results indicated that NTFM is capable of identifying the relationship between traffic flows and capturing traffic phenomena such as queue dynamics. One of the features which represents the capability of a road network is the traffic flow characteristics. By

introducing a reduced flow rate on links of the network then the effects of strategies employed to carry out roadworks can be mimicked. Different traffic management schemes will result in different flow rates past the active work. By comparing the whole network flow and queue characteristics under alternative maintenance strategies the best way of managing the highway repairs, minimising the disruption caused to the road users can be established. Furthermore, a traffic rerouting strategy is incorporated to model journey behaviour when traffic congestion takes place in the highway network.

### 3.1.1 Network Representation

The road network studied in NTFM consists of nodes and links. Links are used to represent roads, i.e. motorway links and urban road links, and nodes are junctions, including signalized intersections and roundabouts, etc. Moreover, parts of the same road with different characteristics such as flow capacity are separated by a node (e.g. when a dual 2 lane carriageway reduces to a single carriageway). Prior to the evaluation of the road network, the relationships between traffic flows and queue build up at the network need to be identified. Models for junctions have links which enable the exit traffic from one junction to enter the second junction; this will work in two directions as for these two junctions two-way traffic flow is deployed. In this way, all the junctions in the network are connected to each other.

In terms of the network flow theory a network can have a number of source nodes and sink nodes, where a source node defines the flow into the network and a sink node defines the flow out of the network. Source and sink nodes can be used to model the edges of the network or include the rest of the network in the model of a sub-network. In addition, the links themselves can have source and sink nodes, which are used to model cumulative traffic entering/leaving the link. This can represent significant traffic

flows to/from the network from such elements as housing estates, airports, railway stations or places of employment. In this way it is possible to avoid the inclusion of all minor roads on the network.

### 3.1.2 Model Data

Data is needed to describe each link by the flow capacity and the link capacitance and each node, which represents a junction, the flow capacity through the node. Such data can be derived from the system description or recording flow data at different times during the day. Data necessary for the modelling is described below:

- $c_{i,j}$  - flow capacity on the link between nodes  $i$  and  $j$  (pcu/hr), where pcu is passenger car unit
- $cp_{i,j}$  - link capacitance, i.e. the maximum number of cars which can queue on the link from node  $j$  (pcu), which is obtained based on the full length from junction  $j$  to junction  $i$
- $sr_{i,j}(t_k)$  - source flow entering the link in time  $t_k$ , for example, cumulative traffic joining the main road from an estate (pcu/hr)
- $sk_{i,j}(t_k)$  - sink flow leaving the link in time  $t_k$ , for example, cumulative traffic leaving the main road for an estate (pcu/hr)
- $d_{i,j,l}(t_k)$  - proportion of flow on the link choosing the outflow direction  $l$ ,  $l$  is expressed either in direction left, right, ahead or in the ID of the destined node, i.e.  $j+1$
- $p_{i,j}(t_k)$  - proportion of flow leaving the motorway link in time  $t_k$

Some data depends on the system structure, for example, the link capacitance, which is described by the room for the queue on the link, and is time independent. Some data

depends on time, for example, the proportion of flow travelling to a certain direction, and can be different throughout the day.

Additional data needed for nodes describing the different types of junction is described in the separate junction models.

### 3.1.3 Main Principle of NTFM

Three main variables are defined in the model:

- $f_{i,j}(t_k)$  - flow on the link in time  $t_k$  (pcu/hr)
- $q_{i,j}(t_k)$  - average number of vehicles queuing on the link in time  $t_k$  (pcu)
- $q_i(t_k)$  - average number of vehicles propagating back to the upstream links of node  $i$  in time  $t_k$  (pcu)

NTFM is constructed based on the principle of the queue model. Firstly, flows from the network source nodes are passed through the network to all the sink nodes, calculating the flow on each link,  $f_{i,j}(t_k)$ . Then the flow on each link is compared with the flow capacity of the link, applying the general equations for the queue on the link and the models for the different types of junction, and the queue is calculated,  $q_{i,j}(t_k)$ . If the queue exceeds the link capacitance, effects of the queue are propagated back through the network,  $q_i(t_k)$ . Finally, in the following time steps, different flows from the network source nodes are propagated through the network, to represent situation such as the rush hour, and their effects are added to the queues present on the network from the previous time steps. If the flow through the network improves, for example, traffic flow rates from the source nodes decrease or traffic lights are adjusted to allow a better flow through the congested links, the queues can decrease and eventually the links can become clear of queues. In this manner the traffic characteristic for a given highway

network throughout a day can be identified by NTFM. Detailed rules for calculating flows and queues on the link are described in the following section.

### 3.1.3.1 Main Equations

Flow on the link  $i$ - $j$  in time  $t_k$  is calculated as a sum of all the flows to node  $i$ , the flow entering the link and the negative flow leaving the link:

$$f_{i,j}(t_k) = \sum_{\text{all } s} d_{s,i,j}(t_k) f_{s,i}(t_k) + sr_{i,j}(t_k) - sk_{i,j}(t_k) \quad (3-1)$$

Once the flow on each link in time  $t_k$  is calculated (for the circumstances that there are no restrictions), the flow value and the queue value on the link  $i$ - $j$  might need to be updated according to the flow capacity on this link  $c_{i,j}$  and the link capacitance  $cp_{i,j}$ .

The updated flow is expressed as  $f'_{i,j}(t_k)$  and the updated queue is expressed as  $q'_{i,j}(t_k)$ .

$\Delta t$  is equal to  $t_k - t_{k-1}$ .

Three cases are considered:

- Flow on the link is higher than the flow capacity and there is no queue on the link in time  $t_k$ :

$$\text{If } f_{i,j}(t_k) > c_{i,j}, \text{ and } q_{i,j}(t_k) = 0, \text{ then} \quad (3-2)$$

$$f'_{i,j}(t_k) = c_{i,j} \text{ and}$$

$$q'_{i,j}(t_k) = (f_{i,j}(t_k) - c_{i,j}) \cdot \Delta t \quad (3-2^a)$$

$$\text{if } q'_{i,j}(t_k) > cp_{i,j}, \text{ then } q'_{i,j}(t_k) = cp_{i,j}, \text{ and } q_i(t_k) = (f_{i,j}(t_k) - c_{i,j}) \cdot \Delta t - cp_{i,j} \quad (3-2^b)$$

- Flow on the link is higher than the flow capacity and there is a queue on the link in time  $t_k$ :

$$\text{If } f_{i,j}(t_k) > c_{i,j}, \text{ and } q_{i,j}(t_k) > 0, \text{ then } f'_{i,j}(t_k) = c_{i,j} \quad (3-3)$$

$$q'_{i,j}(t_k) = q_{i,j}(t_k) + (f_{i,j}(t_k) - c_{i,j}) \cdot \Delta t \quad (3-3^a)$$

if  $q'_{i,j}(t_k) > cp_{i,j}$ , then  $q'_{i,j}(t_k) = cp_{i,j}$  and

$$q_i(t_k) = q_{i,j}(t_k) + (f_{i,j}(t_k) - c_{i,j}) \cdot \Delta t - cp_{i,j} \quad (3-3^b)$$

- Flow on the link is lower than the flow capacity and there is a queue on the link in time  $t_k$ :

If  $f_{i,j}(t_k) \leq c_{i,j}$ , and  $q_{i,j}(t_k) > 0$ , then  $f'_{i,j}(t_k) = c_{i,j}$  and

$$(3-4)$$

$$q'_{i,j}(t_k) = q_{i,j}(t_k) + (f_{i,j}(t_k) - c_{i,j}) \cdot \Delta t \quad (3-4^a)$$

if  $q'_{i,j}(t_k) < 0$ , then  $f'_{i,j}(t_k) = c_{i,j} + \frac{q_{i,j}(t_k)}{\Delta t}$  and  $q'_{i,j}(t_k) = 0$

$$(3-4^b)$$

In the third case, described above, when the flow is less than the flow capacity and a queue is present, the link can be cleared of the queue, depending on the size of the queue and the difference between the flow capacity and the flow. The queue becomes zero, if the difference between the flow capacity and the flow is not greater than the queue size. When the size of the queue is smaller than the difference between the flow capacity and the flow, the flow value is adjusted by the size of the queue, as shown in Equation 3-4<sup>b</sup>.

The flow and the queue on the link with a junction are updated according to the appropriate model of the junction, when the flow capacity for the node is used in the model. More complexity is introduced in the junction models, when separate lanes are modelled on the links and the flow capacity and the capacitance on each lane are considered.

### 3.1.3.2 Queue Propagation back through the Network

Once the queue is larger than the link capacitance, as described in the Equations 3-2<sup>b</sup> or 3-3<sup>b</sup>, the queue at the end of the link,  $q_i(t_k)$ , is passed back to the connecting network, i.e. to the links that contributed to the build-up of the queue. This is done using the queue propagation algorithm. The general idea is that a proportion of the queue is passed to each link that contributed to the build-up of the queue. The proportion of the queue for each link is calculated as the proportion of the flow from that link contributing to the overall flow. For example, if a queue builds up on the link from  $j$  to  $j+1$  and it exceeds the capacity of the link by the number of vehicles  $q_j(t_k)$ , it is proportionally distributed back to all the links that enter node  $j$ . This process is going to increase the size of the queue and decrease the flow on each link that enters node  $j$ :

$$q'_{i,j}(t_k) = q_{i,j}(t_k) + \frac{f_{i,j}(t_k)}{f_{j,j+1}(t_k)} \cdot q_j(t_k) \quad (3-5)$$

$$f'_{i,j}(t_k) = f_{i,j}(t_k) - \frac{f_{i,j}(t_k)}{f_{j,j+1}(t_k)} \cdot \frac{q_j(t_k)}{\Delta t} \quad (3-6)$$

If after this process the size of the increased queue,  $q'_{i,j}(t_k)$ , exceeds the capacity of the link,  $cp_{i,j}$ , the effects of the queue are passed back further through the network until a queue can be accommodated and does not exceed the capacity of the link. For example,

$$\text{if } q'_{i,j}(t_k) > cp_{i,j}, \text{ then } q_i(t_k) = q'_{i,j}(t_k) - cp_{i,j} \text{ and } q'_{i,j}(t_k) = cp_{i,j} \quad (3-7)$$

$q_i(t_k)$  is passed to the upstream links that enter node  $i$ , etc.



### 3.1.3.3 Queue Update in Time

If the queue is present in time  $t_k$ , i.e.  $q'_{i,j}(t_k) > 0$ , it is also present at the beginning of the modelling step  $t_{k+1}$ , i.e.  $q_{i,j}(t_{k+1}) = q'_{i,j}(t_k)$ , which then depends on the flow in time  $t_{k+1}$ ,  $f_{i,j}(t_k + 1)$ , and the relevant equations are applied to update the flow and the queue on the link as necessary.

## 3.2 Junction Sub-models

In addition to the basic link model, the sub-models for each junction type are constructed to express the traffic interaction at junctions. The junction types studied in this model are listed in Table 3-1.

Table 3-1: Junction types in NTFM

<i>Junction groups</i>	<i>Junction types</i>		
Signalized Junctions	Signalized T-junction	Signalized Intersection	Signalized Roundabout
Priority Junctions	T-junction	Urban Roundabout	Motorway Roundabout
One-way Junctions	On-ramp and Off-ramp	Merge and Diverge	Roadwork node

The traffic flow at a signalized junction is influenced by both the flow capacity of the entry arm and the green split time of the traffic signals (proportion of times the signals gives priority to flow in its direction), where the conflicts among competing traffic flows are eliminated owing to the application of traffic lights. For the group of one-way junctions where (except for the on-ramp of motorways), the entering traffic for the one-way junction is only characterized by the corresponding flow capacity. The on-ramp is also evaluated as a priority junction. For priority junctions, the traffic flow is based on

right-of-way rules, where the entering traffic flow for each arm of the junction is restricted by the flow capacity and also by the traffic flows from competing arms. The underlying methodologies for the each junction type stated above are described in detail to explicitly demonstrate these concepts.

### 3.2.1 Signalized Intersection Model

In terms of Figure 3-1, each road to the intersection has two lanes, lane 1 is used for going straight on and turning left and lane 2 for turning right. Assume that the traffic lights are on green for lane 1 and for lane 2 for different lengths of time, but they turn to red for both lanes at the same time.

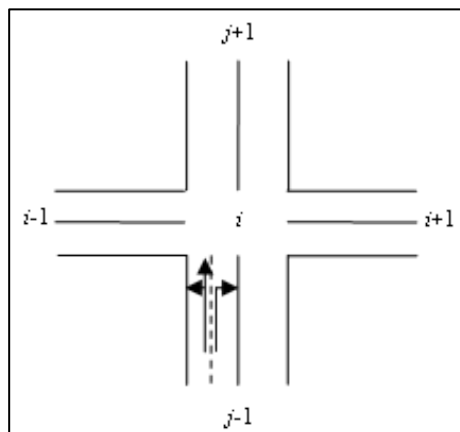


Figure 3-1: Signalized Intersection

#### 3.2.1.1 Junction Specific Data

- $c_i$  - flow capacity for junction  $i$ , depending on gaps between vehicles and vehicle speed (pcu/hr)
- $c_{j-1,i,l}$  - flow capacity for lane  $l$  on the link between nodes  $j-1$  and  $i$  (pcu/hr), depending on gaps between vehicles and vehicle speed (pcu/hr)
- $g_{j-1,l}(t_k)$  - proportion of time on green in lane  $l$  from direction  $j-1$

$cp_{j-1,i,l}$  link capacitance in lane  $l$ , i.e. the maximum number of cars which can queue in lane  $l$  of the link (pcu)

### 3.2.1.2 Model

Since the intersection is controlled by traffic lights, the node flow capacity in the lane is described as a product of the time proportion on green in the lane and the node flow capacity. For example, the flow capacity for node  $i$  in lane  $l$  from direction  $j-1$  is calculated as:

$$c_{j-1,i,l}(t_k) = g_{j-1,l}(t_k) \cdot c_i \quad (3-8)$$

According to the junction description, flow in the lane is described as the appropriate proportion of the flow in the lane. For example, the flow in lane 1 and lane 2 from direction  $j-1$  is described as:

$$f_{j-1,i,1}(t_k) = (d_{j-1,i,j+1}(t_k) + d_{j-1,i,i-1}(t_k)) f_{j-1,i}(t_k) \quad (3-9)$$

$$f_{j-1,i,2}(t_k) = d_{j-1,i,i+1}(t_k) f_{j-1,i}(t_k)$$

A queue might build up in lane 1 and/or lane 2, depending on flow capacity in the lane and lane capacitance. Similarly to the general rules of the queue calculation on the link, three cases are considered:

- Flow on the link in lane  $l$  is higher than the flow capacity in lane  $l$  through the intersection and there is no queue in the lane in time  $t_k$ :

$$\text{If } f_{j-1,i,l}(t_k) > c_{j-1,i,l} \text{ and } q_{j-1,i,l}(t_k) = 0, \text{ then } f'_{j-1,i,l}(t_k) = c_{j-1,i,l} \text{ and} \quad (3-10)$$

$$q'_{j-1,i,l}(t_k) = (f_{j-1,i,l}(t_k) - c_{j-1,i,l}) \cdot \Delta t \quad (3-10^a)$$

$$\text{if } q'_{j-1,i,l}(t_k) > cp_{j-1,i,l}, \text{ then } q'_{j-1,i,l}(t_k) = cp_{j-1,i,l} \text{ and} \quad (3-10^b)$$

$$q_{j-1}(t_k) = (f_{j-1,i,l}(t_k) - c_{j-1,i,l}) \cdot \Delta t - cp_{j-1,i,l}$$

- Flow on the link in lane  $l$  is higher than the flow capacity in lane  $l$  through the intersection and there is a queue in the lane in time  $t_k$ :

$$\text{If } f_{j-1,i,l}(t_k) > c_{j-1,i,l} \text{ and } q_{j-1,i,l}(t_k) > 0, \text{ then } f'_{j-1,i,l}(t_k) = c_{j-1,i,l} \text{ and} \quad (3-11)$$

$$q'_{j-1,i,l}(t_k) = q_{j-1,i,l}(t_k) + (f_{j-1,i,l}(t_k) - c_{j-1,i,l}) \cdot \Delta t \quad (3-11^a)$$

$$\text{if } q'_{j-1,i,l}(t_k) > cp_{j-1,i,l}, \text{ then } q'_{j-1,i,l}(t_k) = cp_{j-1,i,l} \text{ and} \quad (3-11^b)$$

$$q_{j-1}(t_k) = q_{j-1,i,l}(t_k) + (f_{j-1,i,l}(t_k) - c_{j-1,i,l}) \cdot \Delta t - cp_{j-1,i,l}$$

- Flow on the link in lane  $l$  is lower than the flow capacity in lane  $l$  through the intersection and there is a queue in the lane in time  $t_k$ :

$$\text{If } f_{j-1,i,l}(t_k) \leq c_{j-1,i,l} \text{ and } q_{j-1,i,l}(t_k) > 0, \text{ then } f'_{j-1,i,l}(t_k) = c_{j-1,i,l} \text{ and} \quad (3-12)$$

$$q'_{j-1,i,l}(t_k) = q_{j-1,i,l}(t_k) + (f_{j-1,i,l}(t_k) - c_{j-1,i,l}) \cdot \Delta t \quad (3-12^a)$$

$$\text{if } q'_{j-1,i,l}(t_k) < 0, \text{ then } f'_{j-1,i,l}(t_k) = c_{j-1,i,l} + \frac{q'_{j-1,i,l}(t_k)}{\Delta t} \text{ and } q'_{j-1,i,l}(t_k) = 0 \quad (3-12^b)$$

Flow from the intersection to some direction is calculated as a sum of all the flows leaving the link on the appropriate lane. For example, flow from  $i$  to direction  $j+1$  is calculated as a sum of the proportion of the flow from  $i-1$  turning left in lane 1, the proportion of the flow from  $j-1$  going straight in lane 1 and the proportion of the flow from  $i+1$  turning right in lane 2. Also, the flow from the number of cars in the queue in the previous time steps travelling that direction is also added to the expression:

$$\begin{aligned} f_{i,j+1}(t_k) = & \frac{d_{i-1,i,j+1}(t_k)}{d_{i-1,i,j+1}(t_k) + d_{i-1,i,i+1}(t_k)} \cdot f_{i-1,i,1}(t_k) + \frac{d_{j-1,i,j+1}(t_k)}{d_{j-1,i,j+1}(t_k) + d_{j-1,i,i-1}(t_k)} \cdot f_{j-1,i,1}(t_k) \\ & + \frac{d_{i+1,i,j+1}(t_k)}{d_{i+1,i,j+1}(t_k)} \cdot f_{i+1,i,2}(t_k) + f(q_{i-1,i,1}(t_{k-1})) + f(q_{j-1,i,1}(t_{k-1})) + f(q_{i+1,i,2}(t_{k-1})) \end{aligned} \quad (3-13)$$

For example, variable  $f(q_{i-1,i,1}(t_{k-1}))$  describes the flow from the queue in the previous time step between the junctions  $i-1$  and  $i$  that is going to the direction of  $j+1$ .

Once the queue exceeds the link capacitance, its effects are propagated back through the network, following the general algorithm.

### 3.2.2 T-junction Model

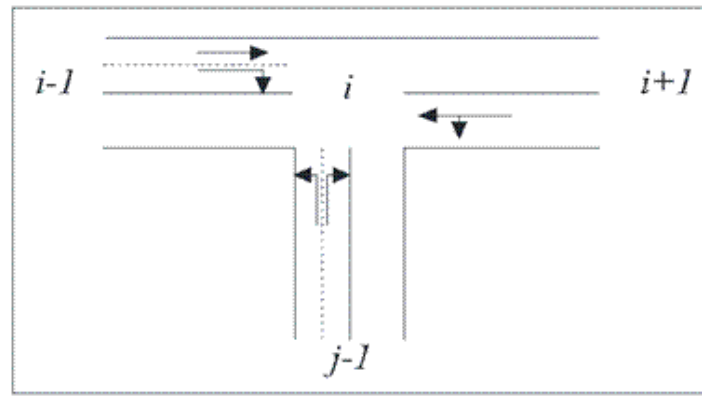


Figure 3-2: T-junction

This junction, shown in Figure 3-2, is controlled assuming that drivers obey the right-of-way rules. On the T-junction a vehicle travelling on the major roads has right-of-way and a vehicle approaching the major road must allow it to pass before joining the flow of traffic. Some roads to the intersection have a single lane, some have two lanes. Two lanes are used on the left part of major road, where lane 1 is used for going straight on and lane 2 is used for turning right and crossing the oncoming traffic on the major road. Also, two lanes are used on the minor road, lane 1 is used for turning left and lane 2 is used for turning right. The rest of the roads have a single lane.

The junction specific data applied in this model is:

$f_{i-1,i,l}(t_k)$  - flow at node  $i$  that coming from direction  $i-1$  and going in lane  $l$ , i.e. 1 represents the left lane, in time  $t_k$  (pcu/hr)

$f_{i+1,i}(t_k)$  - flow at node  $i$  that coming from direction  $i+1$  in time  $t_k$  (pcu/hr)

$c_{i-1,i,l}$  - node  $i$  flow capacity for the traffic coming from direction  $i-1$  and going in lane  $l$ , depending on gaps between vehicles and vehicle speed (pcu/hr)

- $cp_{i-1,i,l}$  - link capacitance in lane  $l$ , i.e. the maximum number of cars which can queue in lane  $l$  of the link (pcu)
- $q_{i-1,i,l}(t_k)$  - average number of vehicles queuing on lane  $l$  of arm  $i-1$  for node  $i$  at the beginning of  $t_k$  (pcu)
- $q_{i-1}(t_k)$  - average number of vehicles propagating back to the upstream links of node  $i$  from arm  $i-1$  in time  $t_k$  (pcu)

The T-junction is controlled by right-of-way rules and a priority is set for certain directions. Therefore, in order to calculate the queue for each direction on the junction, i.e. the major roads  $i-1$  and  $i+1$  and the minor road  $j-1$ , has to be considered separately.

### 3.2.2.1 The Major Road $i-1$

For the flow from direction  $i-1$  to direction  $i+1$ , i.e. in lane 1, no conflicting traffic restriction on the flow exists. Therefore, a queue can only build up due to the flow capacity on the link after the junction, following the general rule described in Section 3.1.3.

For the flow from direction  $i-1$  to direction  $j-1$ , i.e. in lane 2, the conflicting flow is the flow from direction  $i+1$  to  $i$ . A queue builds up if the flow in lane 2 or the conflicting flow is higher than the flow capacity in lane 2 through the intersection. The updated flow is expressed as  $f'_{i-1,i,l}(t_k)$  and the updated queue is expressed as  $q'_{i-1,i,l}(t_k)$ .

Five cases are considered:

- Flow on the link in lane 2 is higher than the flow capacity in lane 2 through the intersection, the conflicting flow from direction  $i+1$  is lower than the flow capacity in lane 2 through the intersection, and there is no queue in the lane in time  $t_k$ :

If  $f_{i-1,i,2}(t_k) > c_{i-1,i,2}$ , and  $f_{i+1,i}(t_k) \leq c_{i-1,i,2}$  and  $q_{i-1,i,2}(t_k) = 0$ , then (3-14)

$$f'_{i-1,i,2}(t_k) = c_{i-1,i,2} \text{ and}$$

$$q'_{i-1,i,2}(t_k) = (f_{i-1,i,2}(t_k) - c_{i-1,i,2}) \cdot \Delta t \quad (3-14^a)$$

if  $q'_{i-1,i,2}(t_k) > cp_{i-1,i,2}$ , then  $q'_{i-1,i,2}(t_k) = cp_{i-1,i,2}$  and (3-14<sup>b</sup>)

$$q_{i-1}(t_k) = (f_{i-1,i,2}(t_k) - c_{i-1,i,2}) \cdot \Delta t - cp_{i-1,i,2}$$

- The conflicting flow from direction  $i+1$  is higher than the flow capacity in lane 2 through the intersection and there is no queue in the lane in time  $t_k$  :

If  $f_{i+1,i}(t_k) > c_{i-1,i,2}$  and  $q_{i-1,i,2}(t_k) = 0$ , then  $f'_{i-1,i,2}(t_k) = 0$  and

$$(3-15)$$

$$q'_{i-1,i,2}(t_k) = f_{i-1,i,2}(t_k) \cdot \Delta t \quad (3-15^a)$$

if  $q'_{i-1,i,2}(t_k) > cp_{i-1,i,2}$ , then  $q'_{i-1,i,2}(t_k) = cp_{i-1,i,2}$ , and (3-15<sup>b</sup>)

$$q_{i-1}(t_k) = f_{i-1,i,2}(t_k) \cdot \Delta t - cp_{i-1,i,2}$$

- Flow on the link in lane 2 is higher than the flow capacity in lane 2 through the intersection, the conflicting flow from direction  $i+1$  is lower than the flow capacity in lane 2 through the intersection and there is a queue in the lane in time  $t_k$  :

If  $f_{i-1,i,2}(t_k) > c_{i-1,i,2}$ , and  $f_{i+1,i}(t_k) \leq c_{i-1,i,2}$  and  $q_{i-1,i,2}(t_k) > 0$ , then (3-16)

$$f'_{i-1,i,2}(t_k) = c_{i-1,i,2} \text{ and}$$

$$q'_{i-1,i,2}(t_k) = q_{i-1,i,2}(t_k) + (f_{i-1,i,2}(t_k) - c_{i-1,i,2}) \cdot \Delta t \quad (3-16^a)$$

if  $q'_{i-1,i,2}(t_k) > cp_{i-1,i,2}$ , then  $q'_{i-1,i,2}(t_k) = cp_{i-1,i,2}$ , and (3-16<sup>b</sup>)

$$q_{i-1}(t_k) = q_{i-1,i,2}(t_k) + (f_{i-1,i,2}(t_k) - c_{i-1,i,2}) \cdot \Delta t - cp_{i-1,i,2}$$

- The conflicting flow from direction  $i+1$  is higher than the flow capacity in lane 2 through the intersection and there is a queue in the lane in time  $t_k$ :

$$\text{If } f_{i+1,i}(t_k) > c_{i-1,i,2} \text{ and } q_{i-1,i,2}(t_k) > 0, \text{ then } f'_{i-1,i,2}(t_k) = 0, \text{ and} \quad (3-17)$$

$$q'_{i-1,i,2}(t_k) = q_{i-1,i,2}(t_k) + f_{i-1,i,2}(t_k) \cdot \Delta t \quad (3-17^a)$$

$$\text{if } q'_{i-1,i,2}(t_k) > cp_{i-1,i,2}, \text{ then } q'_{i-1,i,2}(t_k) = cp_{i-1,i,2}, \text{ and} \quad (3-17^b)$$

$$q_{i-1}(t_k) = q_{i-1,i,2}(t_k) + f_{i-1,i,2}(t_k) \cdot \Delta t - cp_{i-1,i,2}$$

- Flow on the link in lane 2 is lower than the flow capacity in lane 2 through the intersection, the conflicting flow from direction  $i+1$  is lower than the flow capacity in lane 2 through the intersection and there is a queue in the lane in time  $t_k$ :

$$\text{If } f_{i-1,i,2}(t_k) \leq c_{i-1,i,2}, \text{ and } f_{i+1,i}(t_k) \leq c_{i-1,i,2} \text{ and } q_{j-1,i,l}(t_k) > 0, \text{ then} \quad (3-18)$$

$$f'_{i-1,i,2}(t_k) = c_{i-1,i,2} \text{ and}$$

$$q'_{i-1,i,2}(t_k) = q_{i-1,i,2}(t_k) + (f_{i-1,i,2}(t_k) - c_{i-1,i,2}) \cdot \Delta t \quad (3-18^a)$$

$$\text{if } q'_{i-1,i,2}(t_k) < 0, \text{ then } f'_{i-1,i,2}(t_k) = c_{i-1,i,2} + \frac{q'_{i-1,i,2}(t_k)}{\Delta t} \text{ and } q'_{i-1,i,2}(t_k) = 0 \quad (3-18^b)$$

### 3.2.2.2 The Major Road $i+1$

For the flow from direction  $i+1$  to direction  $i-1$  or direction  $j-1$  no conflicting traffic requirement is present. Therefore, a queue can only build up due to the flow capacity on the link after the junction following a general rule described in Section 3.1.3.



### 3.2.2.3 The Minor Road $j-1$

For the flow from direction  $j-1$  to direction  $i-1$ , i.e. in lane 1, the conflicting flow is the flow from direction  $i+1$  to  $i-1$ . While for the flow from direction  $j-1$  to direction  $i+1$ , i.e. in lane 2, the conflicting flow is the sum of the flow from direction  $i+1$  to  $i-1$  and the flow from direction  $i-1$  to  $i+1$ . Both flows on minor road  $j-1$  are evaluated according to the same methodology as the flow in lane 2 from direction  $i-1$ .

### 3.2.3 Signalized T-junction Model

The difference between normal T-junction and signalized T-junction is that signal control is applied to the signalised T-junction so as to leave more entry space for the traffic flow from the minor road. The advantage of this junction is that the flow capacity for the traffic flow from the minor road is specified no matter what value of the traffic flow is on the major road. Typically, there are two phases for the traffic flows through the T-junction, the green times for the traffic flows on the major road and for the traffic flow on the minor road. The green times for each entry time can be adjusted according to the value of traffic flows from each arm, which enables the vehicles to pass through the junction with fewer disturbances.

In terms of Figure 3-2, the entry flow capacity for each lane is restricted both by the green time and original flow capacity, for example, the flow capacity for the 1<sup>st</sup> lane of the major road  $i-1$  is determined by:

$$c'_{i-1,i,1} = g_{i-1,i} \times c_{i-1,i,1} \quad (3-19)$$

In addition to flow capacity, conflicting theory is also adopted to restrict entry capacity. During the green time for major road, there is no conflicting flow for the traffic flows on major road  $i+1$  and on the 1<sup>st</sup> lane of major road  $i-1$ ; whereas, the traffic flow on the

2<sup>nd</sup> lane of major road  $i-1$  is dominated by the conflicting flow on major road  $i+1$ , the underlying methodology is explained in Section 3.2.2.1. On the other hand, during the green time for the minor road, no conflicting flow is presented for both left-turning and right-turning flows.

### 3.2.4 Roundabout Model

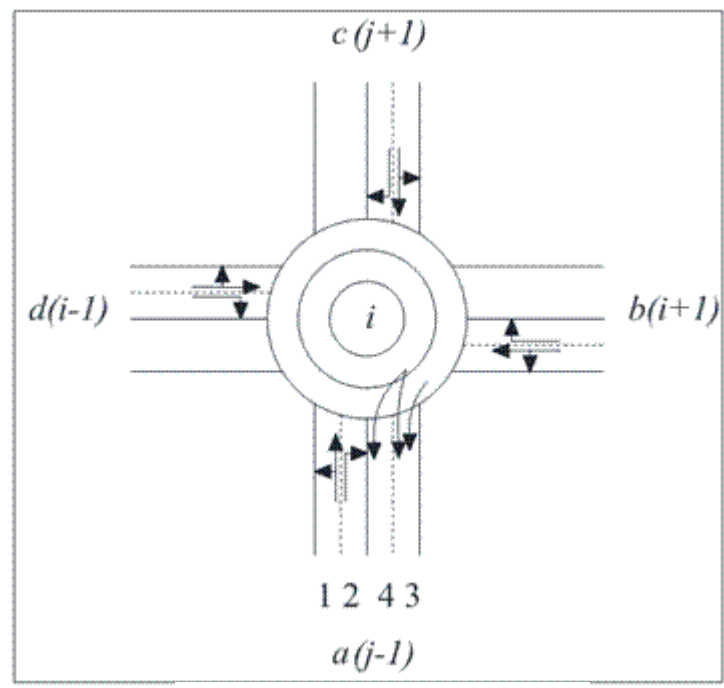


Figure 3-3: Roundabout

A roundabout is modeled in effect as a series of priority junctions with priority to traffic on the roundabout, such as urban roundabout and motorway roundabout. However, signalized roundabout is operated by signal control without consideration of conflicting flows.

Assume a roundabout with four entry arms  $a$ ,  $b$ ,  $c$  and  $d$  for roundabout  $i$ , illustrated in Figure 3-3, used to represent to upstream links  $j-1-i$ ,  $i+1-i$ ,  $j+1-i$ ,  $i-1-i$  respectively. The order in which the traffic flow is calculated at the roundabout progresses clockwise round the roundabout starting at arm  $a$ :  $a-d-c-b$ . The flow from arm  $a$  will be dependent on the flow from arms  $b$  and  $c$ , this is because only parts of flows from arms  $b$  and  $c$

have to pass through the entrance of arm  $a$ , i.e. lane 1 and lane 2, to leave the roundabout, while the flow from arm  $d$  exits at the back of the entrance, i.e. lane 3 and lane 4 of arm  $a$ . The same for other arms, the inflow depends on preceding circulating inflows.

### 3.2.4.1 Urban Roundabout model

Each road to the roundabout is considered to have two lanes; lane 1 is used for turning left and going straight, and lane 2 for turning right. The urban roundabout model assumes that vehicles are not allowed to queue on the roundabout. It also assumes that the vehicles intending to make a u-turn and return in opposite direction along the road in which they approach the roundabout are sufficiently small that they can be ignored.

Because roundabout is usually symmetrically constructed, only the traffic flow on arm  $a$  is analysed. For the flow from arm  $a$  to arm  $d$  and  $c$ , i.e. in lane 1, the flow includes the merged flow from arm  $b$  that is going straight on and the flow from arm  $c$  that is turning right:

$$cf_{a,i,1}(t_k) = \min(f_{b,i,d}, c_{b,i,1} \times \frac{d_{b,i,d}}{d_{b,i,a} + d_{b,i,d}}) + \min(f_{c,i,d}, c_{c,i,2}) \quad (3-20)$$

where

- $cf_{n,i,l}(t_k)$  - conflicting flow for the flow on entry arm  $n$ , i.e.  $a$ ,  $b$ ,  $c$  and  $d$ , in lane  $l$ ,  
i.e. 1 and 2 (pcu/hr)
- $f_{n,i,m}$  - traffic flow from arm  $n$  to arm  $m$  (pcu/hr)
- $c_{n,i,l}$  - flow capacity for lane  $l$  of arm  $n$  (pcu/hr)
- $d_{n,i,m}$  - proportion of the traffic from arm  $n$  to arm  $m$

The *min* formula is used to restrict the inflow from each arm. The first *min* formula in Equation 3-20, takes the minimum of: the first parameter which represents the traffic

flow intending to move from arm  $b$  to arm  $d$ , and the second parameter which corresponds to the flow capacity for this traffic flow in lane 1 of arm  $b$ . While for the second *min* formula in Equation 3-20, it takes the minimum of: the first parameter that demonstrates the traffic flow intending to move from arm  $c$  to arm  $d$ , and the second parameter which indicates the flow capacity for this flow in lane 2 of arm  $c$ . A queue builds up if the flow in lane 1 or the conflicting flow is higher than the flow capacity in lane 1 through the intersection. For the flow from arm  $a$  to arm  $b$ , the conflicting flow is the sum of the flows from arm  $b$  to arm  $d$  and  $c$  and the flow from arm  $c$  to arm  $d$ . A queue builds up if the flow in lane 2 or the conflicting flow is higher than the flow capacity in lane 2 through the intersection. The conflicting flow for lane 2 of arm  $a$  is described as:

$$cf_{a,i,2}(t_k) = \min(f_{b,i,d}, c_{b,i,1} \times \frac{d_{b,i,d}}{d_{b,i,a} + d_{b,i,d}}) + \min(f_{b,i,c}, c_{b,i,2}) + \min(f_{c,i,d}, c_{c,i,2}) \quad (3-21)$$

The first term and the third term in Equation 3-21 are the same as the two terms in Equation 3-20, as the conflicting flow for lane 1 of arm  $a$  also restricts the flow on lane 2 of arm  $a$ . Moreover it is assumed that the traffic on lane 2 of arm  $a$  only takes the inner lane circulating the roundabout, therefore the traffic flow on lane 2 of arm  $a$  is further limited by the traffic moving from arm  $b$  to arm  $c$  that described as the second term in Equation 3-21.

The evaluation of the two traffic flow on arm  $a$  follows the same rule in Section 3.2.2.1. Due to the symmetrical structure of roundabout, the inflow from arm  $a$  depends on the inflows from arm  $b$  and  $c$ , further the flow from arm  $c$  is also influenced by the flow from  $d$  and  $a$ . Rather than modelling this traffic interaction in detail, the evaluation of the traffic on each arm is conducted iteratively, the inflow for each arm on current iteration is derived in terms of the conflicting traffic flow obtained on the current or

previous iteration. Convergence is considered to be reached when the traffic conditions for each arm are unchanged by further iterations.

#### 3.2.4.2 Motorway Roundabout Model

The same methodology is applied to motorway roundabout as the urban roundabout. The extension for this roundabout is that vehicles are allowed to queue on the roundabout. In addition to the restriction of inflows, outflows are also restricted according to the flow capacity of the out-going lanes for each arm. The results from urban roundabout are employed as inputs for the calculation of outflows of the motorway roundabout. It assumes that cars in the queue on the roundabout will leave gaps for entering and exiting the roundabout. Another assumption is that the out-going flow on the outer lane will only take the left lane as exit, while the traffic flow on the inner lane can take either as exit, e.g. considering Figure 3-3, lane 3 for arm *a* is the exit for the traffic flow on the outer lane, and lane 4 is the exit for the traffic flow on the inner lane. If there is residual exiting flow capacity on lane 3, the traffic on the inner lane can also leave the roundabout through lane 3. As the roundabout adopted in this model is assumed to be built symmetrically, the free spaces for each part of the roundabout are the same.

The general rule is that if queue builds up at an exit, it will propagate back through the outer lane or the inner lane, and the length of the queue at each part of the roundabout is derived according to the ratio of mixed inflows through that part of the roundabout. E.g. for part *ab* on outer lane, when the exiting traffic flow, i.e. traffic flow from arm *b* and *c* that intending to exit through arm *a*, exceeds the flow capacity of arm *a*, portion of this traffic would be disturbed on part *ab*, which further resulted in the blockage of traffic from arm *b* that exiting at arm *d*, the amount of this disturbed traffic is obtained in line with the flow ratios on part *ab* of outer lane. On the other hand, if the exiting traffic

flow is less than the flow capacity of the exit arm, the queue formed at previous steps will dissipate that also based on the ratio of mixed flows on each part of the roundabout.

The additional junction data is employed in motorway roundabout:

$cp_{i,ot}$  - capacitance for outer lane of roundabout  $i$  (pcu)

$cp_{i,in}$  - capacitance for inner lane of roundabout  $i$  (pcu)

$c_{a,i,l}$  - flow capacity for lane  $l$  of arm  $a$  (pcu/hr)

$f_{a,i,l}$  - outflow for lane  $l$  of arm  $a$ ,  $l$  includes 3 and 4 (pcu/hr)

$f_{b,i,a}$  - inflow from arm  $b$  to arm  $a$  (pcu/hr), which is calculated as: `

$$f_{b,i,a} = f_{b,i,1} \times \frac{d_{b,i,a}}{d_{b,i,a} + d_{b,i,d}}$$

$of_{b,i,a}$  - outflow of roundabout  $i$  from arm  $b$  to arm  $a$  (pcu/hr)

$d_{b,i,a}$  - proportion of the traffic flow from arm  $b$  to arm  $a$

$oqp_{a,b}$  - length of the queue formed in section  $ab$  of the outer lane of roundabout (pcu), as illustrated in Figure 3-4

$iq_{a,b}$  - length of the queue formed in section  $ab$  of the inner lane of roundabout (pcu), as illustrated in Figure 3-4

$oqp_{a,b}$  - queue propagation of queue for section  $ab$  of outer lane to other links (pcu)

$iqp_{a,b}$  - queue propagation of queue for section  $ab$  of inner lane to other links (pcu)

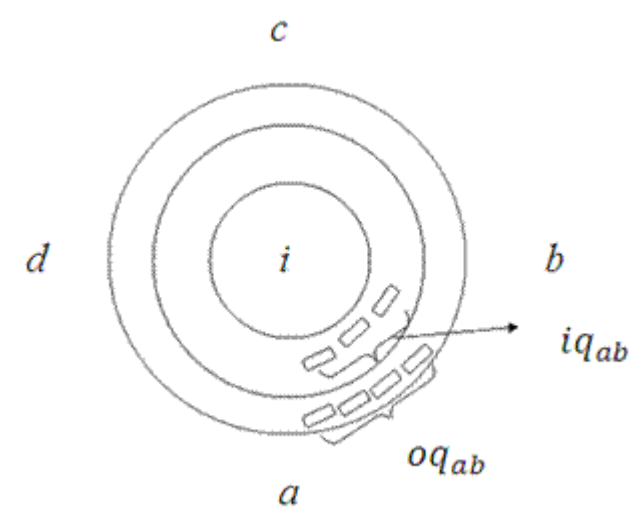


Figure 3-4: The formation of queue at the roundabout

The traffic flows from each arm will occupy their defined path through the roundabout; the mixed traffic flow at each component of the roundabout is represented in Table 3-2.

Table 3-2: Composition of traffic flows

Components of roundabout	Outer lane	Inner lane
ab	$f_{b,i,a}, f_{b,i,d}, f_{c,i,a}$	$f_{d,i,a}, f_{c,i,d}, f_{b,i,c}$
bc	$f_{c,i,b}, f_{c,i,a}, f_{d,i,b}$	$f_{a,i,b}, f_{d,i,a}, f_{c,i,d}$
cd	$f_{d,i,c}, f_{d,i,b}, f_{a,i,c}$	$f_{b,i,c}, f_{a,i,b}, f_{d,i,a}$
da	$f_{a,i,d}, f_{a,i,c}, f_{b,i,d}$	$f_{c,i,d}, f_{b,i,c}, f_{a,i,b}$

The urban roundabout model is initially deployed as the first part of the motorway roundabout model. The results of urban roundabout model are then employed as inputs for the next calculation in the motorway roundabout model. As the roundabout is assumed to be constructed symmetrically, arm *a* is analysed to represent the movement at the roundabout. Lane 3 and lane 4 for arm *a* (as in Figure 3-3) are analysed separately. The evaluation of the exiting traffic in lane 3 for arm *a* is described as follows:

- A queue builds up if flow on the link in lane 3 is higher than the flow capacity of lane 3 of arm *a*,

If  $f_{b,i,a} + f_{c,i,a} > c_{a,i,3}$ ,  $f_{a,i,3} = c_{a,i,3}$

For section  $ab$  of outer lane at the roundabout, the disturbed traffic for exiting traffic at arm  $a$ , i.e.  $f_{b,i,a}$  and  $f_{c,i,a}$  in time  $t_k$  is calculated as

$(f_{b,i,a} + f_{c,i,a} - c_{a,i,3}) \times \Delta t$ , further led to the blockage of flow  $f_{b,i,d}$  derived as

$(f_{b,i,a} + f_{c,i,a} - c_{a,i,3}) \times \Delta t \times \frac{f_{b,i,d}}{f_{b,i,a} + f_{c,i,a}}$  based on the flow proportions on section

$ab$  of outer lane. This is because these three directional traffic flows are mixed together on section  $ab$  of outer lane. As a result, the queue formed on section  $ab$  of outer lane is updated as:

$$oq'_{a,b} = oq_{a,b} + (f_{b,i,a} + f_{c,i,a} - c_{a,i,3}) \times \Delta t \times \frac{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}}{f_{b,i,a} + f_{c,i,a}} \quad (3-22)$$

In addition, the variation of queue in  $ab$  is represented as:

$$\Delta oq_{a,b} = \min(oq'_{a,b}, \frac{cp_{i,ot}}{4}) - oq_{a,b} \quad (3-23)$$

Equation 3-23 is used to restrict the queue increment in section  $ab$ . The first term represents the updated queue length in  $ab$ , when it is greater than the capacitance of section  $ab$ , it is restricted as  $\frac{cp_{i,ot}}{4}$ .

Due to the traffic disturbance, the traffic flows that passing through section  $ab$  of outer lane are decreased by:

$$\Delta f_{b,i,a} = -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{b,i,a}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}}$$

$$\Delta f_{c,i,a} = -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{c,i,a}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}}$$



$$\Delta f_{b,i,d} = -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{b,i,d}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \quad (3-24)$$

If queue is less than the queue capacitance,  $oq'_{a,b} \leq \frac{cp_{i,ot}}{4}$ , the queue is restricted in  $ab$ .

If  $oq'_{a,b} > \frac{cp_{i,ot}}{4}$ , part of the queue propagates back to upstream links, described as:

$$oqp_{a,b} = oq'_{a,b} - \frac{cp_{i,ot}}{4} \quad (3-25)$$

and  $oq'_{a,b}$  is updated as  $\frac{cp_{i,ot}}{4}$ .

As this queue propagation is contributed by traffic flows from arms  $b$  and  $c$ , it spills back to lane 1 of arm  $b$  and section  $bc$  of outer lane. The first part that induced by  $f_{b,i,a}$  and  $f_{b,i,d}$  is added to the queue in lane 1 of arm  $b$ , described as:

$$q'_{b,1} = q_{b,1} + oqp_{a,b} \times \frac{f_{b,i,a} + f_{b,i,d}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \quad (3-26)$$

and the inflow from lane 1 arm  $b$  is decreased as:

$$f'_{b,i,1} = f_{b,i,1} - \frac{oqp_{a,b}}{\Delta t} \times \frac{f_{b,i,a} + f_{b,i,d}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \quad (3-27)$$

If  $q'_{b,1} > cp_{b,i,1}$ , queue propagates back to the source links of arm  $b$ ,

$$qp_{b,1} = q'_{b,1} - cp_{b,i,1} \text{ and } q'_{b,1} = cp_{b,i,1} \quad (3-28)$$

The rest part spills back to section  $bc$  of outer lane, and the queue in section  $bc$  is updated as:

$$oq'_{b,c} = oq_{b,c} + oqp_{a,b} \times \frac{f_{c,i,a}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \times \frac{f_{c,i,a} + f_{c,i,b} + f_{d,i,b}}{f_{c,i,a}} \quad (3-29)$$

The disturbed traffic of  $f_{c,i,a}$  further resulted in the blockage of  $f_{c,i,b}$  and  $f_{d,i,b}$ , as they are mixed in section  $bc$  of outer lane.

Consequently, the amount of traffic that could exit the roundabout at arm  $a$  from arm  $b$  is computed as:

$$of_{b,i,a} = f'_{b,i,1} \times \frac{d_{b,i,a}}{d_{b,i,a} + d_{b,i,d}} + \Delta f_{b,i,a} \quad (3-30)$$

Firstly, the entering traffic at arm  $a$  is reduced by queue propagation in lane 1 of arm  $b$  according to Equation 3-27; and then it is further decreased due to traffic disturbance occurred in part  $ab$  of outer lane based on Equation 3-24.

Subsequently, the same methodology is applied to section  $bc$ ,  $cd$  and  $da$ .

- Flow on the link in lane 3 is lower than the flow capacity of lane 3 through the roundabout

$$\text{If } f_{b,i,a} + f_{c,i,a} \leq c_{a,i,3},$$

- a. If  $q_{a3} = 0$ , no queue.

$$f_{a,i,3} = f_{b,i,a} + f_{c,i,a} \quad (3-31)$$

- b. If  $0 < q_{a3} \leq \frac{cp_{i,ot}}{4}$ , queue located in part  $ab$ .

$$f_{a,i,3} = c_{a,i,3} \quad (3-32)$$

$$oq_{d,a} = oq_{c,d} = oq_{b,c} = 0, oq_{a,b} = q_{a,3} \quad (3-33)$$

$$oq'_{a,b} = oq_{a,b} + (f_{b,i,a} + f_{c,i,a} - c_{a,i,3}) \times \Delta t \times \frac{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}}{f_{b,i,a} + f_{c,i,a}} \quad (3-34)$$

Equation 3-34 is evaluated as Equation 3-22, which is based on the flow ratios on section  $ab$  of outer lane.

Increment of flows due to clearance of queue is evaluated as:

$$\Delta oq_{a,b} = \max(oq'_{a,b} - oq_{a,b}, -oq_{a,b}) \quad (3-35)$$

This max formula is used to restrict the clearance of queue; the maximum value of queue reduction would be the previous queue length. As a result of queue clearance, the traffic flows that passing through part  $ab$  of outer lane are increased by:

$$\begin{aligned} \Delta f_{b,i,a} &= -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{b,i,a}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \\ \Delta f_{c,i,a} &= -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{c,i,a}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \\ \Delta f_{b,i,d} &= -\frac{\Delta oq_{a,b}}{\Delta t} \times \frac{f_{b,i,d}}{f_{b,i,a} + f_{c,i,a} + f_{b,i,d}} \end{aligned} \quad (3-36)$$

If  $oq'_{a,b} \geq 0$ , queue still exists in  $ab$ .

If  $oq'_{a,b} < 0$

$$f_{a,i,3} = c_{a,i,3} + oq'_{a,b} \times \frac{f_{b,i,a} + f_{c,i,a}}{f_{b,i,a} + f_{b,i,d} + f_{c,i,a}} \quad \text{and } oq'_{a,b} = 0 \quad (3-37)$$

Consequently, the traffic flows that passing through  $ab$  are updated as:

$$of_{b,i,a} = f_{b,i,a} + \Delta f_{b,i,a}$$

$$of_{b,i,d} = f_{b,i,d} + \Delta f_{b,i,d}$$

$$of_{c,i,b} = f_{c,i,b} + \Delta f_{c,i,b} \quad (3-38)$$

- c. If  $\frac{cp_{i,ot}}{4} < q_{a3} \leq \frac{cp_{i,ot}}{2}$ , queue located in part *ab*, *bc*. Queue clears from *bc* to *ab*.

The methodology used above will be applied, the first step is to calculate the variation of queue in part *bc*, if the queue in *bc* is cleared, the queue in *ab* need to be evaluated. Then the resulting increments due to queue reduction will be added to out-going flow.

- d. If  $\frac{cp_{i,ot}}{2} < q_{a3} \leq \frac{3}{4}cp_{i,ot}$ , queue located in part *ab*, *bc*, *cd*. Queue clears from *cd* to *ab*.

The sequence for the evaluation of queue at the roundabout would be *cd-bc-ab*.

- e. If  $\frac{3}{4}cp_{i,ot} < q_{a3} \leq cp_{i,ot}$ , queue located in part *ab*, *bc*, *cd*, *da*. Queue clears from *da* to *ab*.

The sequence for the evaluation of queue at the roundabout would be *da-cd-bc-ab*.

As for the lane 4 of arm *a*, the exiting traffic flow is computed as:

$$c'_{a,i,4} = c_{a,i,4} + \max(c_{a,i,3} - f_{a,i,3}, 0) \quad (3-39)$$

Equation 3-39 is used to determine the outflow capacity for lane 4 arm *a*. As flow  $f_{d,i,a}$  can exit the roundabout by either lane 3 or 4, the outflow capacity for  $f_{d,i,a}$  is calculated as the sum of capacity of lane 4 and residual flow capacity of lane 3. The methodology, used to evaluate the traffic condition in lane 3 for arm *a*, is also applied to the outflow through the lane 4 for arm *a*.

### 3.2.5 Signalized Roundabout Model

In order to overcome traffic congestion, signals are introduced to roundabouts. Typically, all the arms of a signalized roundabout are signal controlled; sometimes one arm is left uncontrolled to improve flow capacity. In addition, the signals at roundabouts can be operated full-time or part-time (such as rush hour in the morning or evening). Signalized roundabouts are now quite popular in many urban areas, as well as at motorway interchanges.

The conflict among the traffic flows from the same arm and from competing arms at signalized roundabouts is eliminated by traffic signals. Each traffic flow from one of the arms to another will be assigned a unique green time period. Thus, the flow capacity for traffic flow  $f_{a,i,b}$  is calculated as:

$$C_{a,i,b} = C_i \times g_{a,i,b} \quad (3-40)$$

where

$f_{a,i,b}$  - the traffic flow from arm  $a$  to arm  $b$  (pcu/hr)

$C_{a,i,b}$  - the traffic flow capacity for the traffic flow  $f_{a,i,b}$  (pcu/hr)

$C_i$  - the flow capacity for the lane that the traffic flow  $f_{a,i,b}$  is located on (pcu/hr)

$g_{a,i,b}$  - green period for traffic flow  $f_{a,i,b}$  at one time step

After the evaluation of flow capacity for each traffic flow through the roundabout, the equations in Section 3.2.1 are employed to calculate the outflow, queue and queue propagation at the roundabout.

### 3.2.6 On-ramp and Off-ramp Models

Highway ramp (including on-ramp and off-ramp) or slip road is a short road segment which allows vehicles to enter or exit a motorway.

#### 3.2.6.1 On-ramp

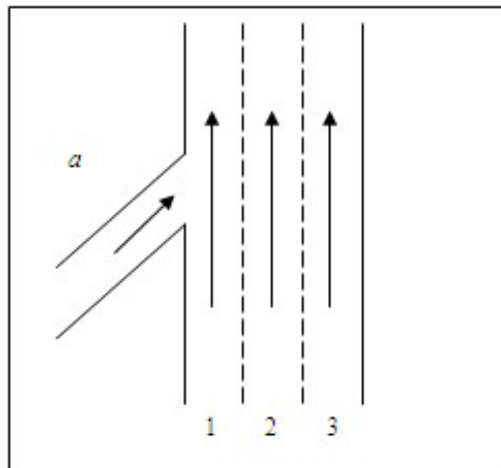


Figure 3-5: On-ramp

Traffic on trunk roads would be divided among lanes 1, 2 and 3 with respect to the observed proportions, and the merging flow from slip road  $a$  needs to find a gap on the left lane of the trunk road is carrying a part of the major traffic. The junction specific data are shown as follows:

$f_a$  - entry flow of slip road  $a$  (pcu/hr)

$F_{on}$  - traffic flow on Motorway (pcu/hr)

$d_i$  - proportions of traffic flows in lane  $i$ , i.e. 1, 2, 3

$f_i$  - traffic flow in lane  $i$  (pcu/hr)

$c_a$  - flow capacity for slip road  $a$  (pcu/hr)

$cp_a$  - link capacitance for slip road  $a$  (pcu)

As there are more than two lanes for the trunk road, only the traffic on the left lane that adjacent to the slip road has an impact on the entry flow. The traffic flow on lane 1 is recognised as the conflicting flow for the flow from slip road  $a$ , which is computed as:

$$f_1 = F_{on} \times d_1 \quad (3-41)$$

If  $f_a$  or  $f_1$  exceeds the flow capacity of the slip road  $a$ ,  $c_a$ , a queue will build up on slip road  $a$ . If the queue exceeds the link capacitance on slip road  $a$ ,  $cp_a$ , the queue will propagate back through the network.

### 3.2.6.2 Off-ramp

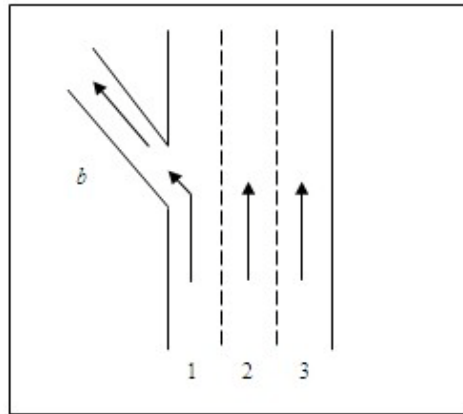


Figure 3-6: Off-ramp

The junction specific data for off-ramp include:

$F_{off}$  - traffic flow on motorway (pcu/hr)

$f_i$  - traffic flow in lane  $i$  (pcu/hr)

$f_b$  - exiting flow on off-ramp link  $b$  (pcu/hr)

$c_b$  - traffic flow capacity for link  $b$  (pcu/hr)

$cp_1$  - link capacitance on lane 1 for major road (pcu)

For the trunk road, assume that only the traffic leaving the major road would take the left lane. Then, the outflow from the trunk road is calculated as:

$$f_1 = F_{off} \times d_1 \quad (3-42)$$

$f_1$  represents the entering flow for off-ramp link  $b$ ; while  $f_b$  demonstrates the exiting flow for off-ramp link  $b$ . If  $f_1$  exceeds  $c_b$ , the queue will be present on lane 1.

### 3.2.7 Merge and Diverge Models

In addition to on-ramp and off-ramp on motorways, merge and diverge are also constructed at where two motorways merge together, as well as at the unidirectional urban roads.

#### 3.2.7.1 Merge

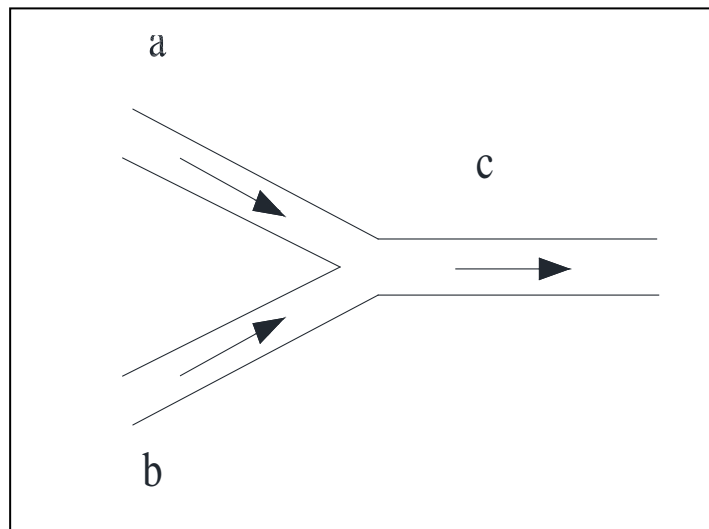


Figure 3-7: Merge

A typical “Y-merge” is illustrated in Figure 3-7, where two streams of traffic from road  $a$  and  $b$  merge into one with equal priority.

There are two main steps involved in the evaluation of a “merge” junction. Firstly, the out-going flows from links  $a$  and  $b$  would be restricted by the corresponding link flow capacity. The general rule listed in Section 3.1.3 is used to calculate the outflow, queue



and queue propagation for links *a* and *b*. Next, the in-coming flows from link *a* and *b* would be further limited by the flow capacity of link *c*. Thus, the same rule is employed to calculate the inflow for link *c* and the queue blocking back to links *a* and *b*. The propagation of queue for each source link is determined by its proportion in the mixed traffic flow.

### 3.2.7.2 Diverge

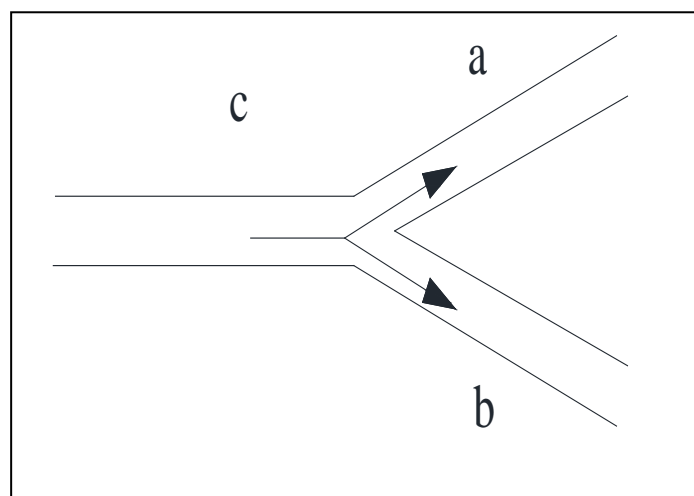


Figure 3-8: Diverge

As for the “diverge” junction, there are usually two lanes for the major road. The exiting flow for each downstream link is evaluated independently.

The left-turning flow is restricted by both the flow capacity of the left lane on link *c* and the flow capacity of link *a*, while the right-turning flow is restricted by both the flow capacity of the right lane on link *c* and the flow capacity of link *b*. The general rule is deployed to evaluate the movement of flows at the “diverge” junction.

### 3.2.8 Roadwork Node Model

The implementation of maintenance work on a road link can be represented by limiting the exiting traffic flow capacity of this link. However, the length and location of worksite, that are used to define the geometry of the worksite, also need to be considered, as does the link capacitance which changes during maintenance. Therefore in NTFM a special roadwork node is employed to represent the part of road link under maintenance, which specifies the location of worksite, the length of the worksite, and the number of lanes in service. A typical roadwork node is shown as:

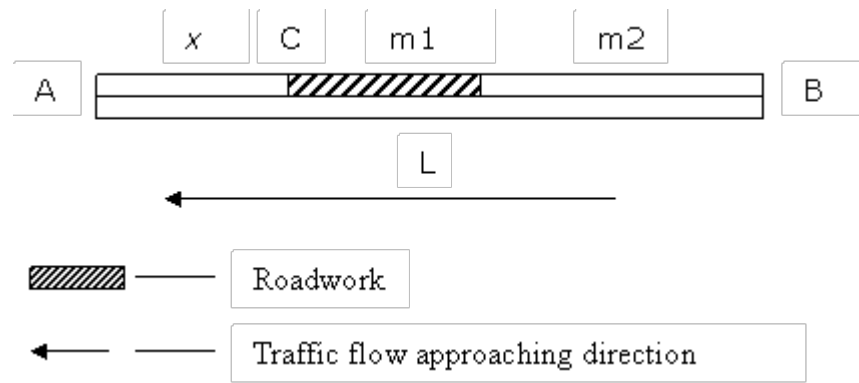


Figure 3-9: A Dual two lane carriageway road link in maintenance

where,

*A*- downstream junction

*B*- upstream junction

*C*- roadwork node

*x*- the distance between downstream junction and worksite

*m<sub>1</sub>, m<sub>2</sub>*- the length of worksite and the rest part of link without maintenance

*L*- the length of link

Two additional parameters *n1* and *n2* are used to define the number of lanes remaining open at the worksite and the rest of the link. For *BA*, its flow capacity is mainly

characterised by its weakest point *C*. The default capacity associated with the type of road is shown in Table 1-2.

### 3.2.8.1 Single Carriageway

On single carriageway roads, the two-way operation of traffic should be maintained. Therefore shuttle working is applied to assign gaps for the two competing flows. At worksites with shuttle working each direction suffers intermittent intervals of flow passing the site, alternating with intervals of zero capacity. The layout of the worksite with shuttle working is illustrated in Figure 3-10.



Figure 3-10: Single carriageway road with shuttle working

For the worksite with shuttle working, the traffic from both directions has to use the only lane open, alternately. Thus, traffic lights are applied to assign the gap for each directional flow, and the cycle time  $C_y$  is represented by:

$$C_y = G + R + A_m \quad (3-43)$$

Where

$G$ - green phase for traffic in the direction with work

$R$ - red phase for traffic in the direction with work

$A_m$ - amber phase that represents the time for the vehicles already flowing through the worksite to exit, which is proportional to the length of worksite

The site capacity for traffic flow in the direction with work is represented by the formula:

$$C_{s,1} = \frac{S \cdot G}{C_y} \quad (3-44)$$

Where  $C_{s,1}$  is the flow capacity and  $S$  is the saturation flow (pcu/hr). The saturation flow is the discharge rate through the worksite during a green phase when a queue is present on the approach to the worksite. The default value of saturation flow in QUADRO is 1800 pcu/hr [58]. Conversely, the site capacity for the other direction is calculated as:

$$C_{s,2} = \frac{S \cdot R}{C_y} \quad (3-45)$$

### 3.2.8.2 Multi-lane Carriageway

For multi-lane carriageway, i.e. dual carriageway and motorway, more than one lane can be maintained concurrently when the traffic passing by is not significant in comparison to the residual flow capacity. If  $N$  lanes are in service through the worksite each with a normal capacity  $C_n$ , then the worksite capacity would be formulated as [58]:

$$C_{M,N} = 0.85 \times N \times C_n \quad (3-46)$$

The default value of  $C_n$  is shown in Table 1-2. The evaluation of the traffic flow through the worksite then follows the general NTFM calculation process in Section 3.1.3.

## 3.3 Link Model

The link model used in NTFM is formulated based on the store-and-forward modelling philosophy. The application of store-and-forward modelling approach to traffic problems was first proposed by Gazis and Potts [59]. It is well suited for queue management on traffic lane under saturated traffic conditions [49, 60]. In terms of the store-and-forward approach, it has been assumed that vehicles experience constant travel times along a road link and are queued at the end of the link if inflow is higher

than the corresponding flow capacity. The outflow from a link is determined in terms of the turning ratios and nature of the corresponding junction. In what follows, the evaluation of travel duration spent on the road link based on store-and-forward philosophy is explicitly described.

### 3.3.1 Calculation of Travel Duration

In NTFM, the total travel time in the network is calculated as the sum of the journey duration for each vehicle spent on each link. For each link on the network, the vehicle journey time is categorised into three parts depending on if a queue is present. These parts are: the time spent on queue dissipation, the travel duration for the non-disturbed traffic that pass through the link without travel delay, and the time spent on queue formation. The disturbed traffic is defined as the traffic that is stored on the link at the current time step; whereas the travel delays at junctions is not taken into account, which is studied in the Section 3.4. One basic assumption is that traffic that is maintained on the link dissipates at a uniform speed and other traffic moves at a different uniform speed elsewhere.

### 3.3.2 Model Formulation

A typical road link, denoted by  $BA$ , is illustrated in Figure 3-11, describing the traffic condition at time  $t$ .  $f_{out}(t)$  (pcu/hr) and  $f_{in}(t)$  (pcu/hr) are used to represent the outflow and inflow on link  $BA$  at time  $t$ , and  $q(t)$  (pcu) is the queue stored on link  $BA$  at time  $t$ . It should be noted that when there is no vehicle stored on link  $BA$ , the value of  $q(t)$  will be 0. Considering the formation and dissipation of the queue on link  $BA$ , the travel duration spent on this link is calculated in three parts;  $t_l$  (hrs) is used to represent the time spent

on queue dissipation,  $t_2$  (hrs) denotes the travel duration for the traffic without travel delay, and  $t_3$  (hrs) is the time spent on queue formation.

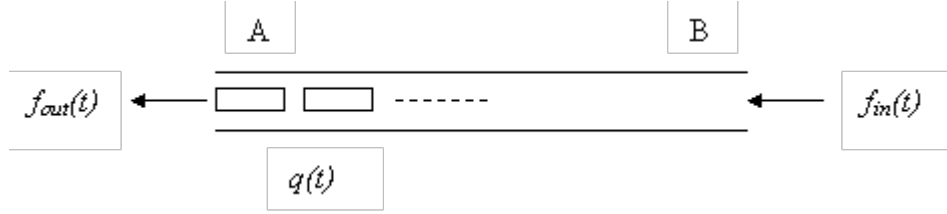


Figure 3-11: Traffic condition for link BA at time  $t$

The evolution of traffic on link  $BA$  is formulated as:

$$q(t + dt) = q(t) - (f_{out}(t) - f_{in}(t))dt \quad (3-47)$$

where  $q(t+dt)$  is the updated queue on this link at time  $t+dt$ . Considering the clearance of the initial queue on link  $BA$ , there are two main traffic conditions involved in evaluating journey time spent on this link: (1)  $q(t)$  is completely cleared at time  $t+dt$ , where  $q(t) \leq f_{out}(t)dt$ , (2) only part of  $q(t)$  is cleared at time  $t+dt$ , where  $q(t) > f_{out}(t)dt$ .

### 3.3.2.1 Case 1

In Case 1  $q(t)$  is less than or equal to the outflow on this link,  $f_{out}(t)dt$ , therefore it can be cleared at time  $t+dt$ . The time spent on queue dissipation is calculated as:

$$t_1 = \frac{\text{headway} \times q(t) \times (q(t) + 1)}{2 \times v_d} \quad (3-48)$$

where headway is defined as 0.007 km in NTFM, including vehicle length and the gap between two adjacent vehicles; and  $v_d$ (km/hr) represents the speed of queue dissipation; and  $\frac{\text{headway} \times q(t) \times (q(t) + 1)}{2}$  (km) is the travel distance for all the vehicles in the initial queue to exit link  $BA$  that follows an arithmetic progression, since the vehicles stored on the link are assumed to be lined evenly.

As the clearance of the initial queue contributed to the outflow of link *BA*, the traffic pass through the link without travel delay is obtained as  $f_{out}(t)dt - q(t)$ , the travel duration for which is computed as:

$$t_2 = \frac{(f_{out}(t)dt - q(t)) \times L}{v} \quad (3-49)$$

where  $L$  (km) is the length of link *BA* and  $v$  (km/hr) denotes the average moving speed on this link.

Moreover, in terms of Equation 3-47 if  $q(t+dt)$  is greater than 0, a new queue is built up on link *BA* at time  $t+dt$ . The time spent on the development of this queue, which is also the travel duration for the traffic delayed on link *BA* at time  $t+dt$ , is represented by the following expression:

$$t_3 = \frac{q(t+dt) \times L}{v} - \frac{headway \times q(t+dt) \times (q(t+dt) + 1)}{2 \times v} \quad (3-50)$$

The first term in Equation 3-50 gives the travel duration for the disturbed traffic that pass through link *BA* at time  $t+dt$ . Actually none of them completed its journey, the second term is provided to rectify this equation by subtracting the extra time spent on the journey that also follows an arithmetic progression as Equation 3-48. Eventually, the total travel duration spent on link *BA* is derived as the sum of  $t_1$ ,  $t_2$  and  $t_3$ .

### 3.3.2.2 Case 2

In Case 2  $q(t)$  is greater than the outflow on this link,  $f_{out}(t)dt$ , therefore only part of the initial queue is dissipated. The part dissipated is equal to the amount of outflow, and the rest part of the initial queue at time  $t+dt$  is determined as  $q(t) - f_{out}(t)dt$ . The journey time for these two parts of the initial queue is evaluated as:

$$t_1 = \frac{\text{headway} \times f_{out}(t)dt \times (f_{out}(t)dt + 1)}{2 \times v_d} + \frac{\text{headway} \times f_{out}(t)dt \times (q(t) - f_{out}(t)dt)}{v_d} \quad (3-51)$$

The first term in Equation 3-51 indicates the time spent on queue dissipation that is evaluated in a similar pattern to Equation 3-48. Owing to the clearance of  $f_{out}(t)dt$  vehicles, the remaining part of the initial queue moves forward to the end of link  $BA$  by  $\text{headway} \times f_{out}(t)dt$  (km). The second term corresponds to the travel time for the rest part of the initial queue.

As there is no non-disturbed traffic in case 2,  $t_2$  is determined as 0. Furthermore the inflow on link  $BA$  at time  $t+dt$ ,  $f_{in}(t)dt$ , is disturbed and added to the tail of the initial queue, the travel duration for which is described as:

$$t_3 = \frac{f_{in}(t)dt \times (L - \text{headway} \times (q(t) - f_{out}(t)dt))}{v} - \frac{\text{headway} \times f_{in}(t)dt \times (f_{in}(t)dt + 1)}{2 \times v} \quad (3-52)$$

where  $L - \text{headway} \times (q(t) - f_{out}(t)dt)$  (km) is the length of link  $BA$  in addition to the portion that occupied by the rest part of initial queue, and  $\frac{\text{headway} \times f_{in}(t)dt \times (f_{in}(t)dt + 1)}{2}$  (km) is the assumed extra journey for the currently disturbed traffic at time  $t+dt$ . As a consequence, the travel duration spent on a link that has limited transportability is achieved.

### 3.3.3 Extension for Multi-lane Link

For a multi-lane link, the queue formed on it is assumed to be uniformly distributed in each service lane. The calculation of the travel duration for the multi-lane link follows the same rules as the single-lane link. The lengths of queue at time  $t$  and  $t+dt$  for each lane are calculated as  $q(t)/n$  and  $q(t+dt)/n$ , respectively, where  $n$  is the number of service lanes; and the remainder modulus for them are represented by  $\text{mod}(q(t),n)$  and



$\text{mod}(q(t+dt),n)$  that can be ignored because of their small magnitude. For instance, when 7 vehicles are delayed on a two lane carriageway, the remainder modulus would be 1. The two cases considered in Section 3.3.2 are investigated.

### 3.3.3.1 Case 1

As stated above in this case the initial queue on link  $BA$ ,  $q(t)$ , is cleared, the time spent on queue dissipation is formulated as:

$$t_1 = \frac{\text{headway} \times \frac{q(t)}{n} \times (\frac{q(t)}{n} + 1)}{2 \times v_d} \times n + \frac{\text{headway} \times \text{mod}(q(t),n) \times (\frac{q(t)}{n} + 1)}{v_d} \quad (3-53)$$

where  $\frac{\text{headway} \times \frac{q(t)}{n} \times (\frac{q(t)}{n} + 1)}{2} \times n + \text{headway} \times \text{mod}(q(t),n) \times (\frac{q(t)}{n} + 1)$  (km) is the travel distance for all the vehicles in the initial queue. Due to the small amount of the remainder modulus, the second term could be neglected, but has been included here.

The travel duration for the non-disturbed traffic is evaluated in the same way as Equation 3-49. Based on Equation 3-47, the queue formed at time  $t+dt$  is obtained. The algorithm used in Equation 3-50 is again applied to compute the time spent on queue formation for multi-lane link, described as:

$$t_3 = \frac{q(t+dt) \times L}{v} - \frac{\text{headway} \times \frac{q(t+dt)}{n} \times (\frac{q(t+dt)}{n} + 1)}{2 \times v} \times n - \frac{\text{headway} \times \text{mod}(q(t+dt),n) \times (\frac{q(t+dt)}{n} + 1)}{v} \quad (3-54)$$

where

$$\frac{\text{headway} \times \frac{q(t+dt)}{n} \times (\frac{q(t+dt)}{n} + 1)}{2} \times n + \text{headway} \times \text{mod}(q(t+dt),n) \times (\frac{q(t+dt)}{n} + 1)$$

(km) is the assumed extra journey for all the disturbed traffic at time  $t+dt$ .

### 3.3.3.2 Case 2

In Case 2 only a portion of the initial queue is dissipated and the remaining part is calculated as  $q(t)-f_{out}(t)dt$ . The time spent on queue dissipation and movement of the rest part of initial queue is:

$$\begin{aligned}
 t_1 = & \frac{headway \times \frac{f_{out}(t)dt}{n} \times (\frac{f_{out}(t)dt}{n} + 1)}{2 \times v_d} \times n \\
 & + \frac{headway \times \text{mod}(f_{out}(t)dt, n) \times (\frac{f_{out}(t)dt}{n} + 1)}{v_d} \\
 & + \frac{headway \times \frac{f_{out}(t)dt}{n} \times (q(t) - f_{out}(t)dt)}{v_d}
 \end{aligned}
 \tag{3-55}$$

In comparison to Equation 3-53, the only difference for the first two terms in Equation 3-55 is that the amount of queue clearance for link *BA* is  $f_{out}(t)dt$  as a result of limited transportability of link *BA*. The last term in Equation 3-55 represents the journey time for the remaining initial queue, where  $headway \times \frac{f_{out}(t)dt}{n}$  (km) is the travel distance for each vehicle in the initial queue.

As mentioned before, the second component for travel duration is 0. Also, all the entering traffic at *time*  $t+dt$  is stored on link *BA* and joins the initial queue. The time spent on the formation of this queue is computed as:

$$\begin{aligned}
t_3 = & \frac{f_{in}(t)dt \times (L - headway \times \frac{q(t) - f_{out}(t)dt}{n})}{v} \\
& - \frac{headway \times \frac{f_{in}(t)dt}{n} \times (\frac{f_{in}(t)dt}{n} + 1)}{2 \times v} \times n \\
& - \frac{headway \times \text{mod}(f_{in}(t)dt, n) \times (\frac{f_{in}(t)dt}{n} + 1)}{v}
\end{aligned}
\tag{3-56}$$

Where  $L - headway \times \frac{q(t) - f_{out}(t)dt}{n}$  (km) is the length of link in addition to the part that occupied by initial queue;  $headway \times \frac{f_{in}(t)dt}{n}$  (km) is the length of the queue caused from the inflow at time  $t+dt$ , while  $headway \times \text{mod}(f_{in}(t)dt, n)$  (km) is the remainder modulus of the new queue divided by the number of service lanes.

### 3.3.4 Extension for Roadwork

When maintenance is performed the link is decomposed into two parts; the worksite and the rest of the link, as shown in Figure 3-12. Also, the numbers of lanes in service for these two parts, i.e.  $BA$  and  $BC$ , are denoted by  $n1$  and  $n2$ , respectively; and the lengths of them are  $m1$  and  $m2$ .

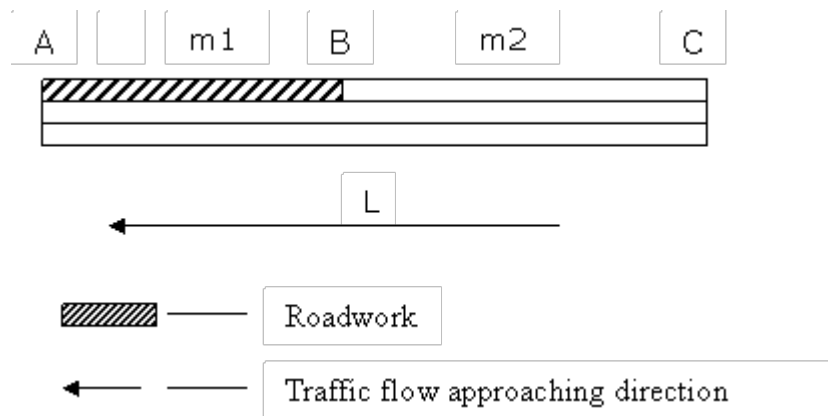


Figure 3-12: A typical road link in maintenance

The exiting flow capacity for link  $CA$  is affected by the worksite  $BA$ , as the service lanes for road users are decreased by at least one lane in this section. It is recognised that a queue is present at  $B$  which propagates back towards  $C$  when the entering flow exceeds the flow capacity of link  $BA$ , this is because there is a sudden reduction of flow capacity at  $B$ . As a result, the evaluation of the journey time for part  $CB$ ,  $t_{CB}$ , follows the same rule as a multi-lane link. Since the queue is not allowed to build up in part  $BA$ , the journey time spent on this section at time  $t+dt$  is computed as:

$$t_{BA} = f_{out}(t)dt \times \frac{L_{BA}}{v_w} \quad (3-57)$$

where  $f_{out}(t)$  (pcu/hr) is the outflow in node  $B$ , and  $v_w$  (km/hr) is the average moving speed on link  $BA$ .

The total travel duration spent on link  $AC$  is achieved as:

$$t_{CA} = t_{CB} + t_{BA} \quad (3-58)$$

### 3.4 Highway Network Performance Metrics

The network performance in NTFM is expressed in outflow, queue and queue propagation for each junction involved in the network. The outflow for a junction is obtained as the sum of the exiting flows through each arm of the junction. The queue for a junction is the sum of queues formed in each lane of each of the entry arms for the junction, which is expressed as queue length. Queue propagation for a junction is indicated as queues at its upstream junctions. In terms of these parameters, the performance metrics for each junction can be travel delay or additional travel cost for road users which is calculated from queue length. In addition, the performance of each road link is evaluated as the travel time spent by vehicles. As a consequence, the performance of the whole network is the aggregation of the performance of all the junctions and links involved, which is demonstrated by the following equations.

The total travel duration spent in the network at time step  $t_k$ :

$$TTD(t_k) = \sum_{i=1}^N T_{D,i}(t_k) + \sum_{j=1}^M T_{T,j}(t_k) \quad (3-59)$$

Where

$T_{D,i}(t_k)$ - delay time spent by road users at junction  $i$  in time  $t_k$  that expressed as a function of queue length, which is determined as

$$T_{D,i}(t_k) = \sum_{a=1}^A \int_{t_{k-1}}^{t_k} q_a(t) dt$$

$q_a(t)$  represents the length of queue on arm  $a$  to junction  $i$  in time  $t$ , and  $A$  is the number of arms at junction  $i$ ; it is noted that the relationship between queue and time is linear

$N$ - the number of junctions on the network

$T_{T,j}(t_k)$ - journey time spent on link  $j$ , which is computed in terms of Section 3.3

$M$ - the number of links on the network

Afterwards, the total road user cost spent on the network is evaluated as:

$$C_u = TTD(t_k) \times v \quad (3-60)$$

$v$ - time value of road user, the UK market price value of time for an average vehicle is £15.38 per hour, 2012 prices and values [61]

Owing to the implementation of roadwork, the transportability of the link must be reduced, which would influence the overall performance of the highway network to a certain degree. Thus, the total travel duration and the total road user cost are provided to indicate the performance of the highway network under normal conditions and maintenance scenarios.

### 3.5 Network Solution Routine

In performing the network level analysis, the inflows at each junction result from the outflows of its upstream junctions and entries to the network; whereas, its outflows are functions of the inflows for the junction. Thus the whole system of traffic flows is interconnected and dependent. It is of vital importance to determine the sequence of node evaluations for traffic flow. These will start from the original node, progress round the network and end at the same node. The simulation of these nodes is processed iteratively until a convergence has been reached when all out profiles are effectively unchanged by further iterations. As for the first iteration, the unknown inputs for each junction in the model are supposed to be zero, which will be updated by the outcomes of its upstream links in the same iteration and applied to the next iteration. Except for the traffic flows from source nodes, traffic flows for each simulation iteration are updated as the inputs for the next iteration. It is worth noting that at the beginning of each iteration queues formed in the network are reset to the condition at the end of previous time period so as to conserve the amount of traffic in the network. For instance, queues are initialized to zero for the simulation iterations at the first time step, since there was no traffic in the network previously.

#### 3.5.1 Convergence of In Profiles and Out Profiles

The basic parameter that is employed to identify the convergence of In and Out profiles for each node, is the average absolute change in individual elements of all In or Out profiles expressed in pcu/hr. These are then averaged to derive the “global” change for the whole network at the end of each iteration. Convergence is assumed to be reached if the “global” change is sufficiently small, i.e. 0.

As for a highway network, the parameters that are used to determine the convergence of traffic condition include the exiting flows for the highway network, the number of vehicles that are stored in the highway network, and the amount of vehicles that propagated back to their upstream junctions.

### 3.5.2 Forecast of Traffic Condition

The purpose of this model is to predict the traffic characteristic for a highway network, including the inflow and outflow for the junctions involved in the network, as well as the travel delay occurred at junctions. At first, it is fed with the initial traffic states, i.e. inflows, turning ratios for junctions and signal control inputs, which are detected by traffic counters on all the links and junctions at each time step. Afterwards, the evaluation of traffic conditions for a highway network falls into two main steps. The first step is to calculate the entering flows, exiting flows and queues for each junction iteratively until those parameters reach a stable state. The next step is to identify the effect of vehicles that are propagating back to their upstream links, which makes the traffic condition even more severe. Consequently, the traffic condition for the highway network at the current time step is obtained, and then this in turn is utilised as the initial traffic condition at next time step. By this means the traffic condition state for a highway network over a time period is evaluated. The whole process of the evaluation of highway network is demonstrated in Figure 3-13, and a more detailed explanation of the NTFM is included in Appendix A.

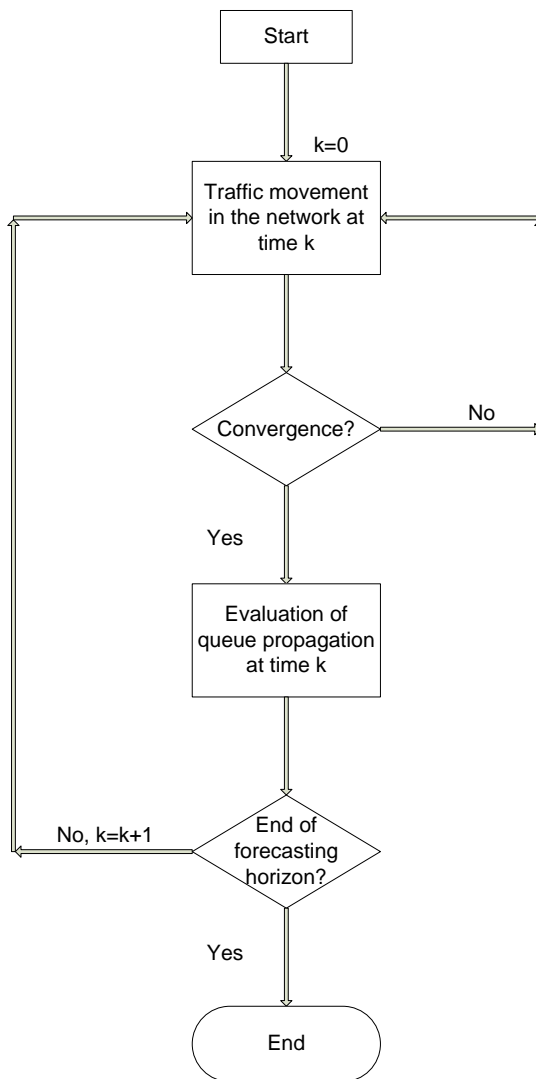


Figure 3-13: The evaluation of Highway Network



### 3.6 Case Study

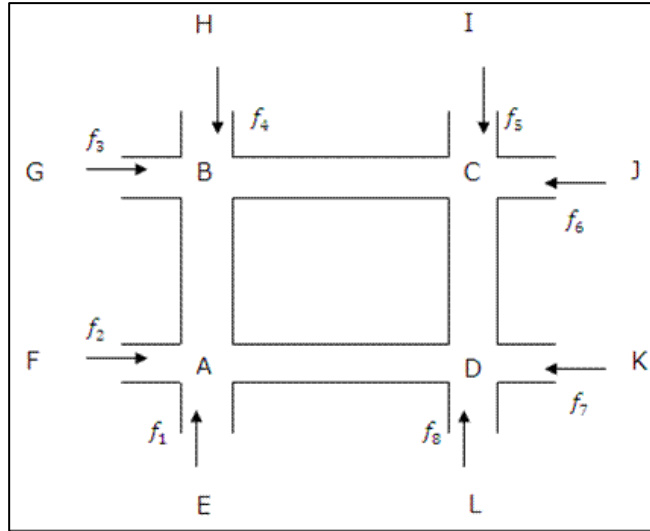


Figure 3-14: The case study network

For ease of exposition, a simple network in Figure 3-14 is provided to illustrate the properties of NTFM. This road network is composed of 4 signalized intersection and 12 dual 2-lane links. The flow capacity for each intersection is defined as 1500 pcu/hr; the capacitance for each inner link is 1000 pcu; the turning ratios for each arm of each junction are defined as 0.3, 0.4 and 0.3 (turning left, going straight, turning right); the green time split for traffic that are turning left and going straight is set to 0.5, and 0.3 for right turning traffic. As  $EA$ ,  $FA$ ,  $GB$ ,  $HB$ ,  $IC$ ,  $JC$ ,  $KD$  and  $LD$  are external links, their capacities are assigned as infinity to avoid queue propagation. In order to illustrate the formation of queues in the network, relatively high traffic flows are employed. The inflows for  $f_1$  to  $f_8$  are 1000, 1000, 1200, 1200, 1400, 1400, 1600, 1600 pcu/hr, respectively. The modelling horizon in this case study is set to 2 time steps, one time step is defined as one hour. Additionally,  $f_{AB}$  is used to denote the flow from node  $A$  to node  $B$ , and  $q_{AB}$  represents the queue that formed at node  $B$  and extended to node  $A$ , which is recognised as a product of  $f_{AB}$ .

### 3.6.1 Network Simulation

The sequence for the evaluation of the nodes in this network is defined as *A-B-C-D-A*. For each iteration the traffic flows in the network are evolved with respect to the following rules until convergence is obtained:

- Junction A,  $f_1$ ,  $f_2$ ,  $f_{DA}$  and  $f_{BA}$  are used to calculate the outflows  $f_{AB}$ ,  $f_{AD}$ ,  $f_{AE}$  and  $f_{AF}$
- Junction B,  $f_3$ ,  $f_4$ ,  $f_{AB}$  and  $f_{CB}$  are used to calculate the outflows  $f_{BC}$ ,  $f_{BA}$ ,  $f_{BG}$  and  $f_{BH}$
- Junction C,  $f_5$ ,  $f_6$ ,  $f_{BC}$  and  $f_{DC}$  are used to calculate the outflows  $f_{CD}$ ,  $f_{CB}$ ,  $f_{CJ}$  and  $f_{CI}$
- Junction D,  $f_7$ ,  $f_8$ ,  $f_{CD}$  and  $f_{AD}$  are used to calculate the outflows  $f_{DA}$ ,  $f_{DC}$ ,  $f_{DK}$  and  $f_{DL}$

It should be noted that  $f_{DA}$ ,  $f_{BA}$ ,  $f_{CB}$  and  $f_{DC}$  are initialised as 0 on the first iteration. After the values of these variables remain steady, queue propagation through the network is evaluated. As a result of this, the traffic condition at the current time step is obtained. Queues formed at this time step are recorded as the initial traffic condition at next time step. In this approach, the traffic condition for this network over the planning span is evaluated.

### 3.6.2 Evaluation of Traffic Condition

At first, the model is fed with the initial traffic states, i.e. inflows, turning ratios for junctions and signal control inputs. The initial traffic condition for the network is depicted in Table 3-3:

Table 3-3: Initial traffic condition in time 1

<i>Junctions</i>	<i>Initial traffic condition in time 1</i>			
A	$q_{AE}$	$q_{AD}$	$q_{AB}$	$q_{AF}$
	0	0	0	0
B	$q_{BA}$	$q_{BC}$	$q_{BH}$	$q_{BG}$
	0	0	0	0
C	$q_{CD}$	$q_{CI}$	$q_{CI}$	$q_{CB}$
	0	0	0	0
D	$q_{DL}$	$q_{DK}$	$q_{DC}$	$q_{DA}$
	0	0	0	0

As for each iteration for time 1, the initial traffic condition states for each junction should be reset according to Table 3-3, so as to keep the balance between traffic inputs and outputs. In order to obtain a stable traffic condition state, the evaluation of the network is run iteratively until convergence is reached, the evolution of traffic for each junction are represented in Table 3-4:

Table 3-4: Evolution of the traffic condition for the case study network in time 1

Junction		Traffic condition states in time 1											
		Entering flow (pcu/hr)				Exiting flow (pcu/hr)				Queue length (pcu)			
A	Iteration	$f_1$	$f_{DA}$	$f_{BA}$	$f_2$	$f_{AE}$	$f_{AD}$	$f_{AB}$	$f_{AF}$	$q_{AE}$	$q_{AD}$	$q_{AB}$	$q_{AF}$
	0	1000	0	0	1000	300	700	700	300	0	0	0	0
	1	1000	1061	789	1000	933	937	1018	962	0	0	0	0
	2	1000	1070	1044	1000	1039	1013	1021	1041	0	0	0	0
	3	1000	1070	1110	1000	1050	1021	1021	1061	0	0	27	0
4	1000	1070	1110	1000	1050	1021	1021	1061	0	0	27	0	
B	Iteration	$f_{AB}$	$f_{CB}$	$f_4$	$f_3$	$f_{BA}$	$f_{BC}$	$f_{BH}$	$f_{BG}$	$q_{BA}$	$q_{BC}$	$q_{BH}$	$q_{BG}$
	0	700	0	1200	1200	789	960	601	570	0	0	90	90
	1	1018	849	1200	1200	1044	1055	983	1005	0	0	90	90
	2	1021	1170	1200	1200	1110	1056	1081	1095	0	69	90	90
	3	1021	1170	1200	1200	1110	1056	1081	1095	0	69	90	90
4	1021	1170	1200	1200	1110	1056	1081	1095	0	69	90	90	
C	Iteration	$f_{DC}$	$f_6$	$f_5$	$f_{BC}$	$f_{CD}$	$f_{CI}$	$f_{CI}$	$f_{CB}$	$q_{CD}$	$q_{CI}$	$q_{CI}$	$q_{CB}$
	0	0	1400	1400	960	1038	705	708	849	0	230	230	0
	1	1089	1400	1400	1055	1066	1070	1167	1170	12	230	230	0
	2	1160	1400	1400	1056	1067	1091	1166	1170	62	230	230	0
	3	1183	1400	1400	1056	1067	1099	1166	1170	78	230	230	0
4	1185	1400	1400	1056	1067	1099	1166	1170	80	230	230	0	
D	Iteration	$f_8$	$f_7$	$f_{CD}$	$f_{AD}$	$f_{DL}$	$f_{DK}$	$f_{DC}$	$f_{DA}$	$q_{DL}$	$q_{DK}$	$q_{DC}$	$q_{DA}$
	0	1600	1600	1038	700	946	1042	1089	1061	400	400	0	0
	1	1600	1600	1067	937	1029	1145	1160	1070	400	400	0	0
	2	1600	1600	1067	1013	1052	1175	1183	1070	400	400	0	0
	3	1600	1600	1067	1021	1054	1179	1185	1070	400	400	0	0
4	1600	1600	1067	1021	1054	1179	1185	1070	400	400	0	0	

Examining Table 3-4, the traffic condition for this network reached a steady state after 5 iterations, as it is quite a small network that only requires a little computational effort. Actually, the simulation is run by 20 iterations to identify the convergence of traffic conditions. Also we found that queues formed at all the four junctions, this is because the entering traffic flow exceeds the corresponding flow capacity. For instance, in terms of the turning ratios for arm  $BA$  of junction  $a$ ,  $f_{BA}$  is divided into two sub-flows for each lane, i.e. 777 pcu/hr, turning left and going straight, in lane 1, and 333 pcu/hr,

turning right, in lane 2. Considering the green splits, the entering flow capacities for these two lanes are 750 pcu/hr and 450 pcu/hr, respectively. As the entering traffic in lane 1 is higher than its entering flow capacity, the portion of the vehicles that exceeds the flow capacity, i.e. 27 pcu, is disturbed in lane 1. For other junctions, because they experienced more traffic than junction A, more serious traffic congestion is suffered.

As for the next time step, the entering flows for the network remains, the only difference is that there are a few queues presented. The initial traffic condition for the network in time step 2 is described in Table 3-5, which is received from the traffic condition at the end of time 1.

Table 3-5: Initial traffic condition in time 2

<i>Junctions</i>	<i>Initial traffic condition in time 2</i>			
A	$q_{AE}$	$q_{AD}$	$q_{AB}$	$q_{AF}$
	0	0	27	0
B	$q_{BA}$	$q_{BC}$	$q_{BH}$	$q_{BG}$
	0	69	90	90
C	$q_{CD}$	$q_{CI}$	$q_{CI}$	$q_{CB}$
	80	230	230	0
D	$q_{DL}$	$q_{DK}$	$q_{DC}$	$q_{DA}$
	400	400	0	0

The initial traffic condition on each iteration in time 2 should be reset according to Table 3-5. The same process is conducted as time 1 and the traffic condition for the network at the end of time 2 is shown in Table 3-6:

Table 3-6: Traffic condition for the case study network in time 2

Junction		Traffic condition states in time 2											
		Entering flow (pcu/hr)				Exiting flow (pcu/hr)				Queue length (pcu)			
A	Iteration	$f_1$	$f_{DA}$	$f_{BA}$	$f_2$	$f_{AE}$	$f_{AD}$	$f_{AB}$	$f_{AF}$	$q_{AE}$	$q_{AD}$	$q_{AB}$	$q_{AF}$
	3,4	1000	1070	1110	1000	1050	1021	1021	1061	0	0	54	0
B	Iteration	$f_{AB}$	$f_{CB}$	$f_4$	$f_3$	$f_{BA}$	$f_{BC}$	$f_{BH}$	$f_{BG}$	$q_{BA}$	$q_{BC}$	$q_{BH}$	$q_{BG}$
	3,4	1021	1170	1200	1200	1110	1056	1081	1095	0	138	180	180
C	Iteration	$f_{DC}$	$f_6$	$f_5$	$f_{BC}$	$f_{CD}$	$f_{CI}$	$f_{CI}$	$f_{CB}$	$q_{CD}$	$q_{CI}$	$q_{CI}$	$q_{CB}$
	3,4	1185	1400	1400	1056	1067	1099	1166	1170	160	460	460	0
D	Iteration	$f_8$	$f_7$	$f_{CD}$	$f_{AD}$	$f_{DL}$	$f_{DK}$	$f_{DC}$	$f_{DA}$	$q_{DL}$	$q_{DK}$	$q_{DC}$	$q_{DA}$
	3,4	1600	1600	1067	1021	1054	1179	1185	1070	800	800	0	0

In comparison to traffic condition in time 1, the entering traffic flows and exiting traffic flows are the same, while the length of all queues in the network are doubled. This is because the traffic inputs for the network remained, which resulted in the weak links suffering more severe congestion.

### 3.6.3 Discussion

Based on the results above, the inflow and outflow for each junction and the queue stored on each link in the case study network are obtained. In the light of these parameters, the total exiting traffic flow and aggregate queue length in the network can be derived by summing up the outflows of all the exits of the network and by summing up the length of queues formed in the network, respectively. The total exiting traffic flow and aggregate queue length are used to measure the transportability of the network, the higher the exiting traffic flow, the better the network transportability, while aggregate queue length is in indirect proportion to the network transportability. Except for the normal road condition, maintenance actions are implemented in the network. NTFM will be deployed to calculate the resulting flow rates in the network to compare

with the original flow rates without maintenance, and additional delay queue length will be evaluated to obtain the cost to road users. Afterwards, by comparing the effects of various maintenance actions on the road network and road users, the best option that resulted in the least maintenance and road user costs will be carried out in the network.

### **3.7 Traffic Rerouting Strategy**

#### **3.7.1 Introduction**

Congestion on motorways and urban roads has significantly reduced the efficiency and transportability of the road network. This has resulted in road users adjusting their pre-defined journey plans to compensate for travel delays, when they occur, and to prevent the possibility that journeys may take significantly longer than anticipated. When excessive traffic congestion exists on a highway network, traffic will take alternative routes and so traffic rerouting takes place. This feature is introduced to NTFM, which is triggered by traffic congestion on the pre-defined route. The queuing on alternative routes is also taken into account. It aims to explore the available flow capacity on alternative routes for the disturbed traffic, and thus minimise the travel delay to road users in congested areas. When the queue formed on any link exceeds a specified threshold, a portion of the incoming traffic will adjust its journey plan if there is available flow capacity on alternative routes.

#### **3.7.2 Methodology**

The traffic rerouting strategy is developed based on the NTFM software capability which is applicable for both motorway and urban roads. One feature of NTFM is that both motorway junctions and urban junctions are taken into account so as to model the

traffic interactions. NTFM forecasts the flow rates and queues formed on any part of a road network. The input configuration of NTFM comprises network geometries, traffic demands, and flow ratios at junctions. In NTFM, traffic rerouting is applied to allow vehicles to circumvent the congested links by detouring on the shortest alternative route. As NTFM works with the turning ratios at junctions, it is realised by varying the traffic demands and flow ratios at the junctions on the alternative routes, while the demands and flow ratios on the rest of the network remain unchanged.

An iterative simulation method is applied to evaluate the rerouted traffic in the network.

The general algorithm is described as follows:

1. NTFM is deployed to simulate traffic movement in the network until the traffic condition has reached convergence, and then to identify congested links and the traffic to be rerouted.
2. Then the source nodes of queues are verified so as to identify the shortest alternative route for each congested link.
3. Determine the amount of the rerouted traffic which should be less than the available flow capacity on the shortest alternative route. In this manner, there would be no queue built up on the alternative route.
4. List the exit numbers for each junction on the shortest alternative route for each congested link.
5. Update flow on first leg of the alternative route.
6. Keeping flow at each junction on the alternative route, realised by updating the outflow proportions at each junction to direct traffic on the alternative route.
7. Repeat Steps 5 and 6 for each other junction on the alternative route.
8. Stop when the end node of the congested link is reached.



9. NTFM is used again to evaluate the traffic in the network on the basis of the updated flow proportions. If traffic congestion still exists on the network, the algorithm return to step 2 and the same procedure is applied.
10. The updated flow proportions should only be temporally and be returned to their original values to the traffic is fine once the queue has completely dissipated.

Within the traffic rerouting strategy, the evaluation of the traffic at the junctions on the pre-defined route and the shortest alternative route is explained below. A typical 4 arm intersection is represented in Figure 3-15, and the data required for traffic evaluation is listed as:

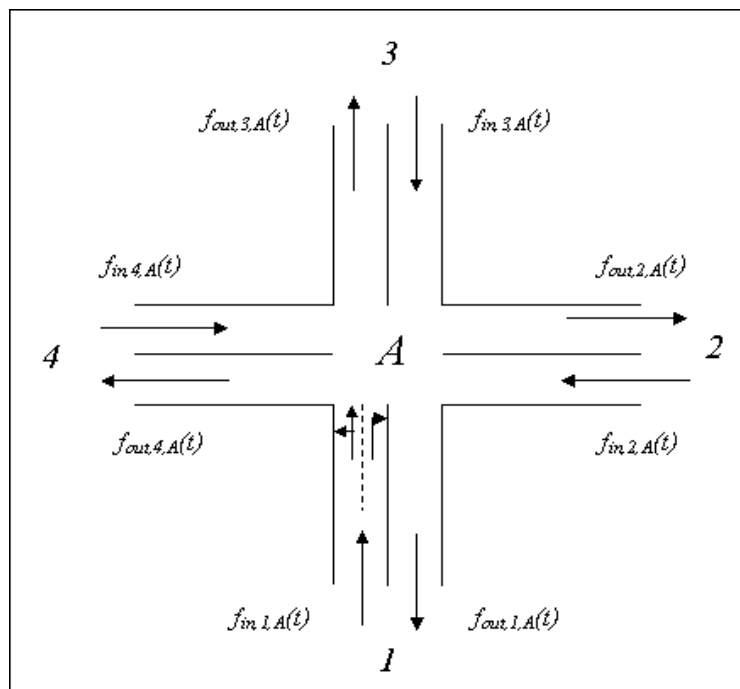


Figure 3-15: A typical 4 arm intersection

$f_{in,i,A}(t)$ - flow at node  $A$  coming from direction  $i$ , i.e. 1, 2, 3, 4 (1 represents the northbound arm, while the rest are used to represent the westbound arm, southbound arm and eastbound arm, respectively), at time  $t$  (pcu/hr)

- $d_{i,A,j}(t)$ - proportion of flow on arm  $i$  for node  $A$  choosing the outflow direction  $j$ ,  $j$  is expressed in direction  $l$  (left),  $a$  (ahead) and  $r$  (right)
- $c_{i,A,l}$  - node  $A$  flow capacity for the traffic coming from direction  $i$  and going in lane  $l$ , i.e. 1 (left lane), 2 (right lane), (pcu/hr)
- $cp_{i,A,l}$  - link capacitance in lane  $l$ , i.e. the maximum number of cars which can queue in lane  $l$ , of the arm  $i$  for node  $A$  (pcu)
- $f_{out,i,A}(t)$ - flow at node  $A$  that going to direction  $i$  at time  $t$  (pcu/hr)
- $q_{i,A,l}(t)$ - average number of vehicles queuing in lane  $l$  of arm  $i$  for node  $A$  at time  $t$  (pcu)
- $F_A(t)$ - the overall inflow at node  $A$  at time  $t$  (pcu/hr). It is calculated as
- $$F_A(t) = \sum_{i=1}^4 f_{in,i,A}(t)$$
- $T_A(t)$ - the overall outflow at node  $A$  at time  $t$  (pcu/hr), obtained as
- $$T_A(t) = \sum_{i=1}^4 f_{out,i,A}(t)$$
- $Q_A(t)$ - amount of the vehicles queuing at node  $A$  at time  $t$  (pcu), evaluated as
- $$Q_A(t) = \sum_{i=1}^4 \sum_{l=1}^2 q_{i,A,l}(t)$$
- $p_{i,A}(t)$ - proportion of the overall traffic at node  $A$  that going to direction  $i$ , calculated as  $p_{i,A}(t) = \frac{f_{out,i,A}(t)}{T_A(t)}$
- $p_{i,A,j}(t)$ - proportion of the traffic to be rerouted on arm  $i$  for node  $A$  choosing the outflow direction  $j$ ,  $j$  is expressed in direction  $l$  (left),  $a$  (ahead) and  $r$  (right). If the traffic is delayed in lane 1, the proportions for each direction is evaluated as  $p_{i,A,l}(t) = \frac{p_{i,A,l}(t)}{p_{i,A,l}(t) + p_{i,A,a}(t)}$ ,

$$p_{i,A,a}(t) = \frac{p_{i,A,a}(t)}{p_{i,A,l}(t) + p_{i,A,a}(t)} \text{ and } p_{i,A,r}(t) = 0 ; \text{ otherwise } p_{i,A,l}(t) = 0 ,$$

$$p_{i,A,a}(t) = 0 \text{ and } p_{i,A,r}(t) = 1$$

The conservation law of traffic flow at any junction, e.g. junction  $A$ , on the network is formulated as:

$$F_A(t)dt + Q_A(t) = T_A(t)dt + Q_A(t + dt) \quad (3-61)$$

Noting that when there is no vehicle stored at junction  $A$  at time  $t$ , the values of  $Q_A(t)$  and  $Q_A(t+dt)$  will be 0, and then  $F_A(t)$  is equal to  $T_A(t)$ .

Three cases are considered in evaluation of the rerouted traffic at junction  $A$ .

### 3.7.2.1 Source Node

If junction  $A$  is the source node of the rerouted traffic, and the rerouted traffic,  $rt$  (pcu/hr), is transferred from direction 1 to 4, the traffic at junction  $A$  needs to be modified:

- As there is no queue stored at junction  $A$ ,  $F_A(t)=T_A(t)$ . The initial outflow

proportions for junction  $A$  is evaluated as 
$$\begin{bmatrix} p_{1,A}(t) \\ p_{2,A}(t) \\ p_{3,A}(t) \\ p_{4,A}(t) \end{bmatrix} = \begin{bmatrix} f_{out,1,A}(t) \\ f_{out,2,A}(t) \\ f_{out,3,A}(t) \\ f_{out,4,A}(t) \end{bmatrix} / F_A(t), \text{ which}$$

can be expressed as  $P_A(t) = f_A(t) / F_A(t)$ .

- After traffic rerouting, outflows on arm 1 and 4 are varied by  $rt$ , calculated as  $f_{out,,iA}(t)-rt$  and  $f_{out,,iA}(t)+rt$ , respectively.
- The outflow proportions are updated in terms of

$$\begin{bmatrix} p'_{1,A}(t) \\ p'_{2,A}(t) \\ p'_{3,A}(t) \\ p'_{4,A}(t) \end{bmatrix} = \begin{bmatrix} f'_{out,1,A}(t) \\ f'_{out,2,A}(t) \\ f'_{out,3,A}(t) \\ f'_{out,4,A}(t) \end{bmatrix} / F_A(t), \text{ also expressed as } P'_A(t) = f'_A(t) / F_A(t).$$

### 3.7.2.2 Other Junction on the Alternative Route

If junction  $A$  is located on the shortest alternative route, and the traffic to be rerouted entered it on arm 2 and exited through arm 1, the traffic is evaluated as:

- Calculate the initial outflow proportions  $P_A(t)$ .
- When traffic rerouting is performed, both the overall inflow and outflow at junction  $A$  are increased to  $F_A(t) + rt$ .
- The outflow on arm 1 is raised to  $f_{out,1,A}(t) + rt$ .
- The updated outflow proportions are obtained as

$$\begin{bmatrix} p'_{1,A}(t) \\ p'_{2,A}(t) \\ p'_{3,A}(t) \\ p'_{4,A}(t) \end{bmatrix} = \begin{bmatrix} f'_{out,1,A}(t) \\ f'_{out,2,A}(t) \\ f'_{out,3,A}(t) \\ f'_{out,4,A}(t) \end{bmatrix} / (F_A(t) + rt) \text{ or } P'_A(t) = f'_A(t) / F'_A(t).$$

### 3.7.2.3 End Node on the Alternative Node

If junction  $A$  is the end node on the alternative route and the initial queue is stored on arm 3, while arm 4 is on the alternative route.

- Calculate the initial outflow proportions  $P_A(t)$ .
- After traffic rerouting, the overall outflow is increased to  $T_A(t) + rt$ .
- Due to traffic rerouting, the outflow on arm 1, 2 and 4 are updated as:

$$f'_{out,1,A}(t) = f_{out,1,A}(t) + rt \times p_{3,A,2}$$

$$f'_{out,2,A}(t) = f_{out,2,A}(t) + rt \times p_{3,A,1}$$

$$f'_{out,4,A}(t) = f_{out,4,A}(t) + rt \times p_{3,A,3}$$

- The updated outflow proportions are represented by

$$\begin{bmatrix} p'_{1,A}(t) \\ p'_{2,A}(t) \\ p'_{3,A}(t) \\ p'_{4,A}(t) \end{bmatrix} = \begin{bmatrix} f'_{out,1,A}(t) \\ f'_{out,2,A}(t) \\ f'_{out,3,A}(t) \\ f'_{out,4,A}(t) \end{bmatrix} / (T_A(t) + rt) , \text{ or } P'_A(t) = f'_A(t) / T'_A(t).$$

As for other junctions on the network, the traffic demands and flow ratios remain unchanged. By this means the traffic condition state for a highway network in consideration of traffic rerouting over a time period is evaluated. It is noted that not only 4-arm intersections, but 2-arm intersections and 3-arm intersections, e.g. merge and T-junction, are also widely used in transportation, the methodology described above can be applied to evaluate the traffic at such intersections, where the outflow proportion for one or more direction is defined as 0.

### 3.7.3 Case Study

For convenience of illustration, a simple road network is provided to investigate the variation of flow rates and proportions on any part of the network due to traffic rerouting. In this study the example network is signal controlled, so the traffic interactions among the flows from different sources are eliminated; and the links on the network belong to dual 2 lane carriageway.

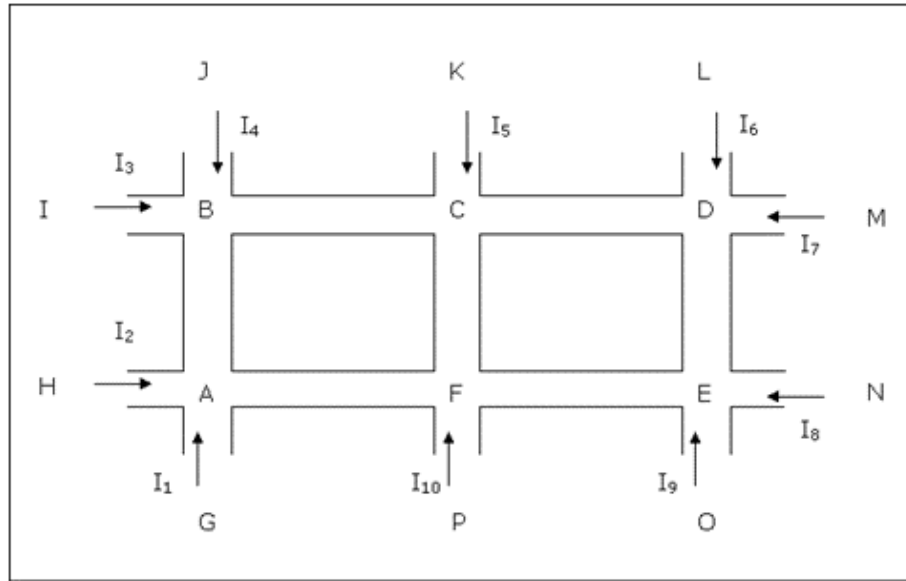


Figure 3-16: A simple road network

Data necessary for this network is described below:

$I_i(t)$ - inflow at external node  $i$  at time  $t$  (pcu/hr, pcu stands for passenger car unit)

$O_i(t)$ - outflow at external node  $i$  at time  $t$  (pcu/hr)

$fc_{i,A,l}$ - available flow capacity in lane  $l$  of arm  $i$  at junction A, calculated as

$$fc_{i,A,1}(t) = cp_{i,A,1} - f_{in,i,A}(t) \times (d_{i,A,l} + d_{i,A,a}) \text{ and } fc_{i,A,2}(t) = cp_{i,A,2} - f_{in,i,A}(t) \times d_{i,A,r} \quad (3-62)$$

The turning ratios on each arm for the junctions on this network are set to 0.3 for direction  $l$  (left), 0.4 for direction  $a$  (ahead) and 0.3 for direction  $r$  (right). The links  $CD$  and  $FE$  are defined as 0.7 km, while the rest links are 0.35 km; and then the capacitance for the service lane on each link is calculated in terms of the following Equation:

$$cp = \frac{L}{\text{headway}} \quad (3-63)$$

where headway is defined as 0.007 km.

The flow capacity for junction  $C$  is fixed to 2000 pcu/hr, while the flow capacity for other junctions is 3000 pcu/hr, and the green splits for the traffic that going straight and

turning left and the traffic turning right are defined as 0.5 and 0.3. The inflows (pcu/hr) for the example network are represented in Table 3-7:

Table 3-7: Inflows (pcu/hr) for the example network

$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$
1200	1200	1200	1200	1200	1200	1200	1200	1200	2000

Based on the parameters described above, NTFM is used to evaluate the network performance during a period of one hour, and the outflow proportions for each junction are obtained, which are depicted below:

Table 3-8: Traffic condition at each junction without traffic rerouting

Junction A								
Arm	1 (GA)		2 (FA)		3 (BA)		4(HA)	
$f_{in,i,A}$	1200		1448		1224		1200	
$f_{out,i,A}$	1285		1207		1274		1306	
$p_{i,A}$	0.2533		0.238		0.2512		0.2575	
$q_{i,A,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	0	0	0	0	0	0	0	0

Junction B								
Arm	1 (AB)		2 (CB)		3 (JB)		4(IB)	
$f_{in,i,B}$	1274		1278		1200		1200	
$f_{out,i,B}$	1224		1222		1253		1253	
$p_{i,B}$	0.2472		0.2468		0.253		0.253	
$q_{i,B,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	0	0	0	0	0	0	0	0

Junction C								
Arm	1 (FC)		2 (DC)		3 (KC)		4(BC)	
$f_{in,i,C}$	1525		1222		1200		1222	
$f_{out,i,C}$	1213		1306		1304		1278	
$p_{i,C}$	0.2378		0.256		0.2556		0.2506	
$q_{i,C,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	68	0	0	0	0	0	0	0

Junction D								
Arm	1 (ED)		2 (MD)		3 (LD)		4(CD)	
$f_{in,i,D}$	1274		1200		1200		1306	
$f_{out,i,D}$	1232		1264		1262		1222	
$p_{i,D}$	0.2474		0.2538		0.2534		0.2454	
$q_{i,D,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	0	0	0	0	0	0	0	0

Junction E								
Arm	1 (OE)		2 (NE)		3 (DE)		4(FE)	
$f_{in,i,E}$	1200		1200		1232		1447	
$f_{out,i,E}$	1287		1308		1274		1210	
$p_{i,E}$	0.2534		0.2575		0.2509		0.2382	
$q_{i,E,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	0	0	0	0	0	0	0	0

Junction F								
Arm	1 (PF)		2 (EF)		3 (CF)		4(AF)	
$f_{in,i,F}$	2000		1210		1213		1207	
$f_{out,i,F}$	1210		1447		1525		1448	
$p_{i,F}$	0.2149		0.257		0.2709		0.2572	
$q_{i,F,l}$	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2	Lane 1	Lane 2
	0	0	0	0	0	0	0	0



In the light of the traffic condition at each junction, the inflows and outflows at each external node on the network are demonstrated in Table 3-9:

Table 3-9: Traffic inputs and outputs (pcu/hr) for the example network

Node	G	H	I	J	K	L	M	N	O	P	Total
$I_i$	1200	1200	1200	1200	1200	1200	1200	1200	1200	2000	12800
$O_i$	1285	1306	1253	1253	1304	1262	1264	1308	1287	1210	12732

It was found that a queue formed in lane 1 of arm 1, i.e. link  $FC$ , at junction  $C$ . Based on the flow ratios on this arm, the inflows in lane 1 and lane 2 on link  $FC$  are calculated as 1068 and 457 pcu/hr; where the inflow in lane 1 exceeds the corresponding flow capacity, 1000 pcu/hr, therefore 68 vehicles are queued on link  $FC$ . In terms of Equation 3-63, the capacitance for lane 1 on link  $FC$  is calculated as 50 pcu, thus the disturbed traffic filled the current link and propagated back to upstream links. In order to avoid the traffic congestion occurred on link  $FC$ , road users would seek any other routes to leave junction  $C$ . Two alternative routes are applicable, route  $F-A-B-C$  and route  $F-E-D-C$ ; as the former is shorter, it is ascertained as the diversion route for link  $FC$ .

The next step is to determine the amount of the traffic to be rerouted. The alternative route is composed of lane 2 of link  $FA$ , lane 2 of link  $AB$  and lane 1 of link  $BC$ , and the available flow capacities on them are obtained as 466, 518 and 645 pcu/hr using Equation 3-62. As the disturbed traffic is less than the least available flow capacity on the alternative route, all the vehicles in the queue can be diverted.

Junction  $F$  is recognised as the source of the rerouted traffic, and the rerouted traffic is transferred from direction 3 to direction 4. In terms of the equations developed in Section 3.7.2.1, the traffic conditions at this junction are varied as:

Table 3-10: Traffic condition at junction F with traffic rerouting

Junction F				
Arm	1 (PF)	2 (EF)	3 (CF)	4(AF)
$f_{out,i,F}$	1210	1447	1525	1448
$f'_{out,i,F}$	1210	1447	1457	1516
$p_{i,F}$	0.2149	0.257	0.2709	0.2572
$p'_{i,F}$	0.2149	0.257	0.2588	0.2693

The rerouted traffic entered junction A on arm 2 and exited through arm 3. Consequently, the outflow on arm 3 is increased by 68 pcu/hr, while the outflows on other arms are stable. In terms of Section 3.7.2.2, the traffic at junction A is updated as:

Table 3-11: Traffic condition at junction A with traffic rerouting

Junction A				
Arm	1 (GA)	2 (FA)	3 (BA)	4(HA)
$f_{out,i,A}$	1285	1207	1274	1306
$f'_{out,i,A}$	1285	1207	1342	1306
$p_{i,A}$	0.2533	0.238	0.2512	0.2575
$p'_{i,A}$	0.25	0.2348	0.2611	0.2541

Due to the accommodation of the rerouted traffic, the traffic at junction B is evaluated as:

Table 3-12: Traffic condition at junction B with traffic rerouting

Junction B				
Arm	1 (AB)	2 (CB)	3 (JB)	4(IB)
$f_{out,i,B}$	1224	1222	1253	1253
$f'_{out,i,B}$	1224	1290	1253	1253
$p_{i,B}$	0.2472	0.2468	0.253	0.253
$p'_{i,B}$	0.2438	0.257	0.2496	0.2496

As for junction *C*, the disturbed traffic on arm 1 exited the junction by detouring on the alternative route. As the queue at junction *C* is located in lane 1 of arm 1, the destination for the rerouted traffic would only be junctions *B* and *K*. According to the flow proportions in lane 1 of arm 1, the flows to each direction are calculated as 29 and 39 pcu/hr, respectively. The updated traffic condition at junction *C* is represented as:

Table 3-13: Traffic condition at junction *C* with traffic rerouting

Junction C				
Arm	1 (FC)	2 (DC)	3 (KC)	4(BC)
$f_{out,i,C}$	1213	1306	1304	1278
$f'_{out,i,C}$	1213	1306	1343	1307
$p_{i,C}$	0.2378	0.256	0.2556	0.2506
$p'_{i,C}$	0.2347	0.2527	0.2598	0.2528

After the evaluation of the rerouted traffic, NTFM is applied again to evaluate the traffic in the network based on the updated outflow proportions for each junction rather than the flow ratios on each link so as to consume less computational effort. The updated inflows and outflows for the network are depicted in Table 3-14:

Table 3-14: Traffic inputs and outputs (pcu/hr) for the example network with traffic rerouting

Node	G	H	I	J	K	L	M	N	O	P	Total
$I_i$	1200	1200	1200	1200	1200	1200	1200	1200	1200	2000	12800
$O_i$	1285	1306	1270	1265	1343	1262	1264	1308	1287	1210	12800

The traffic inputs for the network are equal to the outputs. It means that the disturbed traffic left the network by detouring on the shortest alternative route. Consequently, traffic rerouting strategy is realised in NTFM by varying the flow rates and ratios on the alternative routes. It is worth noting that when the traffic condition on the network improved the outflow proportions should be reset to their original values.

### 3.8 Summary

In this chapter, we have proposed a macroscopic network level traffic flow model and its associated junction models, which are based on a queue model and follow the right of way rules. The model calculates the flows through the junctions and the queues which build up and disperse at different points during the day. It has also been demonstrated how to calculate the journey spent on road links and how to measure network performance according to the aggregate queue length and additional road user cost.

The modelling capability developed provides advances on the previously developed macroscopic traffic flow models in the following features:

- a. It accounts for both motorway and urban roads in the same road network reflecting the interactive nature of the two systems.
- b. It copes with two-way traffic flow by employing iterative simulation method to determine the value of the dependent traffic flows in the network.
- c. Entry and exit points in the model are used to model traffic flow along each urban network section link (road). This simulates traffic exiting/joining the network at housing estates or work place locations.
- d. It deploys shared lane to illustrate the traffic interaction among mixed directional traffic flows.
- e. Traffic rerouting strategy is developed to model the travel behaviour when traffic congestion is present in the network.

The results showed that this model has the capability to describe the evolution of dependent traffic flows and forecast the traffic movement and queue dynamics through a simple case study network, as well as model the traffic diversion taken place in the network when a queue is presented.

## **4 Literature Review on Highway Asset Modelling**

### **4.1 Introduction**

As highways agencies have turned their attention from construction of highway networks to pavement maintenance and rehabilitation (M&R), the need for optimal M&R strategy is emerging. Any M&R activity aims to improve the pavement condition and performance, and can have two effects on the pavement: an immediate influence on the pavement condition and an impact on the future rate of pavement deterioration.

M&R activities for a pavement network can be classified in terms of their frequency and their impact on the serviceability of the road. Minor maintenance, surface treatments and major maintenance are the main categories of the pavement maintenance activities. Minor maintenance activities are defined as localized repairs of pavement surface and shoulder defects, maintenance of road drainage, side slopes and roadsides. Surface treatments of the existing pavement are used to restore pavement structural strength and integrity in order to ensure serviceability of the pavement. Major maintenance actions are performed to restore or improve the existing pavement strength beyond its original strength, such as resurfacing, deep overlays, reconstruction and new construction.

At the network level, an M&R strategy is a set of operations applied to a pavement network or segment during its service life, which involve some of the M&R operations stated above. M&R operations can be applied individually or in combination each year. As for the M&R strategy, it should be evaluated by analysing its associated benefits and costs for the whole pavement system. A typical life cycle of a pavement is presented in Figure 4-1, which illustrates the effects of different maintenance actions on the pavement condition.

This chapter is organised as follows. Some widely adopted pavement maintenance actions are explained in Section 4.2. Section 4.3 covers the selection of M&R strategy for a pavement system. In Section 4.4 we describe the philosophy of Genetic Algorithms (GAs) and, as well as some advanced GA techniques.

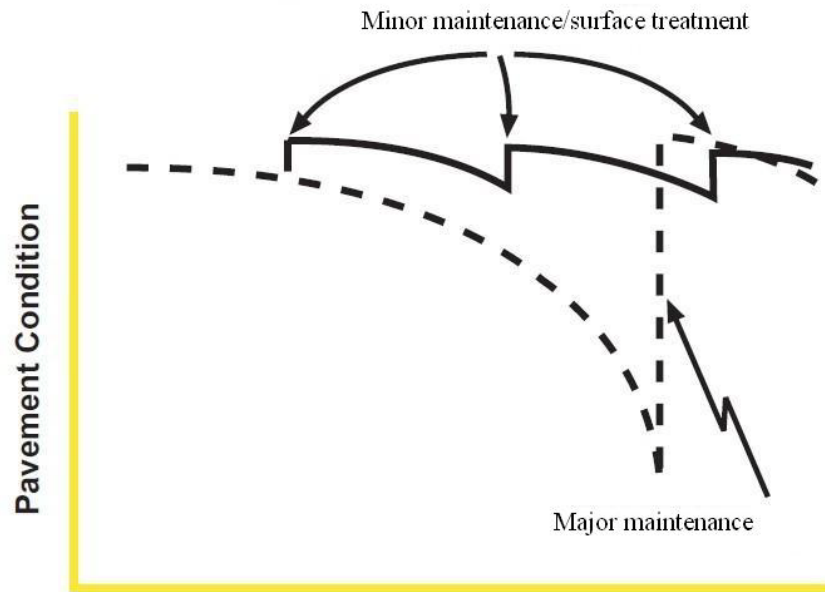


Figure 4-1: Typical pavement life cycle

## 4.2 M&R Operations for Pavement

In order to maintain the structural and functional performance of the pavement, various M&R operations are adopted. Those operations are applied to pavements according to the occurred distress types and corresponding severity levels. The following maintenance operations are usually conducted in the restoration of pavements [62].

#### 4.2.1 Minor Maintenance

##### 4.2.1.1 Potholing Filling

Potholing is recognized as the loss of material from the surface layer of the pavement because of water freezing and thawing beneath the surface layer, which has been identified as a severe condition. Until routine maintenance, e.g. patching, is conducted, potholing filling is used as a temporary repair to remedy the pavement condition.

##### 4.2.1.2 Patching

Patching is usually conducted to replace the flawed materials in the pavement surface. The first step for patching is to mark the pavement in square or rectangle patterns beyond the area of potholing, and then to remove the damaged marked asphalt. The next step is to clean and treat the exposed lower layer, which is followed by the placement of new asphalt material.

#### 4.2.2 Surface Treatments

##### 4.2.2.1 Crack Treatment

Crack treatment is an effective way to treat the pavement surface using bituminous materials, including crack sealing and crack filling. Crack sealing is conducted by injecting bituminous materials into cracks; crack filling is the placement of emulsion and aggregate or fine graded hot mix asphalt into cracks. Crack sealing is usually conducted on the relatively good quality pavement where the cracks are less than 20 mm in width. Crack filling is used to deal with the cracks with width more than 20 mm. Crack filling is regarded as a longer-term treatment than crack filling.

#### 4.2.2.2 Surface Dressing

Surface dressing is performed by spraying an emulsion bituminous binder to the surface of the pavement, and aggregate chippings are spread and rolled to make sure chippings are embedded into the pavement surface [63]. Surface dressing protects the pavement from infiltration of water and oxidation by sealing the surface and restores pavement skid resistance.

#### 4.2.2.3 Thin Surfacing

Thin surfacings are manufactured by machine laying a hot bituminous bound mixture onto an emulsion tack or bond coat, which are typically 15-40 mm thick [64]. They are used to perform minor regulation of existing pavement surfaces, e.g. remove rutting, increase pavement texture and improve skid resistance.

### 4.2.3 Major Maintenance

#### 4.2.3.1 Pavement Overlay

Pavement overlays are usually applied to the pavements in a poor condition that can improve the strength of pavement [64]. Typically thin overlay or deep overlay is applied. Thin overlay is defined as the overlay thickness is less than 50mm. It is usually applied to relatively sound pavement to provide a new protective surface and improve pavement serviceability. However, it has no significant influence on the structural strength of the pavement. As for deep overlay, it aims to improve pavement condition to a large extent, thus to ensure a better pavement quality. Due to their significant impact on the pavement condition and performance, pavement overlays are termed major maintenance.



#### 4.2.3.2 Resurfacing

Resurfacing can be carried out if the highway engineers determine that an existing pavement is in reasonably good condition. It is used to increase the strength of pavement structure, compensate for loss of skid resistance, as well as remove defective materials and enhance riding quality.

#### 4.2.3.3 Reconstruction

Under the condition that the pavement has exceeded its service life, reconstruction is provided to renew the existing pavement. Reconstruction is recognised as major maintenance intervention. It is applied as a full-depth, full-length reconstruction of existing pavement on its road alignment, including rehabilitation of all drainage systems.

#### 4.2.3.4 New Construction

New construction is also categorised as major maintenance, which involves a full-width, full length construction of new pavement on a road alignment in parallel to the existing pavement, updating an existing pavement to a higher standard, and providing additional lanes to the existing pavement.

### **4.3 Decision Making Strategies for M&R**

An M&R strategy is a set of M&R operations applied to a pavement network or a segment during its service life, which includes surface treatment, minor and major treatments. Each M&R strategy is associated with the present value cost and present value benefit of the pavement network or segment. The benefit can be measured by the impact of each M&R on the performance of the pavement. Typically, there are two

main approaches applied in the evaluation of M&R strategy, the prioritisation model and the optimisation model.

#### 4.3.1 M&R Prioritisation Model

For the prioritisation model, the first step is to build an inventory of pavement conditions of all the segments on the network, and then allocate M&R funds to pavement segments that are in the worst condition or experience an accelerated deterioration rate. As it will be described in the optimisation model below, it is possible to postpone the M&R operations of the pavement segment to later years. The benefits are evaluated making a comparison between the present value of the pavement condition of the segments maintained without a delay and with a delay.

Prioritisation method is the traditional way applied to the decision making, which is defined as the “bottom-up” approach because the M&R treatments for each project are evaluated first and then they are prioritized and conducted according to the needs and resources of the highways agency [65].

In the pavement management system (PMS), the prioritisation model is used to identify the corresponding M&R treatments for each pavement segment in the pavement network individually. Further, with the incorporation of pavement performance prediction model in the PMS, the prioritisation model is able to estimate when each segment of the network will reach the threshold of its service level. The selection of M&R treatments for each segment is usually evaluated according to pavement distress measures or functional performance indices. After the identification of projects for each segment in the network, their ranking needs to be estimated based on the agency decision criterion. The ranking methods of the projects can be categorised into different general groups in terms of various pavement condition measures [66]. The first group of

prioritisation methods is mainly based on the present value of the current pavement condition, which involves [67]:

- Prioritisation based on functional performance indices. In this approach, a functional performance index is applied to evaluate the pavement condition and develop a priority index together with other factors. The mostly adopted functional performance indices include pavement serviceability index (PSI) and pavement condition index (PCI), which are determined by visual assessment of the various pavement surface distresses;
- Prioritisation based on pavement distress. The priority in this approach is determined by the amount that the measured pavement distress exceeds its relevant intervention level, the larger the exceedance the higher the priority. All the individual pavement distresses are combined in a functional relationship as the priority index. The higher value of the priority index represents the higher priority for the corresponding project;
- Priority based on treatment. The priority for each project is related to the maintenance treatment programmed in terms of its influence on the pavement condition, together with road class and its traffic volume considered as the basis for defining priority. In this case, the heavy traffic roads are given higher priority.

As for the second group of prioritisation methods, it considers both current and future pavement condition [67]. The future condition for each pavement segment is achieved by the pavement performance prediction model and evaluated in terms of its related costs and benefits. The most commonly adopted approach among this group is the cost-effective method, which is obtained as the ratio of the area under the pavement

performance versus time curve to the M&R treatment life-cycle cost [68]. Thus, the priority for each project is defined by its generated cost effectiveness.

After the implementation of the priority rank of the pavement network, projects can be conducted according to their ranking under the budget constraint. If the total budget requirements identified exceed available budget, the projects that cannot be performed in the current year should be deferred to later years.

The superiority of prioritisation models is that it is simple, easy to understand and can be implemented quickly. However, there are several inbuilt disadvantages of the models [69]. Firstly, they do not yield an optimal M&R strategy since the decisions are determined at the project level, which results in that the final network level decisions are just the combination of a set of project level decisions [65]. Secondly, decisions are conducted sequentially according to the priority rank without the consideration of their overall influence on the pavement network; they are not capable of estimating the consequences of the pavement network after the implementation of treatments. Therefore, the prioritisation models are only suitable for short-term decision making problems.

#### 4.3.2 M&R Optimisation Model

The determination of an optimal M&R strategy is the main objective of the optimisation model, which is dependent on the decision policy selected, such as maximizing the proportion of “good” pavements and minimizing the proportion of “bad” pavements subject to budget constraints, or minimizing the total life-cycle cost under the constraint of pavement quality [70]. Both deterministic and probabilistic pavement performance prediction models can be incorporated into the optimisation model. The deterministic models utilize the regression relationships to describe the deterioration of the pavement condition, which expresses the pavement performance in detailed and quantitative terms,

with reference to the various pavement surface distresses that characterise the pavement condition [71]. Probabilistic models employ Markov models to forecast pavement performance [72]. The latter models are more preferable because they recognise the stochastic nature of the pavement system with the consideration of the uncertainty in inspection and prediction of pavement condition.

The optimisation models can be applied at both network level and project level pavement management system. In PMS, network level decisions are made first, followed by project level decisions. The network level PMS should consider the following tasks: (1) identifying the overall pavement network performance in terms of physical pavement condition and pavement service level over time; (2) developing an estimate of budget requirements and allocating resources over the entire pavement network; (3) establishing a set of pavement construction, maintenance and rehabilitation treatments. The network level decisions only provide the budget allocation and maintenance strategies for the portions of the pavement network, but lack of a list of specific treatments for individual pavement segments. The project level analysis is needed for individual segments based on the network level analysis, which involves: (1) monitoring and estimating the pavement performance at specific sites; (2) selecting appropriate remedial actions for pavement segments and evaluating their cost and impact on pavement performance; (3) establishing maintenance strategies for individual pavement segments for a finite planning horizon under constraints.

#### 4.3.2.1 Segment-linked Optimisation Model

In the case of network level maintenance optimisation model, there is no description of pavement condition and performance in detailed and quantitative terms, and the M&R treatments selected are assigned to portions of the pavement systems. No specific treatment is applied on individual road segments. Therefore, segment-linked

optimisation model is employed to overcome such problem [71, 73]. One feature of this model is that the M&R activities are associated with road segments, and not just with road categories. The model explicitly recognizes the segments of the pavement network where the M&R activities should be conducted. It means that pavement segments are defined as the decision-making units to which the M&R treatments apply. In general, the size of pavement segment can be various. However, if the segments are too long, their pavement condition cannot be predicted accurately, which present low homogeneity than short segments. In contrast, if they are too short, it may result in that the M&R treatments achieved from the model may make pavement condition vary excessively within the pavement system. Also, it makes the optimisation model more intensive.

Due to the decomposition of pavement network into separate segments, the pavement system becomes much more intensive and complicated. For instance, even for a 20-segment road network, a 5-year planning horizon and 9-alternative M&R activities optimisation problem, the number of alternative M&R strategies to be considered is huge (approximately  $2.66 \times 10^{95}$ ). This model may be very complicated and time-consuming for the integer linear programming, so GAs and some other soft computing approaches are usually applied to solve this kind of optimisation problem.

#### 4.3.2.2 Multi-Objective Optimisation Model

In a typical PMS, there is often more than one objective that needs to be achieved, and those multiple objectives, often conflicting, have significant different effects on the resulting M&R strategy. Traditionally, researchers and practitioners treat this kind of problem as a single objective optimisation problem while imposing other competing objectives as constraints. There are typically two limitations of the single objective optimisation [74]. Firstly, it is important to select the objective for optimisation among

the involved objective. Also, the proper ranges for the objectives that are set as constraints in the formulation need to be defined. Consequently, the solutions achieved from the single-objective optimisation model are suboptimal compared to the ones obtained from the multi-objective optimisation model.

The multi-objective optimisation model aims to find a vector of decision variables which optimizes the involved objective functions and satisfies the constraints [74]. Typically, the parameters that are involved in the pavement maintenance optimisation problem include budget, time, pavement quality requirement and work delivery, etc. The constraints of those factors are employed to ensure the feasibility of the final global optimal solution. However, there is no single global optimal solution for the multi-objective optimisation model. Therefore, the means to solve this model is to develop Pareto optimal solutions for all involved competing objectives, and then to identify a Pareto frontier by the selection of all the non-dominated solutions. Then the acceptable solution in line to the decision makers' criteria is selected. There are many optimisation methods applied in the generation of Pareto optimal solutions, such as the multi-objective simplex method [75], GAs [76], the weighting sum method [77] and the normal constraint method [78], which are the most commonly adopted ones [74]. However, there is no single optimisation method that is superior to other optimisation methods. The selection of the optimisation method depends on the preference of users and information available.

Multi-objective simplex method is applied to solve multi-objective linear problems subjected to multiple linear constraints, which is not sufficient for non-linear optimisation problems. The weighting sum method is the simplest multi-criteria decision making method which has been used extensively to provide Pareto optimal solutions by varying the weights of the involved objectives; however, the weights for each objective need to be determined properly [79]. As pavement maintenance problems

belong to non-linear optimisation problems, GAs are applied to perform optimisation and the normal constraint method is used to make a balance between the involved objectives.

#### 4.3.2.3 Maintenance and Rehabilitation Optimisation using Genetic Algorithms

Genetic Algorithms (GAs) can be applied to the M&R optimisation problem because of their robust search capabilities that overcome the combinatorial explosion of large-size optimisation problems and are based on the survival-of-the-fittest concept of Darwinian evolution. The major barrier of the optimisation problem is that the solution space grows exponentially with the size of the problem, so the conventional optimisation approach can be inefficient to find the optimal solution. While GAs, which incorporate a set of initial solutions and generate new and better solutions according to the probabilistic rules, can be more effective and the likelihood of achieving the optimal solution is increasing [80].

### 4.4 Genetic Algorithms

#### 4.4.1 Introduction

Genetic algorithms were proposed by John Holland at the University of Michigan and applied as a computational technique in 1975 [76]. GAs are developed based on the survival-of-the-fittest concept of Darwinian evolution. Goldberg [81] constructed the initial framework for GAs, known as the binary-coded genetic algorithm (BGA), and exhibited its robustness of optimisation and search. GAs imitate the natural process of biological evolution, can be more efficient and the probability of reaching the optimal solution is increasing. During the last two decades GAs have been widely studied and experimented, and a significant contribution was achieved within pavement management engineering [71, 80, 82-84]. GAs are attractive to pavement engineers



because of their robust search capabilities and ease of implementation. Furthermore, GAs are applicable to resolve multi-objective optimisation problems. During the last few years multi-objective genetic algorithms (MOGAs) were successfully applied to a large amount of pavement maintenance scheduling problems [85-87]. Recently, many different MOGAs have been proposed, among which the non-dominated sorting genetic algorithm II (NSGA-II) is recognised as one of the most advanced method for solving multi-objective problems [88, 89].

#### 4.4.2 Simple Genetic Algorithms

In biological terms, natural evolution takes place in chromosomes that build up the structure of creatures. As for GAs, they involve encoding the decisive parameters of the optimisation problem as a finite-length string, i.e. chromosome. An individual GA chromosome is used to represent a solution to the optimisation problem that is known as a genotype and the value of a gene contributing to the chromosome is called an allele.

##### 4.4.2.1 Coding

Holland [76] introduced binary coding approach for mapping a chromosome to the parameters of an optimisation problem, in which the value of a gene can only be either 0 or 1. A binary string as an unsigned integer where the string  $A = a_l a_{l-1} \dots a_2 a_1$  is decoded to the parameter value  $x$ :

$$x = \sum_{i=1}^l a_i \cdot 2^{i-1} \quad (4-1)$$

For example, for a string  $A = 10010$ , it is decoded to:

$$x = 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 18$$

If upper and lower bounds are applied, denoted by  $u_0$  and  $l_0$ , this binary string is decoded as:

$$x = l_0 + \frac{\sum_{i=1}^l a_i \cdot 2^{i-1}}{2^l - 1} \cdot (u_0 - l_0) \quad (4-2)$$

When  $u_0$  and  $l_0$  are defined as 0 and 10, respectively, the string value is decoded to:

$$x = 0 + \frac{18}{2^5 - 1} \times (10 - 0) = 5.14$$

Apart from binary coding, real coding and integer coding are also used to map the parameter space of an optimisation problem [90, 91].

#### 4.4.2.2 General Operation

GAs begin by the generation of an initial pool of genotypes to represent a set of feasible solutions, and then each individual within the initial pool is evaluated using the objective function and ranked in terms of its fitness that obtained as its value of the objective function. With the aid of genetic operators, i.e. reproduction, crossover and mutation, each genotype is allowed to create a certain number of offspring depending on its fitness. With respect to the values of the objective function, relatively better solutions would be retained, while the rest would be deleted. As a consequence, a new parent pool is formed by selecting the desired number of offspring. The genetic selection process, solution-pool selection process, and offspring generation are repeated until the result reaches convergence or a maximum iteration set by the user is met. The general framework for GAs is described in Figure 4-2.

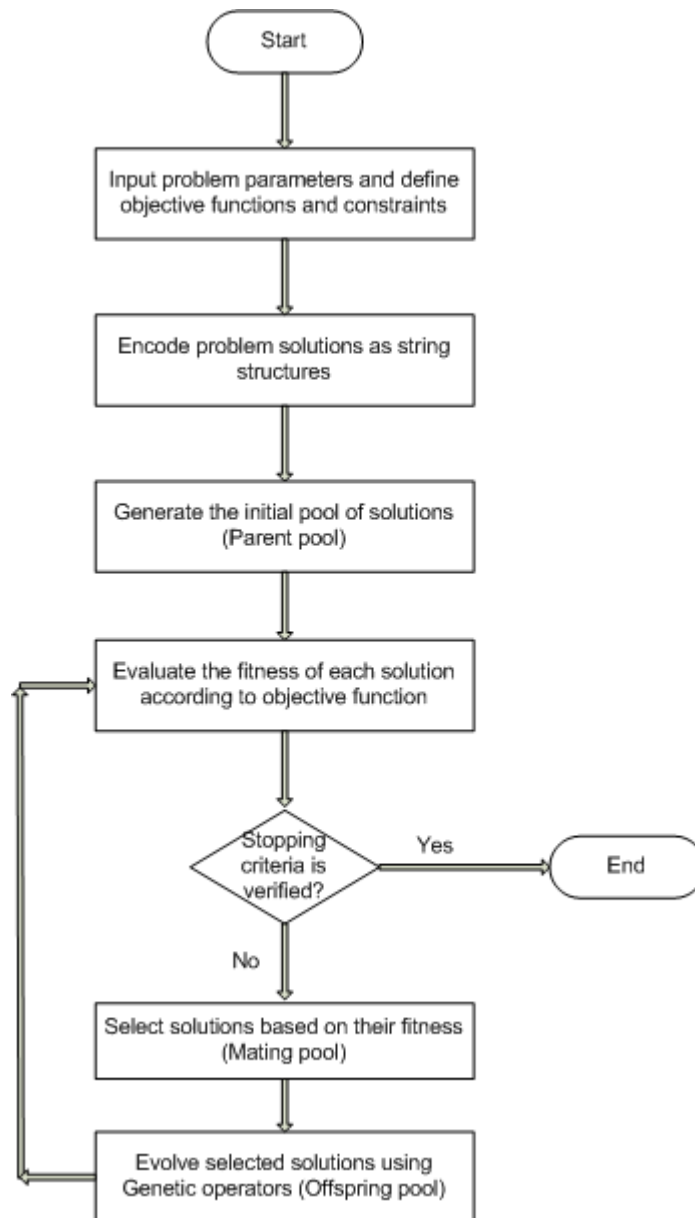


Figure 4-2: General framework for GAs

#### 4.4.2.3 Genetic Operators

After the generation of the initial solutions, parent pools are selected in accordance with their fitness using a selection process. The following step is to generate offspring pools by mating of the parent pools. There are three operators applied, reproduction, crossover and mutation.

##### 4.4.2.3.1 Reproduction

Once the parent pool is created reproduction is performed. Reproduction is a procedure in which individuals in the population pool are copied in terms of their fitness values. The fitness value for each chromosome is obtained by transforming its objective function value into a measure of relative fitness. Afterwards, selection methods are used to determine how chromosomes are chosen for reproduction. One of the most popular and simple method is roulette wheel selection [81]. Each chromosome in the pool is assigned a roulette wheel slot sized in proportion to its fitness value. The allocated space on the roulette wheel for each individual is derived in terms of the selection function, described as:

$$p_i = \frac{f_i}{\sum_{i=1}^N f_i}, i = 1, 2, \dots, N \quad (4-3)$$

Where

$f_i$ - the fitness value of chromosome  $i$

$N$ - the number of chromosomes in the parent pool

When the generation of offspring is required, a simple spin of the roulette wheel produces the reproduction participants. Therefore, chromosomes with a higher fitness value are more likely to be succeeded in the next generation. The selected chromosomes are then placed into the mating pool for further genetic operator processes.

For instance, there are five chromosomes in the parent pool, 00011, 00110, 01100, 11000, and 10001. The objective function values of them are calculated using binary coding method, described in Table 4-1:

Table 4-1: Sample population of chromosomes and fitness values

ID	Chromosome	Fitness value	% of Total
1	00011	3	4.8
2	00110	6	9.7
3	01100	12	19.4
4	11000	24	38.7
5	10001	17	27.4
Total		62	100

The fitness value of chromosome 1 (00011) is calculated as 3, which occupies 4.8% of the total fitness. Therefore, chromosome 1 is allocated 4.8% of the biased roulette wheel space, and its probability for reproduction is 4.8%.

#### 4.4.2.3.2 Crossover

There are two main steps involved in crossover. On one hand, the newly reproduced chromosomes in the mating pool are mated at random. On the other hand, each pair of chromosomes undergoes crossover by selecting a crossing point in the chromosome at random and then combining the genes from one mate before the crossing point and the genes from the other mate after the crossing point. For example, consider two binary parent chromosomes:

$$P_1 = 001|0100,$$

$$P_2 = 010|1101.$$

Assuming the randomly selected crossover point is 3 from the left end of the chromosomes and the resulting two offspring are:

$$O_1 = 001|1101,$$

$$O_2 = 010|0100.$$

This method described above is single point crossover, which is the simplest form of genetic operator. However, it becomes insufficient for the long chromosomes. Thus, multi-point crossover and uniform crossover are proposed to generate more string combinations so as to overcome this deficiency [92, 93].

#### 4.4.2.3.3 Mutation

Mutation takes place after crossover, it allows for the genes on the chromosomes in the mating pool to be altered. It is processed by probabilistically selecting chromosomes from the mating pool in accordance with their fitness values, and the mutation point at the selected chromosome is picked randomly and the gene at that point is altered at random. The mutated chromosome is then entered into the parent pool in the next generation. For instance, mutating the 3<sup>rd</sup> bit in a binary string  $P_3 = 010\underline{1}100$  results in the offspring  $O_3 = 01\underline{1}1100$ .

The mutation operator when only one gene on the chromosome is altered at a time is termed uniform mutation. In addition, multiple uniform mutation is utilised in GAs, described as uniform mutation of  $n$  randomly selected genes on the selected chromosome. The value of  $n$  is defined at random from 1 to the length of the chromosome.

#### 4.4.2.4 Genetic Algorithms Parameters

There are some program variables that affect the performance of the GAs need to be defined, including:

- The population size,
- The maximum number of generations,
- The crossover rate,
- The mutation rate.

With the increase of population size and maximum number of generations, the search space is more thorough explored and more feasible solutions might be generated. However, this increased the computational complexity significantly and a trade-off between computational effort and space exploration must be made.

As for the crossover rate, given a higher value more of the solution space can be investigated and the chance of achieving a local optimal solution is reduced. Nevertheless, when the crossover rate is too high more computation is required.

For mutation rate, if it is too low, some useful genes may never be tried out; on the contrary, if it is too high, it leads to the fact that the offspring will start losing resemblance to their parents. It is suggested that mutation rate should be set within the interval [0.001, 0.01] [94].

#### 4.4.3 Multi-objective Genetic Algorithms

In pavement engineering, there is often more than one objective that needs to be taken into account, and those multiple objectives, often conflicting, have significant different impacts on the resulting M&R strategy, such as maximisation of pavement condition and minimisation of maintenance cost.

The aim of multi-objective optimisation model is to achieve a vector of decision variables which optimizes the involved objective functions and satisfies the constraints. Typically, the approach to solve this model is to develop Pareto optimal solutions in terms of the involved objectives, and then to identify the corresponding Pareto frontier. Finally, the acceptable solution is selected according to the decision maker's criteria. As MOGAs are easy to implement, robust, and they need less information about the problem being resolved, they are particularly suitable to achieve the Pareto optimal set than other multi-objective optimisation techniques [95].

#### 4.4.3.1 Problem Formulation and Pareto Optimality

A general multi-objective optimisation problem with  $m$  decision variables and  $n$  objectives is expressed as:

$$\text{Min } F(X) = (f_1(X), f_2(X), \dots, f_n(X)) \quad (4-4)$$

Where  $X = (x_1, x_2, \dots, x_m)$  is an  $n$ -dimensional variable vector, and  $f_1(X), f_2(X), \dots, f_n(X)$  are the  $n$  objective functions. The solution space  $X$  is generally restricted by a set of constraints, including both equality and inequality constraints, stated as:

$$h_i(X) = 0, i = 1, 2, \dots, I; \quad (4-5)$$

$$g_j(X) = 0, j = 1, 2, \dots, J. \quad (4-6)$$

In Pareto based approaches, a solution vector  $a \in X$  is said to dominate a solution vector  $b \in X$  if  $a$  is better or equal to  $b$  in all attributes, expressed as:

$$\forall i \in \{1, 2, \dots, n\} : f_i(a) \leq f_i(b) \text{ and } \exists j \in \{1, 2, \dots, n\} : f_j(a) < f_j(b) \quad (4-7)$$

A solution vector  $c \in X$  is said to be Pareto optimal if and only if it is non-dominated regarding the whole solution space  $X$ . If a solution is not within the Pareto optimal set, it can be improved in terms of any objective without leading to deterioration in other objectives. The combination of all feasible non-dominated solutions within solution space  $X$  is defined as the Pareto optimal set and the objective function values of Pareto optimal solutions in the objective space is recognised as the Pareto optimal frontier. For instance, the Pareto optimal frontier for a typical two-objective minimisation problem is illustrated in Figure 4-3:



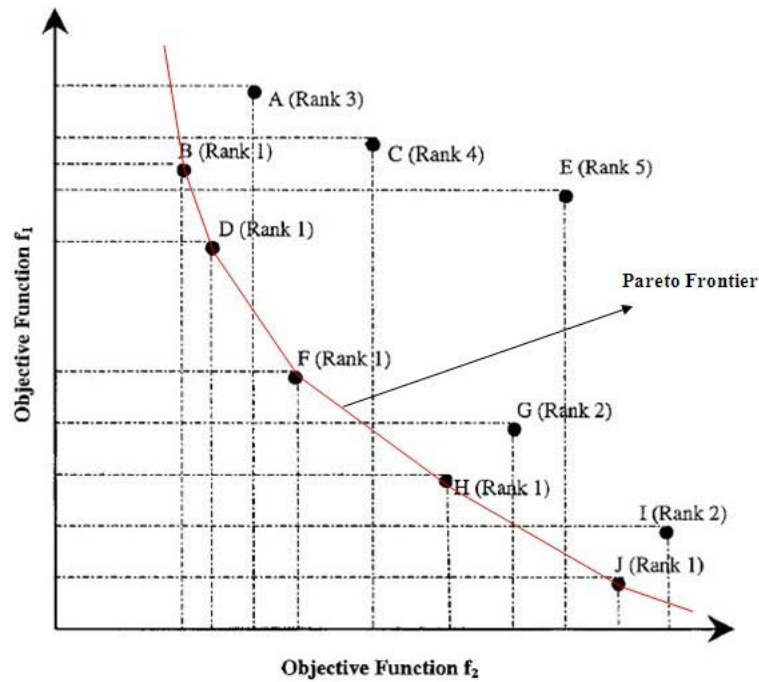


Figure 4-3: Pareto optimal frontier in two objective spaces

Figure 4-3 is removed from [86]. The solutions  $B$ ,  $D$ ,  $F$ ,  $H$  and  $J$  are recognized as non-dominated solutions, while other solutions are dominated by them. Eventually, the global optimal solution is defined as the point in the Pareto optimal frontier that has the minimum Euclidian distance to the origin point along the Pareto Frontier.

#### 4.4.3.2 General Framework for MOGAs

Schaffer [96] proposed the first MOGA, called vector evaluated genetic algorithm (VEGA). VEGA applies the selection operation of the GAs to generate non-dominated solutions. Individual objectives are deployed as the selection metrics. It is found that this method cannot cover the whole Pareto frontier. Later, Goldberg [81] introduced the non-dominated sorting procedure to MOGAs as the rank-based approach, three main steps are involved:

- All the non-dominated individuals in the population pool are identified. They are assigned the best rank, i.e. rank 1, that considered as the best solutions, and then are eliminated from the population pool.
- The non-dominated individuals in the residual population pool are captured and assigned rank 2, they are also deleted from the population pool.
- This process of evaluating non-dominated individuals in the pool continues until all the individuals in the pool have been ranked.

Figure 4-4 shows the Pareto ranking procedure for the small population with two minimisation objectives, which is removed from [97].

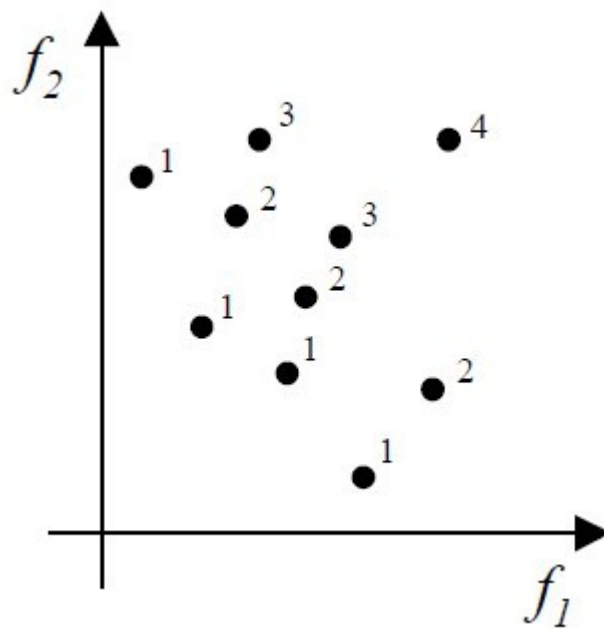


Figure 4-4: Population ranking based on non-dominated sorting procedure

Since then many different MOGAs based on various Pareto ranking approaches have been provided for solving engineering problems [89]. The general framework for these MOGAs is described in Figure 4-5.

MOGAs start from the generation of an initial pool of chromosomes, and the objective values for each chromosome are calculated like in the single objective GAs. Then Goldberg's non-dominated sorting procedure is applied to find all the non-dominated

individuals that have the same reproductive potential, in which the fitness value of each individual is calculated according to the non-dominated criterion instead of the objective values themselves. If convergence is not verified, the individuals in the parent pool are processed by genetic operators, i.e. reproduction, crossover and mutation, to generate offspring solutions, and the algorithm returns to the fitness evaluation step for the next generation. This process continues to produce new solutions that would generate a superior frontier that dominates the existing frontier that achieved in the previous iterations until a set of globally non-dominated solutions is found or the maximum iteration is exhausted.

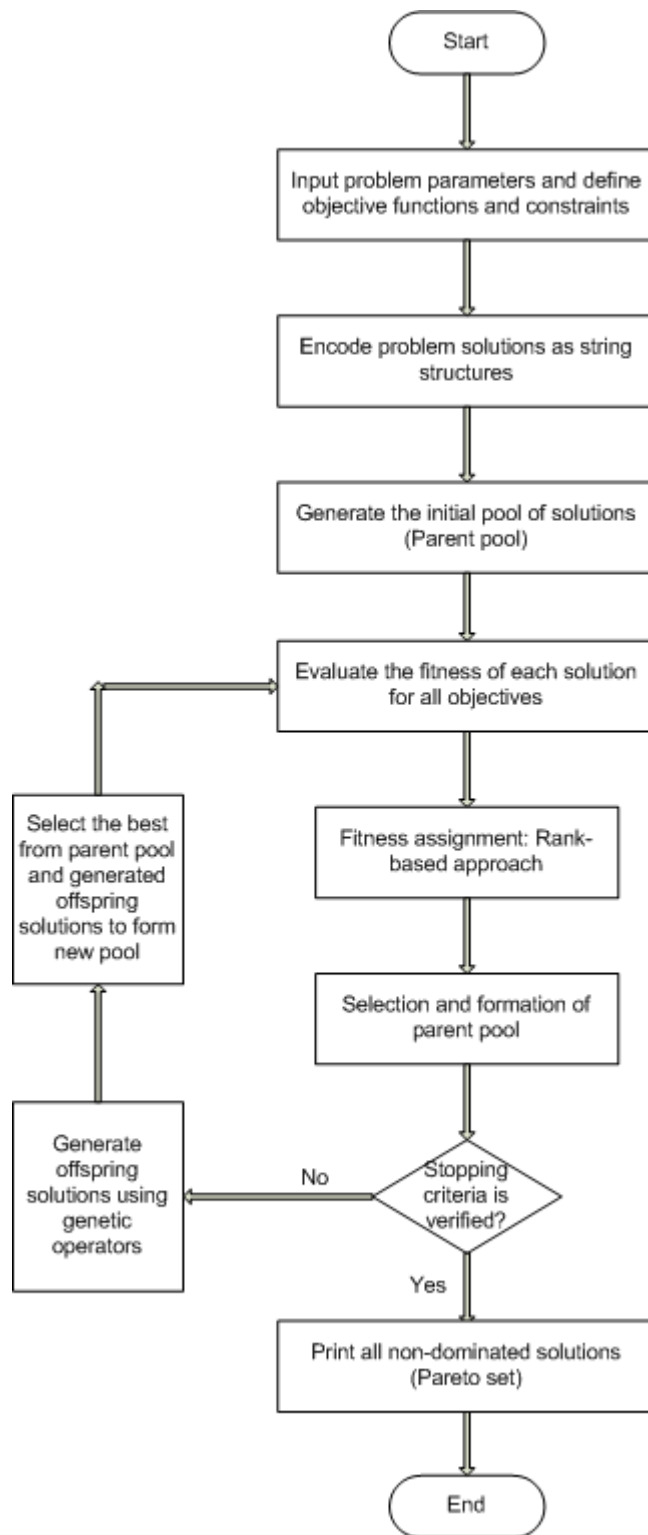


Figure 4-5: General framework for MOGAs

### 4.4.3.3 The Non-dominated Sorting Genetic Algorithm

#### 4.4.3.3.1 NSGA Overview

The non-dominated sorting genetic algorithm (NSGA) is developed based on Goldberg's non-dominated sorting procedure and sharing approach [98]. The only difference between NSGA and simple GAs is the selection operation, while the procedure of crossover and mutation are the same. In contrast to the general non-dominated sorting procedure described in Section 4.4.3.2, the identified non-dominated individuals in the population are given a large dummy fitness value, and then are removed from the population. The same dummy fitness value is given to the non-dominated individuals with an equal reproductive potential. Then sharing method [99] is used to process the individuals with the same ranking, realised by implementing selection process using reduced fitness values which are calculated by dividing the original dummy fitness value of an individual by a quantity proportional to the amount of points around it. This results in a set of optimal solutions to present in the pool. Next, the non-dominated individuals in the reduced population are identified and assigned lower than the minimum value of the previous ranked individuals, and then sharing is applied to this second set of non-dominated individuals. This process is repeated until the whole population is ranked.

When assigning fitness value and sharing is complete, the population is reproduced in terms of the dummy fitness value, and genetic operators are used to process the individuals in the population. The individuals that have the maximum dummy fitness value are located in the first front, and they have a higher probability of generating offspring. This process is intended to explore the Pareto optimal frontier, which leads to quick convergence of the population towards the Pareto optimal frontier and sharing method aids in distributing it over this frontier.

Figure 4-6 shows the general framework for NSGA, which is removed from [98]. As stated above, NSGA varies from simple GAs only in the classification of non-dominated frontiers and the sharing procedure. The sharing for each frontier is evaluated by calculating the share function value between two individuals that belong to the same frontier, formulated as [98]:

$$Sh(d_{ij}) = \begin{cases} 1 - (\frac{d_{ij}}{\sigma_{share}})^2, & \text{if } d_{ij} < \sigma_{share} \\ 0, & \text{otherwise} \end{cases} \quad (4-8)$$

where

$d_{ij}$ - the phenotypic distance between two individuals  $i$  and  $j$  in the same Pareto frontier

$\sigma_{share}$ - the maximum phenotypic distance defined between any two individuals in the same Pareto frontier to become members of a niche

Subsequently, another parameter niche count is provided by summing up the sharing function values for all the individuals in the current frontier from an individual. Eventually, the shared fitness value for each individual in the current frontier is obtained by dividing its dummy fitness value by the corresponding niche count.

The Goldberg's non-dominated sorting procedure is taken into account by NSGA in a better way. The ranking method is implemented in terms of the non-dominated individuals in the population and then the distribution of the non-dominated individuals is filtered using a niche formation method, resulting in that the distinct non-dominated individuals to be identified in the population.

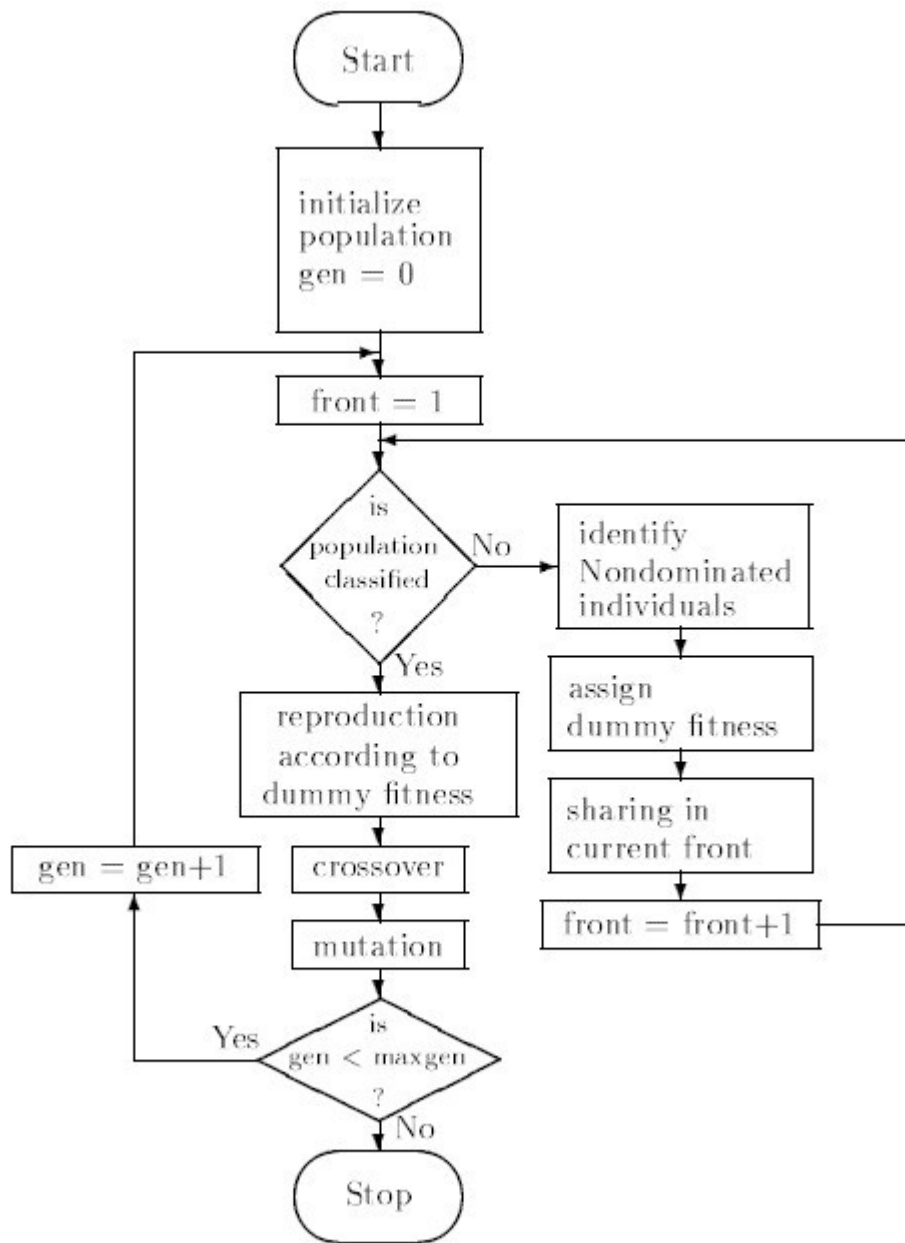


Figure 4-6: General framework for NSGA

#### 4.4.3.3.2 NSGA-II: the improved NSGA

Over the past years, the main weaknesses of NSGA are depicted as follows [88]:

- High computational complexity of the selection operation of non-dominated solutions. The computational complexity is  $O(MN^3)$  (where  $M$  is the number of objective functions and  $N$  is the size of the population), which is resulted from the non-dominated sorting procedure.
- Lack of elitism. It is reported that elitism can improve the performance of GAs substantially and can help maintaining good individuals once they are identified [100].
- Requirement for defining the sharing parameter  $\sigma_{share}$ . The performance of the sharing technique in achieving a set of solutions depends significantly upon the value of  $\sigma_{share}$ . In addition, the sharing technique leads to more computational complexity.

Deb et al. [88] developed the improved version of NSGA that named NSGA-II, which alleviates the difficulties stated above.

#### 4.4.3.3.3 Elitist Non-dominated Genetic Algorithm

A fast non-dominated sorting approach is proposed in NSGA-II, which requires  $O(MN^2)$  computational complexity. In this approach, two parameters are calculated for each individual in the population: (a) domination count  $n_p$ , the amount of individuals which dominate the individual  $p$ , and (b)  $S_p$ , a series of individuals that the individual  $p$  dominates. All individuals with a domination count of zero are located in the first non-dominated frontier. Next for each non-dominated individual  $p$  with  $n_p=0$ , the members of its domination set  $S_p$  are processed by reducing their domination count by one. Any member with a current domination count of zero is placed in the second non-dominated frontier. This process is repeated until the whole population is evaluated, which requires



$O(MN^2)$  computations that is much less than that of the non-dominated sorting procedure applied in NSGA. Also, the sharing function method presented in NSGA is replaced by a crowded-comparison method. This method has no need for specifying the sharing function value, and gives rise to a better computational complexity. In addition, a selection operator is proposed that generates a mating pool by mixing the parent and offspring populations and identifying the best  $N$  individuals in terms of their fitness and spread.

#### 4.4.3.3.4 Constraint Handling

NSGA-II can also be applied to solve constrained multi-objective problems using a constraint handling approach. This constraint approach deploys the binary tournament selection, where two individuals are selected from the population and the better one is chosen. In the presence of constraints, each individual in the population can be either feasible or infeasible. As a consequence, three cases are taken into account: (a) both individuals are feasible; (b) one is feasible and the other is not; (c) both individuals are infeasible. With the consideration of infeasible individuals, the definition of dominance between two individuals in the population is modified as:

*Definition 1:* An individual  $a$  is said to constrained-dominate an individual  $b$ , if one of the following situations is fulfilled.

1. Individual  $a$  is feasible and individual  $b$  is infeasible.
2. Both are infeasible, but individual  $a$  has a smaller overall constraint violation.
3. Both are feasible and individual  $a$  dominates individual  $b$ .

It is worth noting that the modification in the definition of dominance has no effect on the computational complexity of NSGA-II.

## 4.5 Summary

This Chapter introduces the maintenance actions that widely applied in the UK, which can be classified into three main groups according to their frequency and their impact on the serviceability of road, i.e. minor maintenance, surface treatment and major maintenance.

This chapter provides two major approaches employed in the decision making of M&R strategy: the prioritisation model and the optimisation model. Both of them can be applied to the PMS at the network and project levels. The prioritization model is regarded as the “bottom-up” approach because it is applied to individual segments in the pavement network firstly and then prioritized according to the needs and resources of the highways agency. The advantage of such a model is that it is easy to use and it can be implemented quickly. However, it cannot yield an optimal M&R strategy because of no consideration of the influence of the whole M&R strategy on the pavement network. Conversely, the optimisation model utilises the “top-down” methodology, which is applied at the network level pavement system at first, and then extended to the project level to make specific decisions for individual segments. It aims to find the optimal M&R strategy for the pavement system according to its defined objective. Also, it can solely be applied to pavement system at the project level, such as the segment-linked optimisation model. Both deterministic and probabilistic pavement performance prediction models can be incorporated into the optimisation model. The deterministic models apply the regression equations to describe the deterioration of the pavement condition, while the probabilistic models utilise the Markov models for the same purpose.

Traditionally, the optimisation models deploy only one objective while imposing other competing objectives as constraints. Such models suffer two limitations [74]: how to

define the objective that deserves the most attention and determine the proper range values for the other objectives that formulated as constraints. As a result, the solution obtained is usually suboptimal. Therefore, multi-objective optimisation model can be used to overcome this problem. Since there is no single global optimal solution for the multi-objective optimisation model, it is achieved by generating the set of Pareto optimal solutions. Then, the global optimal solution is selected from the Pareto solution set, which is a vector of decision variables which optimizes the defined objectives and satisfies the constraints involved.

The methods that are usually employed as the solution of optimisation model include linear programming, integer programming and dynamic programming. However, the solution space obtained in the optimisation model increase exponentially with the increase of the size of the pavement system, especially for the segment-linked and multi-objective optimisation models, which are far more intensive and complicated. Therefore, GAs and some other soft computing applications can be deployed as the solution for such models, because they can overcome the combinatorial explosion of large-sized optimisation problem. In the presence of conflicting objectives, multi-objective genetic algorithms can be applied owing to their robust search capabilities and ease of implementation. With the adoption of an advanced non-dominated sorting approach and a simple constraint-handling method, NSGA-II is selected for pavement maintenance optimisation in this thesis.

## 5 Scheduling of Pavement Maintenance

### 5.1 Introduction

In this chapter we investigate pavement maintenance problem at both the network level and the project level. The proposed models are used to determine the optimal M&R strategy with the aim of minimising total cost and maximising pavement condition, taking into account the budgetary and applicable technical constraints.

For the network level optimisation model two main components are involved: (1) the pavement deterioration model used to predict pavement performance; and (2) the technique used to determine the M&R strategy. The pavement deterioration model employed in this thesis is a deterministic pavement age gain model which requires less historical data to calibrate than other pavement deterioration models. In order to capture the uncertainty inherent in the pavement deterioration process, the deterministic model can be transformed to a probabilistic one by incorporating normal distributions. The approach used to achieve the optimal M&R strategy is an optimisation model based on GAs.

At the project level, in addition to the maintenance cost to the highways agency, pavement maintenance works and constructions on the highway network also disrupt traffic flow and increase safety hazards to vehicle drivers and road workers, which further resulted in additional cost to road users. It is expected that the delay cost to road users may exceed the corresponding maintenance cost. As a consequence, it is greatly desirable to optimise the worksite arrangements so as to cause less disruption and delay to the road users.

Subsequently, NTFM is applied to model the delays and disruption caused by maintenance activities, the resulting road user costs together with maintenance costs can

be used to optimise maintenance actions for road sections on the network in the long-term. Afterwards, the maintenance arrangements for road sections can be optimised so as to minimise road user cost and to make sure that the corresponding maintenance cost is acceptable to highways agencies, where the single-objective GA is employed to perform optimisation.

## **5.2 Scheduling of Pavement Maintenance using NSGA-II at the Network Level**

In the presence of aging and deteriorating highway networks and inadequate budgets, it is of substantial importance for highways agency to operate, preserve and enhance highway networks in a cost-effective way. This thesis presents a road section based optimisation model to determine the optimal maintenance and rehabilitation (M&R) strategy for a highway network in the planning period using GAs. The model is aimed at minimising the total agency cost and maximising the remaining pavement life of a highway network over a given planning period. Road sections between junctions are the decision making units to which maintenance actions apply. This model employs age gain to indicate the improvement of road sections resulting from M&R actions. In addition, integer encoding is deployed to define various maintenance action possibilities. This coding method reduces the length of the chromosome compared to the more commonly used binary encoding of variables.

### **5.2.1 Age Gain Maximisation Model**

The developed model applies the expected age gain associated with the M&R actions as the pavement improvement indicator [85]. Five maintenance actions are considered, do nothing, patching, surface dressing, resurfacing and overlay, which are represented by 0, 1, 2, 3 and 4 using the integer coding method. The maintenance cost and duration associated with each maintenance action is shown in Table 5-1 [101]:

Table 5-1: Maintenance duration and cost for each type of maintenance action

M&R	Single carriageway (S2)		Dual 2 lane carriageway (D2AP)		Dual 3 lane motorway (D3M)	
	Duration (days)	Cost (£'000s)	Duration (days)	Cost (£'000s)	Duration (days)	Cost (£'000s)
Do nothing	0	0	0	0	0	0
Patching	2	50	3	100	4	140
Surface dressing	4	70	5	140	6	170
Resurfacing	8	200	14	550	20	900
Overlay	16	320	23	820	33	1350

1. Costs and days are for 1km of road, that is, both carriageways.
2. Costs are £'000s expressed in average 2012 price, which are derived in terms of inflation rate and DfT [101].

The age gain for each M&R action is characterised by the remaining pavement life of the road section, given in Table 5-2 [101]:

Table 5-2: Expected ages associated with pavement maintenance actions versus remaining pavement life

Remaining pavement life (years)	Pavement condition	Expected age gain (years)				
		Do nothing	Patching	SD	Re	Ov
		(0)	(1)	(2)	(3)	(4)
>6	0	0	2	5	7	10
(4,6]	1	0	2	3	7	10
(2,4]	2	0	1	3	6	9
(0,2]	3	0	0	1	6	9

1. (0), (1)... (4) represent the ID of maintenance actions.

## 5.2.2 Classification of Road Sections

When each road section is deployed as the decision making unit, it becomes computationally inefficient for a highway network that is composed of a large amount of sections, which is the case of most highway networks. To deal with this problem, the proposed formulation of the maintenance optimisation problem categorizes road sections on a highway network into groups with similar properties with regard to some characterising variables, such as road classes, traffic categories, and climatic regions, which govern the pavement performance. By this means, all individuals from the same group are assumed to have the same performance characteristics and should be studied in a similar manner. In this thesis, climatic condition is not considered, as the road sections on a relatively small-sized highway network belong to the same climatic region. The classification of road sections is illustrated in Table 5-3 [101]:

Table 5-3: Road section types

Road section group	Road class	Traffic flow (000s)	Initial Pavement life (years)
1	S2	5-15	10
2	S2	15-20	8
3	D2AP	10-20	10
4	D2AP	20-30	9
5	D2AP	30-40	8
6	D3M	20-30	11
7	D3M	30-40	10
8	D3M	40-80	9

1. Flows are opening year Annual Average Daily Traffic (AADT).
2. The initial pavement life for each road class is assumed, which can be calibrated according to historical data.

Considering both road class and traffic category, at least 8 types of road sections are investigated, including 2 types for S2, 3 types for D2AP and 3 types for D3M. Moreover, the initial pavement life for each type of new road section is also provided. In terms of the classification of road sections, the road sections on a constructed highway network are divided into 8 groups. Also the individuals from the same group have the same initial pavement life. As a result, maintenance actions are assigned to each road section group at each year during the planning horizon.

### 5.2.3 Model Formulation

The two objective functions of this model are formulated as follows:

$$\text{Min} \sum_{n=1}^N \sum_{t=1}^T \frac{1}{(1+d)^t} C_{n,t} \quad n = 1, \dots, N; t = 1, \dots, T \quad (5-1)$$

$$\text{Max} \sum_{n=1}^N rpl_{n,T} \times L_n \quad n = 1, \dots, N; t = 1, \dots, T \quad (5-2)$$

Where

$C_{n,t}$  - maintenance cost for road section group  $n$  at year  $t$

$rpl_{n,T}$  - remaining pavement life for road section group  $n$  at year  $T$

$L_n$  - the total length of road sections that belong to group  $n$

$N$  - number of road section groups

$T$  - planning horizon

$d$  - discount rate, expressed as a fraction

The discount rate is used to transform costs and benefits arising in different years to their present value. For example, £100 to be received in  $n$  years from now is worth  $\frac{£100}{(1+d)^n}$  in today's money. It is worth noting that this expression ignores the effect



of inflation and assumes that £100 has the same real value at each year. A discount rate of 0.035 is adopted in COBA [102].

The evolution of the remaining pavement life for road section  $i$  at year  $t$  is evaluated as:

$$rpl_t(i) = rpl_{t-1}(i) + ag_{t,i}[m][k] - 1$$

$$\text{If } rpl_t(i) > pl(i), rpl_t(i) = pl(i) \quad (5-3)$$

Where

- $rpl_t(i)$  - remaining pavement life for road section  $i$  at year  $t$
- $m$ - pavement condition of road section  $i$  at year  $t$ , i.e. 0,1,2 and 3, evaluated based on the remaining pavement life that depicted in Table 5-2
- $k$ - maintenance action performed in road section  $i$  at year  $t$
- $ag_{t,i}[m][k]$ - age gain associated with the maintenance action performed in road section  $i$  at year  $t$  based on remaining pavement life
- $pl(i)$ - initial pavement life for road section  $i$

Equation 5-3 indicates that the remaining pavement life of road section after maintenance cannot exceed its initial pavement life. The constraints involved are stated as:

$$rpl_t(i) \geq rpl_{\min} \quad (5-4)$$

$$\sum_{n=1}^N \frac{1}{(1+d)^t} C_{n,t} < B_t \quad (5-5)$$

Where  $B_t$  is the budget for year  $T$ , £'000.

The first constraint is that the remaining pavement life for each road section group  $n$  is greater than the minimum remaining pavement life. The second constraint is the budget constraint, where the maintenance cost on the whole network at year  $t$  should be less than the available budget. In addition, the constraint of major M&R actions for the

whole network can be included; it means that the number of major maintenance actions applied to the network at each year cannot exceed its threshold.

#### 5.2.4 Case Study

The requirement is to maintain a newly constructed road network during a planning horizon  $T$ , of 20 years. For ease of explanation, an example network that consists of 4 road sections is analysed, which is illustrated in Figure 5-1:

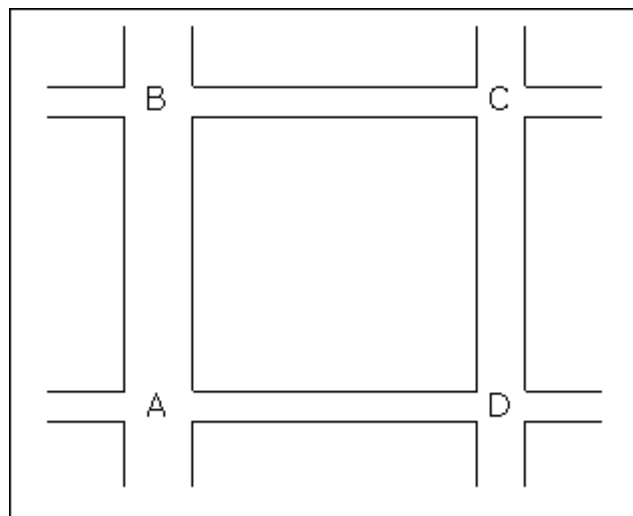


Figure 5-1: An example road network

The pavement condition of the road sections on the network is depicted in Table 5-4:

Table 5-4: Pavement condition for the example network

Road section	Road section group (Table 5-3)	Initial pavement life (years)	Length (km)
AB	6	11	1
BC	1	10	1
CD	3	10	1
AD	2	8	1

As there are 4 road sections on the network, 80 decision variables are incurred in this problem. For instance, the decision variable  $I_{nt}$  represents the maintenance action

implemented in road section  $n$  at year  $t$ . The decision string structure for road section  $n$  is represented by:

$I_{n1}$	$I_{n2}$	$I_{n3}$	$I_{n4}$	$I_{n5}$	$I_{n6}$	$I_{n7}$	$I_{n8}$	...	$I_{n15}$	$I_{n16}$	$I_{n17}$	$I_{n18}$	$I_{n19}$	$I_{n20}$
----------	----------	----------	----------	----------	----------	----------	----------	-----	-----------	-----------	-----------	-----------	-----------	-----------

Figure 5-2: The part of chromosome for road section  $n$

#### 5.2.4.1 Objectives and Constraints

As stated in Section 5.2.3, two objective functions are considered, which are formulated as:

$$\text{Min} \sum_{n=1}^4 \sum_{t=1}^{20} \frac{1}{(1+d)^t} C_{n,t} \quad (5-6)$$

$$\text{Max} \sum_{n=1}^4 \text{rpl}_{4,20} \times L_n \quad (5-7)$$

and the constraints are evaluated as:

$$\text{rpl}_t(i) \geq 3 \quad (5-8)$$

$$\sum_{n=1}^n \frac{1}{(1+d)^t} C_{n,t} < 2000 \quad (5-9)$$

The minimum remaining pavement life is defined as 3 years so as to ensure the serviceability of pavements, and the available annual budget is £2,000,000. The solutions that violate the constraints are defined as infeasible solutions. NSGA-II [88] deploys the constrained-domination principle to rank individual solutions. Firstly, each feasible solution is assigned a better non-domination rank than any other infeasible solutions. Secondly, all feasible solutions are ranked in terms of their non-domination level based on the objective functions values. As for infeasible solutions, the solution with a smaller constraint violation has a better rank.

#### 5.2.4.2 GA Parameters

In GAs, there are four main GA parameters: population size ( $P$ ), maximum simulation generations ( $M$ ), crossover rate ( $p_c$ ), and mutation rate ( $p_m$ ). As 80 variables need to be optimised, the population size is set to 500. Based on NSGA-II, an analysis was conducted to investigate the effect of changing GA parameter values. A limited set of values for each GA parameter was selected as follows:

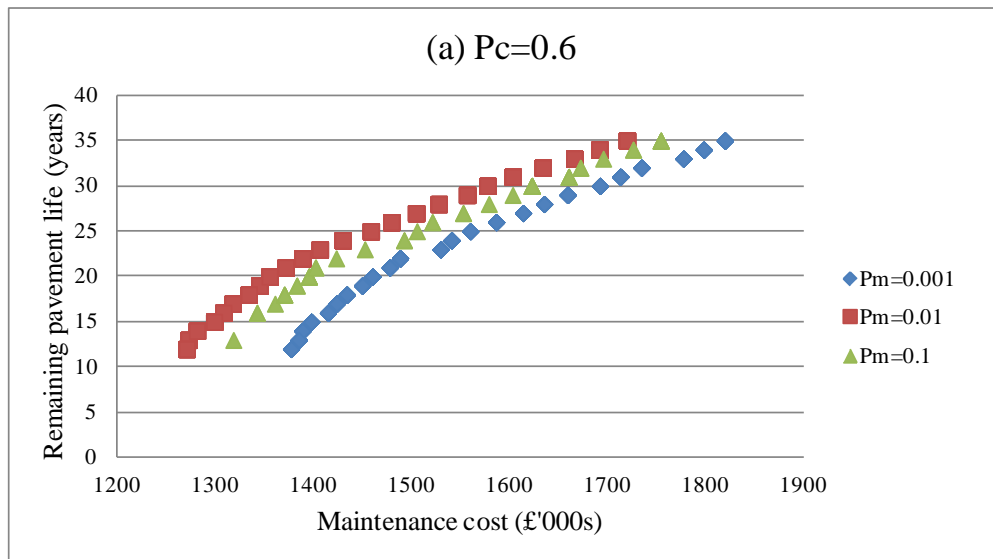
$p_c$ : 0.6, 0.7, 0.8, 0.9

$p_m$ : 0.001, 0.01, 0.1

$M$ : 5000

$P$ : 500

NSGA-II is used to determine the optimal Pareto frontiers using different combinations of crossover rates and mutation rates. Figure 5-3 is used to analyse the effect of mutation rates on the optimisation results when specific crossover rates are chosen.



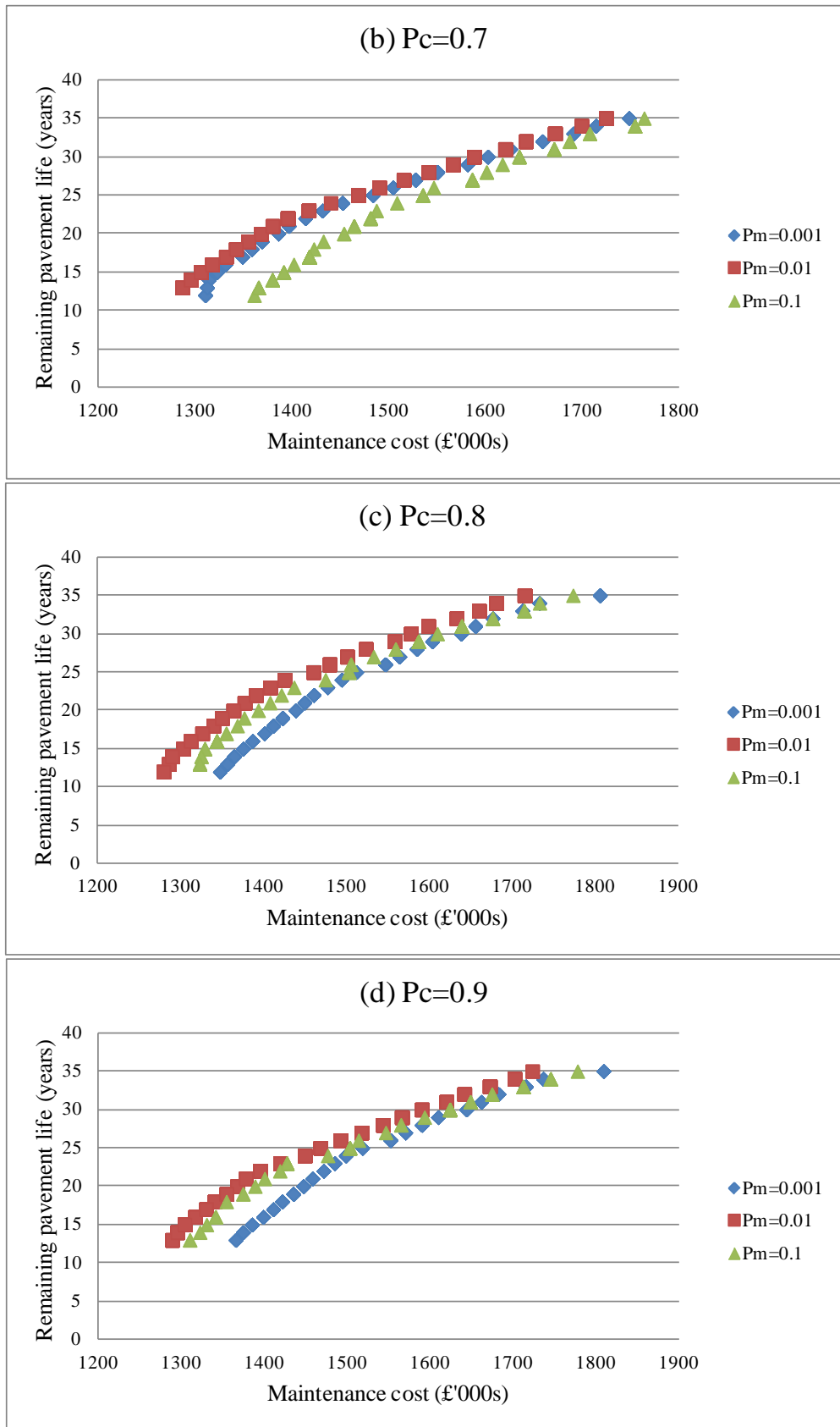


Figure 5-3: Optimal Pareto frontier for different mutation rates under each crossover rate

The results show that the mutation rate 0.01 lead to a better performance. This is because when mutation rate is too high, the offspring solutions cannot maintain some good genes of their parents. While for lower mutation rate, some helpful genes may never be explored. Figure 5-4 shows how crossover rates influence the optimisation process when mutation rate is 0.01.

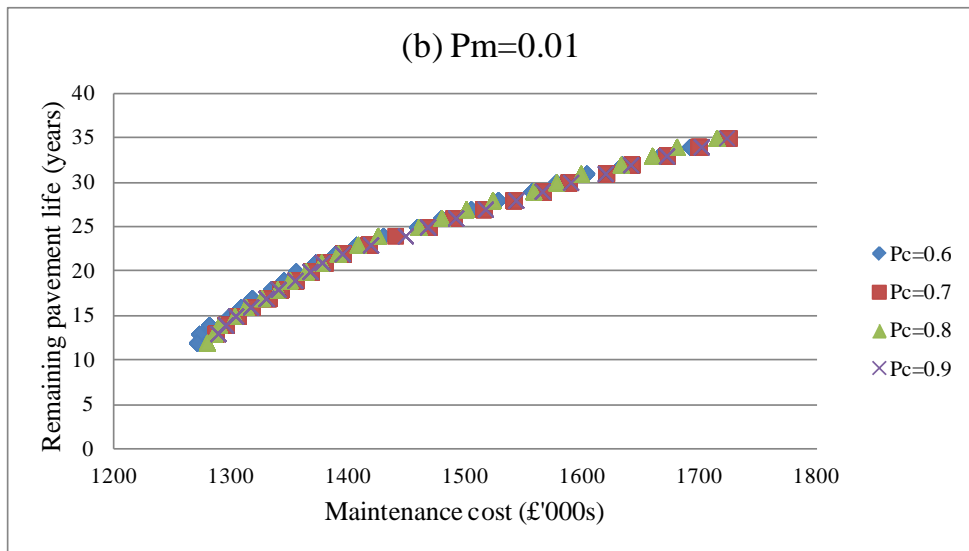


Figure 5-4: Optimal Pareto frontier for different crossover rates when  $P_m=0.01$

Almost the same optimal Pareto frontiers are obtained for each combination of crossover and mutation rates. In order to consume less computational effort, the optimal crossover rate is chosen as 0.6. Consequently, crossover rate, 0.6, and mutation rate, 0.01, are applied.

#### 5.2.4.3 Testing

To perform further testing the specified set of GA parameters values are used and 5 simulation runs are conducted. The maximum generation in a single run of the test is also set to 5000 generations which takes around 5 minutes to run on a PC.

The average fitness function value in each generation is deployed as the indicating factor when considering the convergence of the optimisation problem. For instance, the results obtained at each generation for runs 1 and 5 are presented in

Figure 5-5. The results indicate that the average population fitness function value converges towards the optimal solution. It is identified that the results are scattered in a relatively narrow range and the convergence to the optimal solution is obtained.

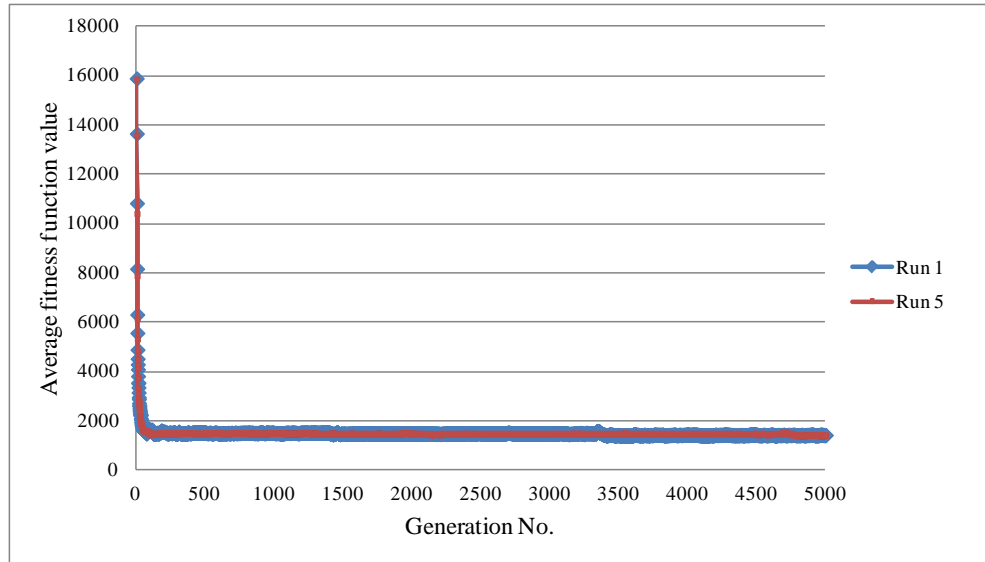


Figure 5-5: Average fitness function value at each generation

#### 5.2.4.4 Optimal Pareto Solution

Based on the specified GA parameter values, NSGA-II [88] is employed to achieve the optimal M&R strategy, and the resulting optimal Pareto frontier after different maximum generations is illustrated in Figure 5-6.

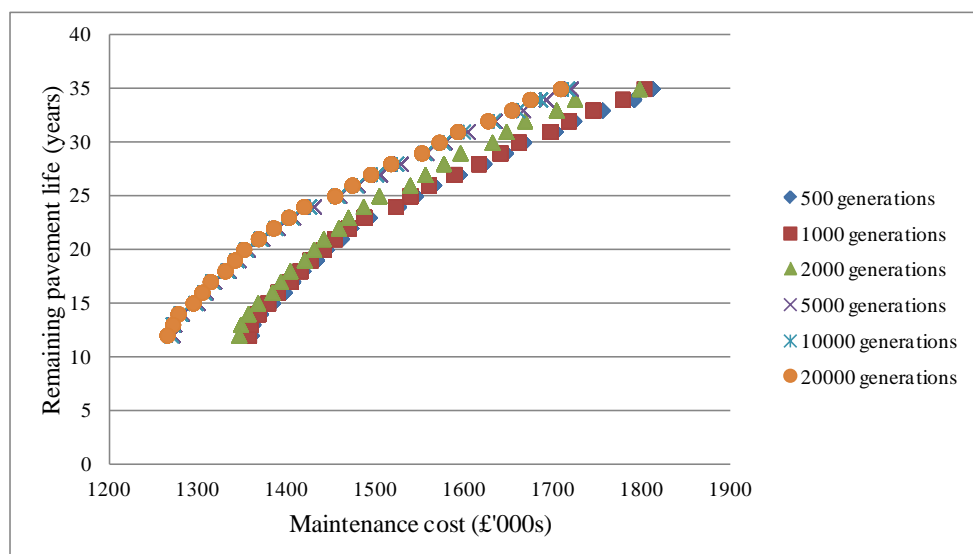


Figure 5-6: The optimal Pareto frontier

The optimal Pareto frontier obtained is approaching the real Pareto frontier with increasing simulation generations. The results demonstrate that the optimal Pareto frontier is achieved at generation 5000, which were not improved significantly with further iterations.

The next step is to normalise the obtained optimal Pareto frontier. The normalised distance for each non-dominated solution is calculated in terms of its objective function values. A normalised objective function value is achieved over a range of 0-100 for each objective based on the following mapping rules [86]:

$$O_N(i, x) = \frac{O(i, x) - O(i, \min)}{O(i, \max) - O(i, \min)} \times 100 \quad (5-10)$$

Where

$O_N(i, x)$  - the normalised parameter value of objective  $i$  for solution  $x$

$O(i, x)$  - the actual value of objective  $i$  for solution  $x$

$O(i, \max)$  - the maximum parameter value of objective  $i$  for non-dominated solutions

$O(i, \min)$  - the minimum parameter value of objective  $i$  for non-dominated solutions

For this optimisation problem,  $O(1, \min) = \text{£}1,264,076$  and  $O(1, \max) = \text{£}1,707,796$  for the objective of minimising maintenance cost;  $O(2, \min) = 12$  years and  $O(2, \max) = 35$  years for the objective of maximising remaining pavement life. The purpose of this mapping approach is to describe the two objective parameters on an identical space. Based on Equation 5-10, the normalised optimal Pareto frontier is illustrated in Figure 5-7:



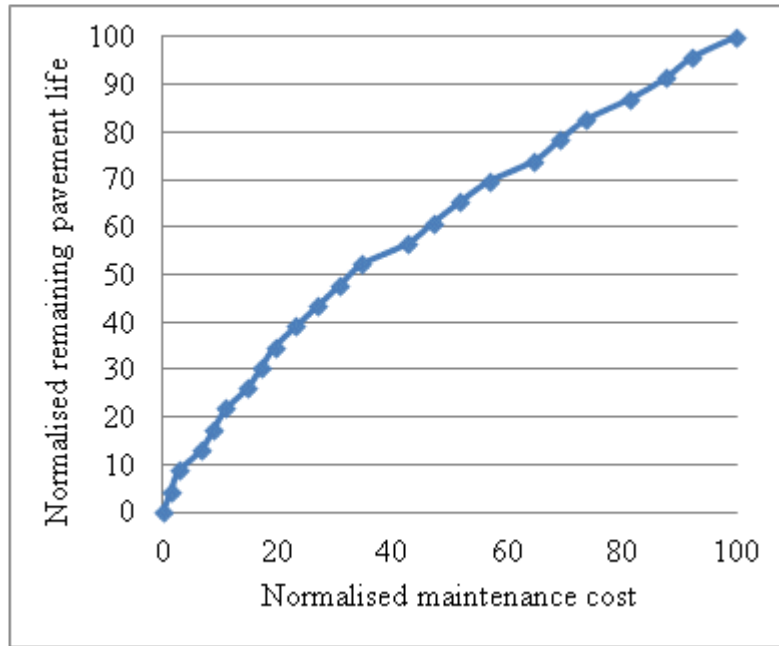


Figure 5-7: Normalised optimal Pareto frontier

The final step is to identify the solution that has the smallest Euclidean distance,  $d_k$ , from  $(O_N(1, \min), O_N(2, \max))$ , represented by:

$$d_k = \sqrt{[O_N(1, x) - O_N(1, \min)]^2 + [O_N(2, x) - O_N(2, \max)]^2} \quad (5-11)$$

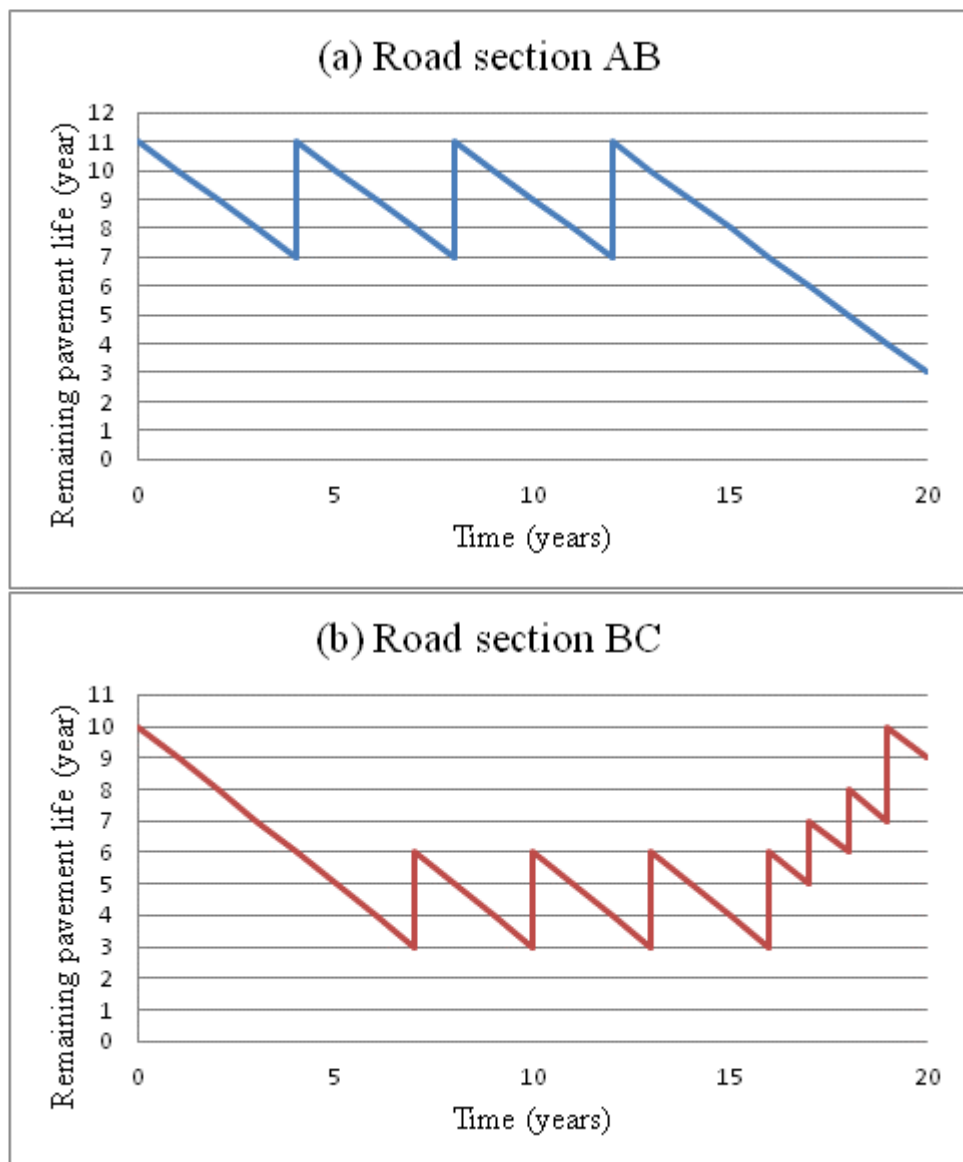
The optimal solution based on the shortest normalised distance is determined as maintenance cost of £1,418,291 and total remaining pavement life of 24 years, chosen from the optimal Pareto solution set within the space of £1,264,076-£1,707,796 for maintenance cost, and 12-35 years remaining pavement life for the network. The resulting optimal Pareto solution is described in Table 5-5:

Table 5-5: The optimal M&R strategy

Road section	Maintenance scheduling during the planning horizon T (20 years)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AB	0	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0
BC	0	0	0	0	0	0	0	2	0	0	2	0	0	2	0	0	2	1	1	2
CD	0	0	0	0	0	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2
AD	0	0	0	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2	1	1

Examining Table 5-5, the maintenance plan for each road section on the network is provided. For road section *AB*, three maintenance actions are implemented on it over 20 years; three SD actions at years 5, 9 and 13.

In terms of the M&R strategy provided in Table 5-5, the evolution of remaining pavement life for each road section during the planning period is illustrated in Figure 5-8:



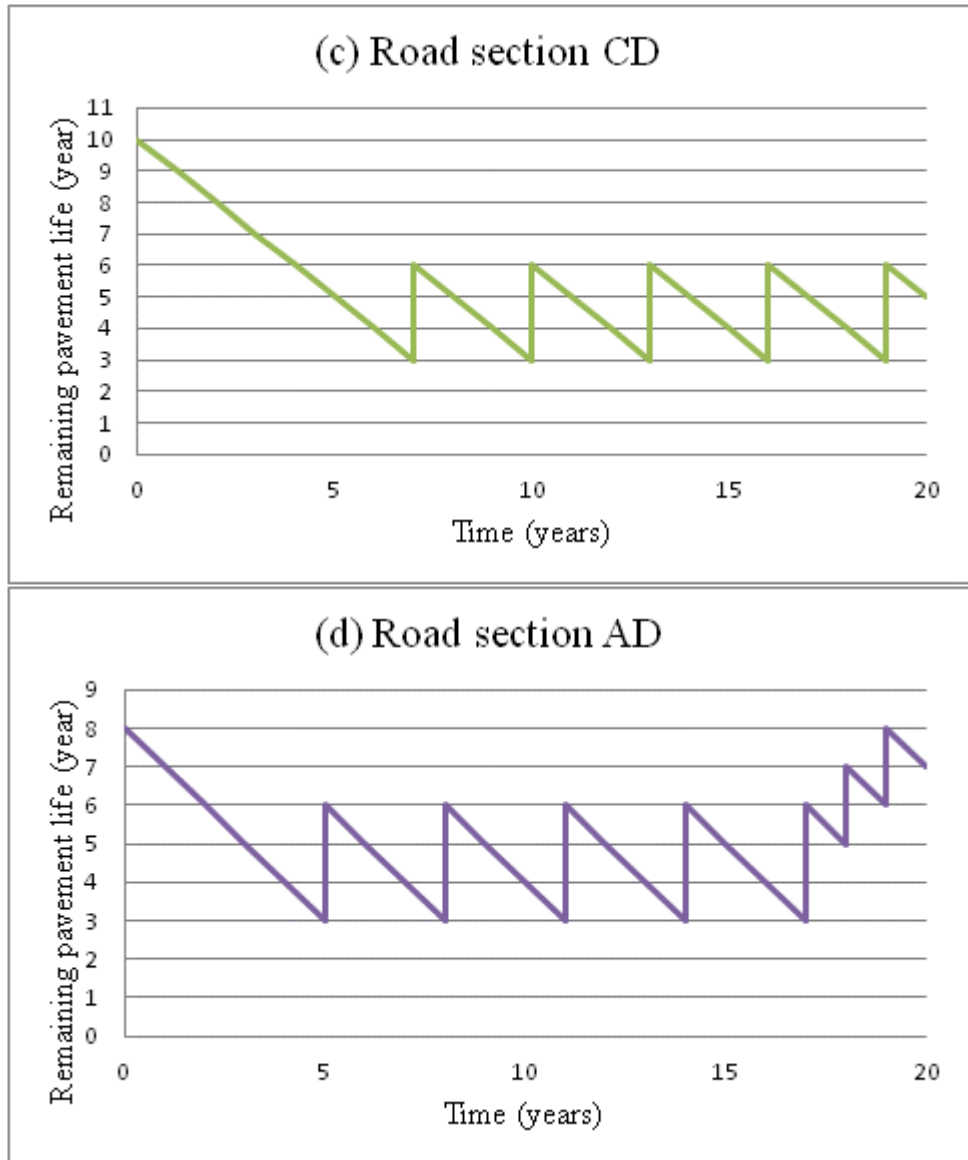


Figure 5-8: Pavement life cycle: (a) Section AB, (b) Section BC, (c) Section CD, (d) Section AD

For road section *AD*, one maintenance action, i.e. 2, is scheduled at year 6. The remaining pavement life at the end of year 5 is 3 years, and then SD is applied to prolong the pavement life by 3 years according to Table 5-2, which led to the remaining pavement life of 6 years at the beginning of year 6. At the end of year 6, the remaining pavement life is decreased to 5 years as one year elapsed. There are 7 maintenance actions implemented to maintain the serviceability level of road section *AD* with the remaining pavement life of 7 years at the end of the planning period.

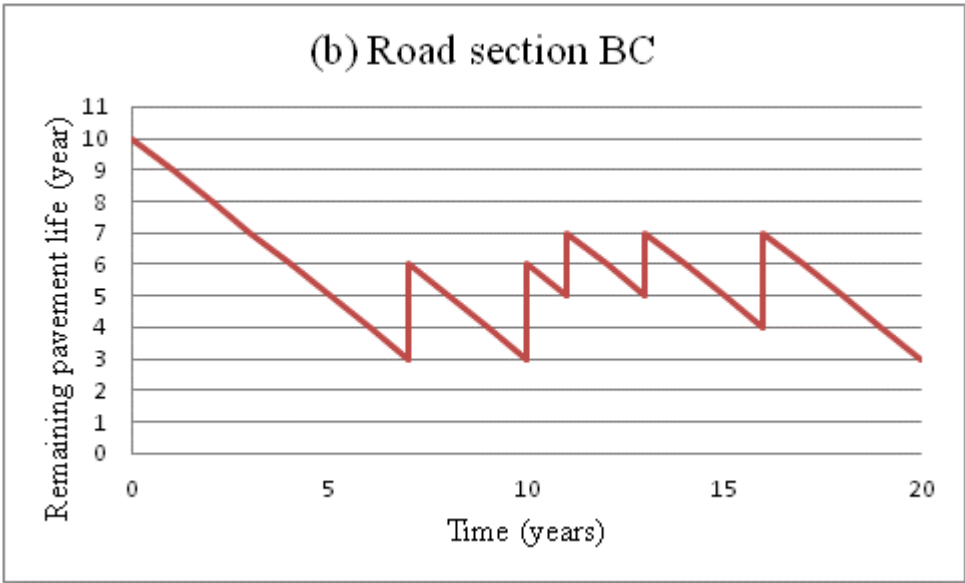
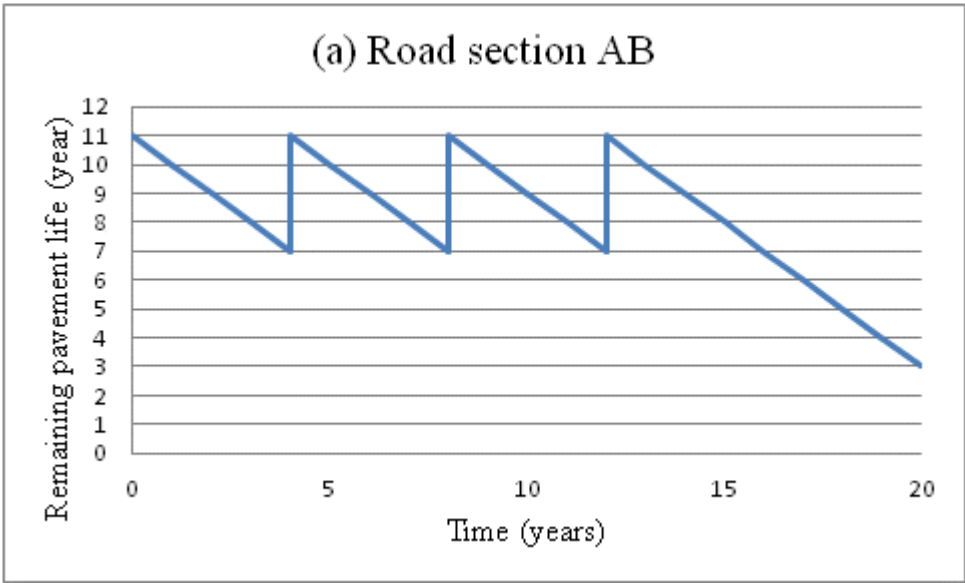
We also found that road sections *BC* and *AD* are maintained frequently at the late state of the planning period; this is because the maximisation of remaining pavement life is employed as one objective, and road sections *BC* and *AD* belong to single carriageway which requires less maintenance cost than dual carriageway and motorway to restore pavement condition.

When only the minimisation of maintenance cost is considered, the optimal M&R strategy is identified as the [0, 0] point at the normalised optimal Pareto frontier shown in Figure 5-7, which is described in Table 5-6; and the corresponding maintenance cost and total remaining pavement life are £1,264,076 and 12 years, respectively.

Table 5-6: The optimal M&R strategy with the objective of minimising maintenance cost

Road section	Maintenance scheduling during the planning horizon T (20 years)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
AB	0	0	0	0	2	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0
BC	0	0	0	0	0	0	0	2	0	0	2	1	0	1	0	0	2	0	0	0
CD	0	0	0	0	0	1	0	0	0	2	0	0	2	0	0	2	1	0	0	0
AD	0	0	0	0	0	2	0	0	2	0	0	2	0	0	2	0	0	2	0	0

According to Table 5-6, the evolution of remaining pavement life for road sections during the planning period is illustrated in Figure 5-9:



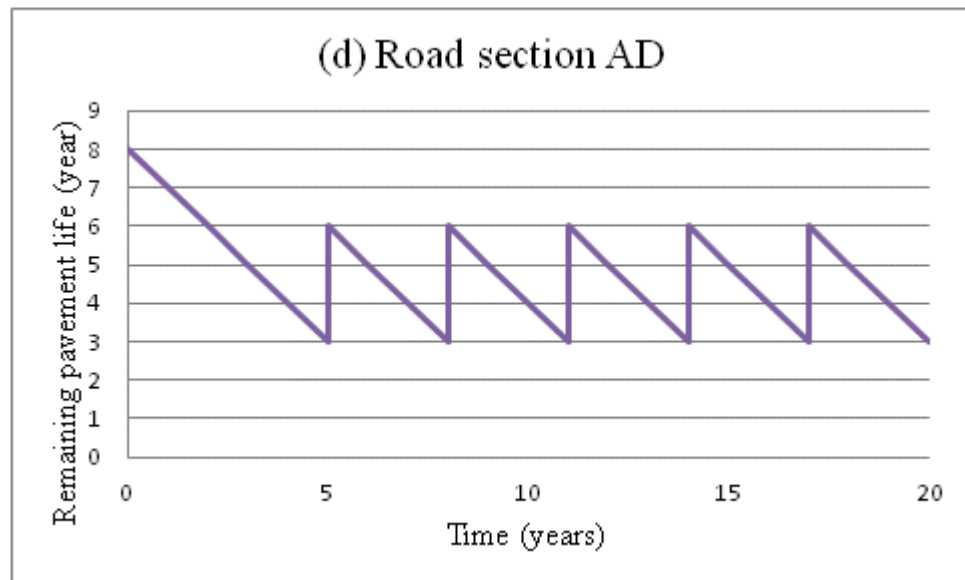
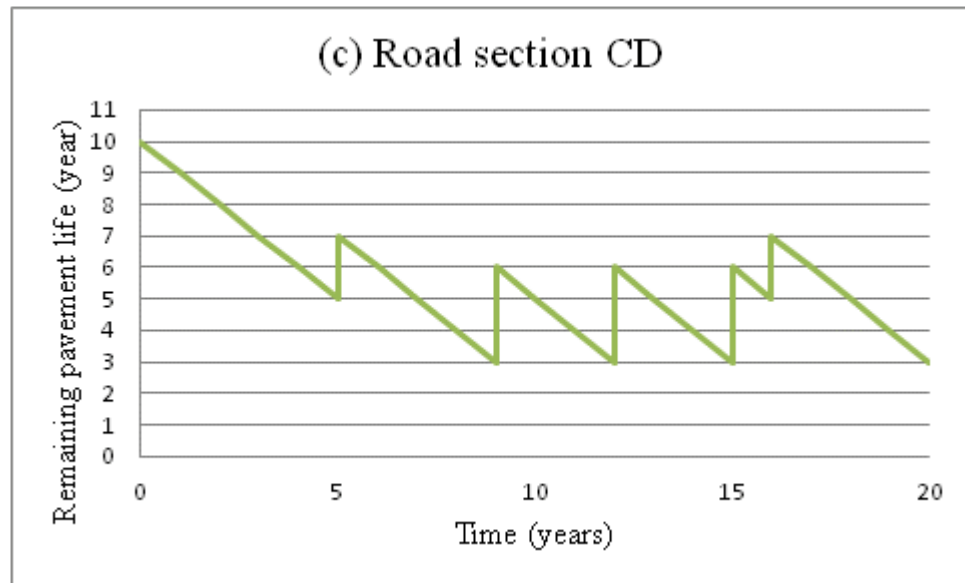


Figure 5-9: Pavement life cycle when only the minimisation of maintenance cost is considered: (a) Section AB, (b) Section BC, (c) Section CD, (d) Section AD

Examining Figure 5-9, we found that the remaining pavement life at the end of planning period for each road section is 3 years; this is because the objective of optimisation problem is selected as the minimisation of maintenance cost which led to the minimum total remaining pavement life. In reality, the selection of the optimal Pareto set is determined according to the decision making criteria of highways agencies.

### 5.2.5 Probabilistic Pavement Deterioration Model

As a matter of fact, the expected life for a new pavement will have variability and be characterised by a distribution rather than a definite value. The normal distribution is recognised as one of the most widely used probabilistic distributions in engineering problems. Its probability density function is symmetric about the mean,  $\mu$ , and shape is determined by its mean and standard deviation,  $\sigma$ . For instance, the initial life for a pavement with mean 10 years and standard deviation 0.5 is illustrated as:

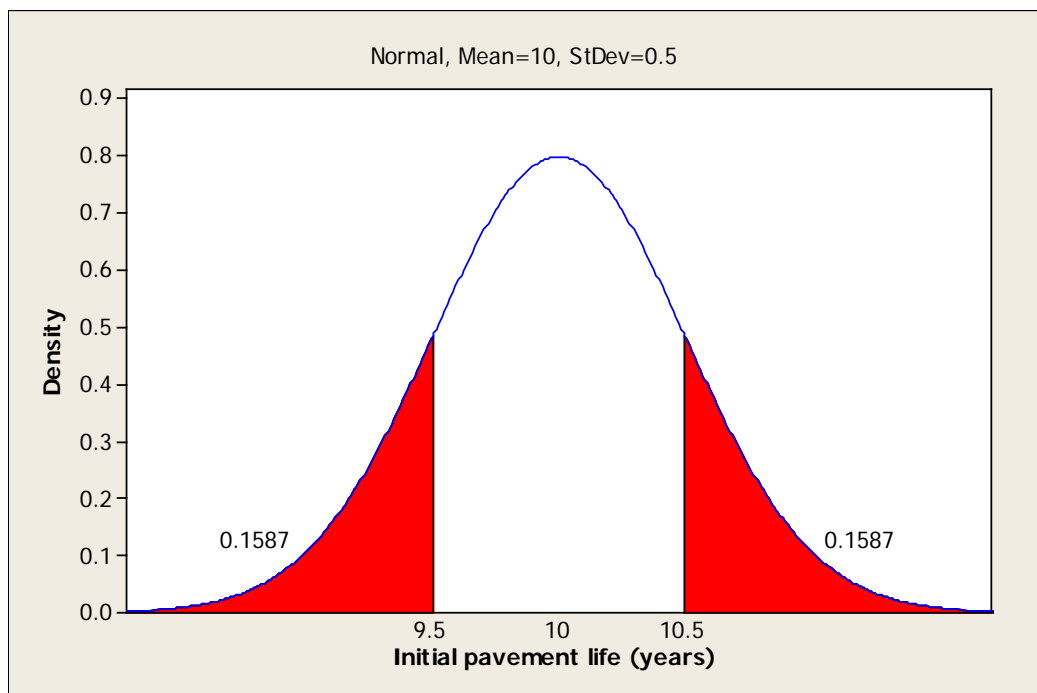


Figure 5-10: The distribution of initial pavement life

The probability density function for the normal distribution  $N(\mu, \sigma^2)$  is formulated as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty \leq x \leq +\infty \quad (5-12)$$

The cumulative density function can be achieved by integrating  $f(t)$ . However, Equation 5-12 cannot be integrated to derive a universal formula which is applicable for any given  $\mu$  and  $\sigma$ . As the integration of cumulative density function is performed using numerical methods and the results tabulated for different limits of the integration, it is

very difficult if this required a different probability distribution table for any possible set of  $\mu$  and  $\sigma$ . This obstacle is overcome by transforming any normal distribution  $N(\mu, \sigma^2)$  into a normal standard distribution with mean 0 and standard deviation 1, i.e.  $N(0, 1)$ .

This translation requires a new variable  $z$ , defined as:

$$z = \frac{x - \mu}{\sigma} \quad (5-13)$$

Consequently,  $z$  is normally distributed with mean 0 and standard deviation 1, termed the standard normal variate.

#### 5.2.5.1 Generation of Normal Variables

In pavement engineering, it is often desirable to generate variables that are normally distributed. The generation of any normal variate  $Y(\mu, \sigma^2)$  can be converted to the generation of the standard normal variate  $Z$  by  $y = \sigma \times z + \mu$ . All the approaches for generating standard normal variate are developed based on the availability of a random number generator capable of producing uniform random variables over the range 0 to 1, i.e.  $U(0,1)$ ; some methods widely used are described explicitly as follows.

##### 5.2.5.1.1 Central Limit Theorem

Suppose  $X_1, X_2, \dots, X_n$  are independent variables that follow the identical distribution with mean  $\mu$  and standard deviation  $\sigma$ . Consequently, when  $S_n = \sum_{i=1}^n X_i$ , the random variable  $\frac{S_n - n\mu}{\sigma\sqrt{n}}$  is characterised by the normal standard distribution.

For simplicity, the uniformly distributed variables  $U(0,1)$  are used to construct  $S_n$ . With the increase of  $n$ , the resulting  $S_n$  will be more approximate to a real normal distribution. In terms of the mathematical point,  $n=12$  is usually adopted. As  $U(0,1)$  has a mean 0.5 and standard deviation  $1/12$ , the resulting  $S_{12}$  is determined as  $N(6,1)$ .

By this means, a random sample for the normal distribution  $S_{12}$  is produced as:



$$X = \sum_{i=1}^{12} X_i \quad (5-14)$$

In terms of the central limit theorem,  $X$  is determined as a normal distribution with mean 6 and standard deviation 1. Afterwards, the random sample from the normal distribution  $Y(\mu, \sigma^2)$  is obtained as:

$$Y = (X - 6)\sigma + \mu \quad (5-15)$$

#### 5.2.5.1.2 The Box–Muller Method

The first step is to generate two independent random variables  $U$  and  $V$  that are uniformly distributed on  $(0, 1)$ . The next step is to compute two random variables  $X$  and  $Y$  using  $U$  and  $V$ :

$$\begin{aligned} X &= \sqrt{-2 \ln U} \cos(2\pi V) \\ Y &= \sqrt{-2 \ln U} \sin(2\pi V) \end{aligned} \quad (5-16)$$

Both  $X$  and  $Y$  follow the normal standard distribution, and will be correlated. The final step is to calculate the random sample from the normal distribution using  $X$  and  $Y$ .

#### 5.2.5.1.3 Marsaglia Polar Method

Marsaglia polar method is developed based on the Box-Muller method. In this approach, two variables  $U$  and  $V$  that uniformly distributed on  $(-1, 1)$  are required, and then  $S=U^2+V^2$  is obtained. If  $S$  is greater than or equal to one this algorithm starts over, otherwise two normal standard variates  $X$  and  $Y$  are calculated as:

$$\begin{aligned} X &= U \sqrt{\frac{-2 \ln S}{S}} \\ Y &= V \sqrt{\frac{-2 \ln S}{S}} \end{aligned} \quad (5-17)$$

Based on  $U$  and  $V$ , two normal variables are generated.

### 5.2.5.2 Probabilistic Pavement Age Gain Model

Probabilistic pavement models are able to incorporate the uncertainty and predict the pavement performance as the probability of occurrence of a range of possible outcomes. On one hand, the deterministic approach is superior to the probabilistic approach if only limited data of pavement condition is available. On the other hand, the probabilistic approach presents a probability distribution of pavement states which describes the stochastic nature of pavement deterioration.

The deterministic pavement age gain model applied in this thesis is transformed to a probabilistic one using the normal distribution. The pavement condition for each road section group is depicted in Table 5-7:

Table 5-7: Road section types in probabilistic pavement model

Road section group	Road class	Traffic flow (000s)	Initial pavement life (years)	
			$\mu$	$\sigma$
1	S2	5-15	10	0.5
2	S2	20	8	0.5
3	D2AP	10-20	10	0.5
4	D2AP	30	9	0.5
5	D2AP	40	8	0.5
6	D3M	20-30	11	0.5
7	D3M	40	10	0.5
8	D3M	80	9	0.5

and the expected age gain for each maintenance action is evaluated in Table 5-8:

Table 5-8: Expected ages associated with pavement maintenance actions versus remaining pavement life in probabilistic pavement model

Remaining pavement life (years)	Pavement condition	Expected age gain (years)									
		Do nothing		Patching		SD		Re		Ov	
		(0)		(1)		(2)		(3)		(4)	
		$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
>6	0	0	0	2	0.3	5	0.3	7	0.3	10	0.3
(4,6]	1	0	0	2	0.3	3	0.3	7	0.3	10	0.3
(2,4]	2	0	0	1	0.3	3	0.3	6	0.3	9	0.3
(0,2]	3	0	0	0	0.3	1	0.3	6	0.3	9	0.3

For the pavement described in Figure 5-10, its remaining pavement life after 5 years would be characterised by a normal distribution with mean 5 and standard deviation 0.5, described in Figure 5-11:

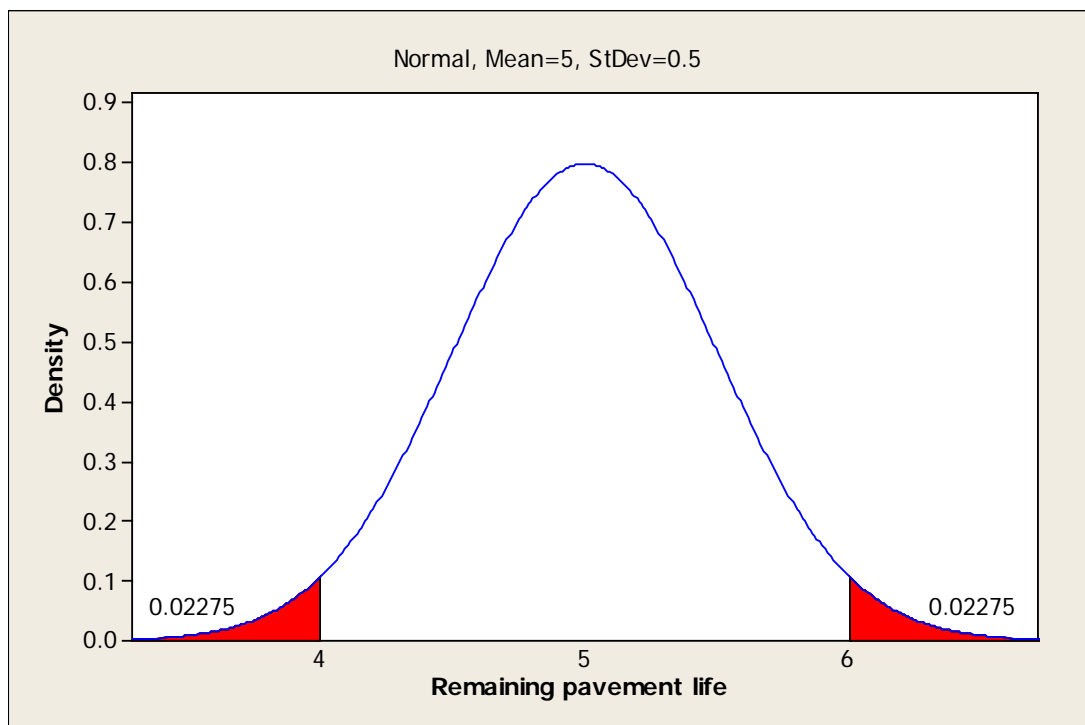


Figure 5-11: Probabilistic remaining pavement life

The probability of the remaining pavement life within the scale of [0, 6] is 97.725%. Afterwards, SD is applied to maintain the pavement. Within this specified scale, SD is dominated by a normal distribution with mean 3 and standard deviation 0.3. While for the rest of the whole space, SD is characterised by a normal distribution with mean 5 and standard deviation 0.3.

### 5.2.5.3 Example

For the road section *AB* in Figure 5-1, the optimal M&R strategy depicted in Table 5-5 is used to maintain this road in 20 years. To test the performance of section *AB* based on probabilistic pavement age gain model, 5 runs are implemented using the central limit theorem as only one normal variable is generated at each time, described in Figure 5-12:

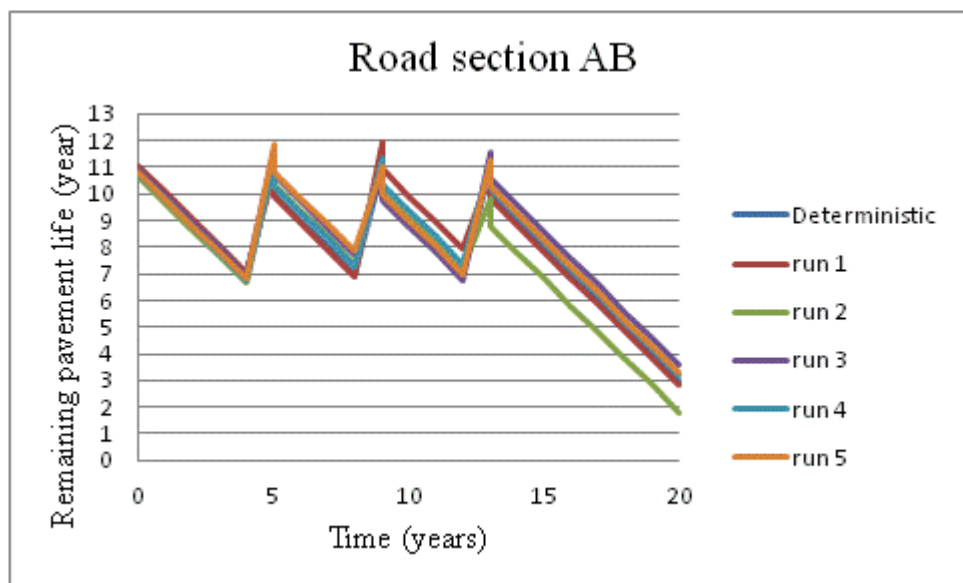


Figure 5-12: Probabilistic pavement life cycle

Figure 5-12 shows that the results obtained in each run possess the similar pattern. However, the variations of remaining pavement life are identified. Especially for run 2, the remaining pavement life at year 20 is less than 2 years, which indicates an obvious difference from the deterministic one. Thus the same M&R strategy can result in quite different pavement conditions. Afterwards, the distribution of the remaining pavement

life at year 20 is investigated, illustrated in Figure 5-13. It demonstrates that most of the samples are within the range of [2.2, 4.0]. The mean of distribution is identified as 3.039 years, which is slightly higher than the deterministic remaining pavement life, 3 years.

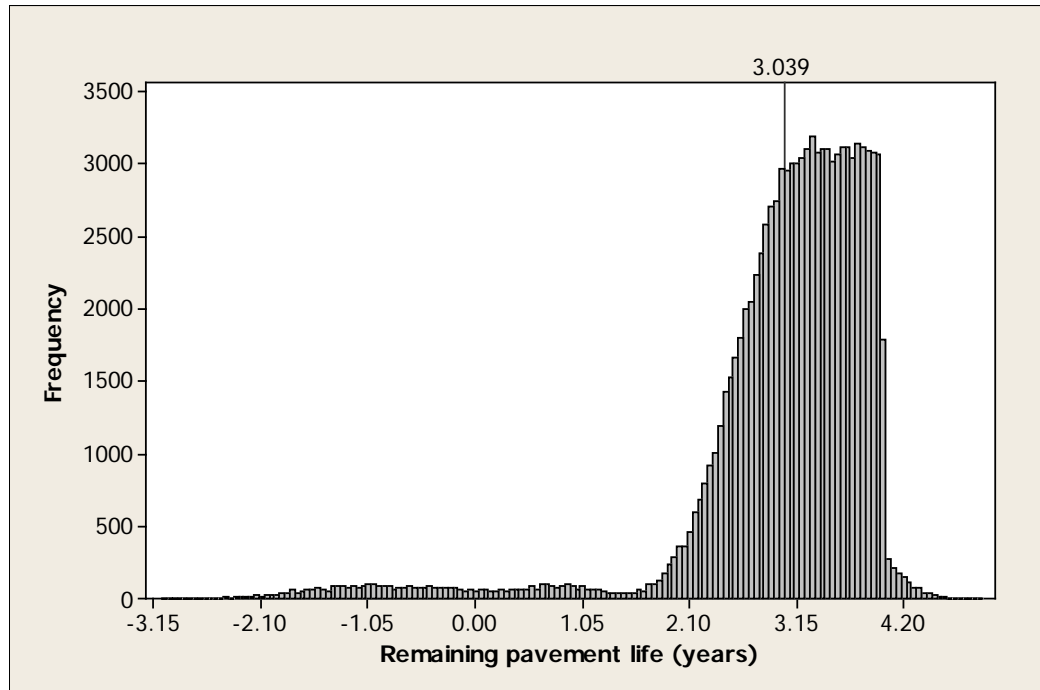


Figure 5-13: Distribution of the remaining pavement life of road section AB

Subsequently, the road network in Section 5.2.4 is investigated; the only difference is that pavement lives and expected age gain associated with each maintenance action are characterised by normal distributions rather than definite values. Based on the central limit theorem, these normal variables are generated randomly. NSGA-II is used to find the optimal Pareto frontier using probabilistic pavement model and 10 runs are carried out. The GA parameters are set to:

$p_c$ : 0.6

$p_m$ : 0.01

$M$ : 5000

$P$ : 500

The optimal Pareto frontier obtained in each run is illustrated in Figure 5-14. It shows that each optimal Pareto frontier follows the same trend with a little variation. Also the average optimal Pareto frontier is identified, which is slightly above the deterministic optimal Pareto frontier that was obtained in Section 5.2.4.4. This is because for the probabilistic model, the remaining pavement life is allowed to exceed the maximum remaining pavement life. In addition, the GAs aim to find the solutions with less cost and long remaining pavement life, so the solutions that with the same maintenance cost and more remaining life would be retained on the optimal Pareto frontier. The variation of the probabilistic optimal Pareto frontier indicates the stochastic nature of pavement, and demonstrates that the probabilistic model is more realistic. However, the probabilistic model requires more computational effort than the deterministic one.

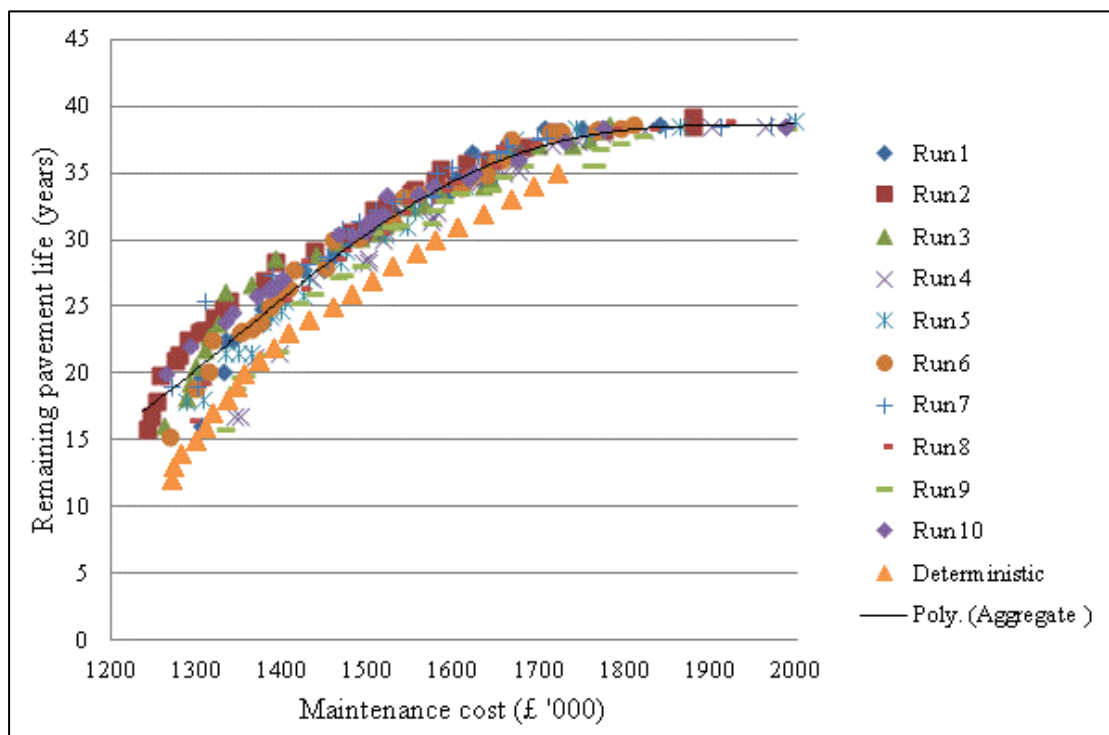


Figure 5-14: Probabilistic Pareto optimal frontier

### **5.3 Scheduling of Pavement Maintenance using the NTFM at the Project Level**

M&R activities on highways usually occupy the road, disrupt traffic flow and increase safety hazards for road users and workers. Therefore, M&R activities are not only expensive in terms of the agency cost but also from the viewpoint of the user cost. The user cost due to traffic delays frequently far exceeds the maintenance cost. Efficient scheduling of M&R through a worksite may greatly reduce the total cost, including agency and user cost. A “worksite” is an area on a two-lane or multi-lane highway, where one or more lanes are closed, so that the selected M&R activities can be conducted [103].

Since the scheduling of M&R activities is a major concern of the highways agency and local authorities, many studies have been performed in this area. Few studies have focused on the optimisation of worksite length considering the impact of agency cost and road user cost. A mathematical model was proposed in [104] to optimise the worksite length and related traffic control on two-lane, two-way highways where one lane closure is applied. Afterwards, Chien and Schonfeld [105] applied this model to optimise the worksite length on four-lane, two-way highways where one lane in one direction at a time is closed. The underlying assumption for these studies is that the traffic flow heading to the worksite was steady and given.

Some studies have been devoted to the optimisation of the worksite schedule, which yielded the minimal total cost, i.e. agency cost and road user cost. The mathematical approach developed in [104] was also used to optimise worksite schedules [103]. Considering lane closure situations, Fwa and Cheu [106] optimised the worksite schedules to minimise traffic delays subject to constraints of maintenance operational requirements using GAs. Jiang and Adeli [107] proposed a Boltzmann-simulated annealing neural network to optimise the worksite arrangements, i.e. worksite length,

start time of the worksite, when considering lane closure options and working in darkness.

Considering the significant traffic delays resulting from the implementation of roadworks, drivers may change their journey behaviour, such as taking alternative routes to circumvent a congested road section with a maintenance worksite. The methodology developed in [105] was deployed to jointly optimise worksite length and diversion fraction considering the traffic condition of the alternative route bypassing the worksite [108]. Afterwards, an analytical model is constructed to find the optimal worksite length, maintenance schedule and the traffic diversion fraction, while considering time-varying traffic demand, variable maintenance cost and the production rates of maintenance teams. In addition, the capacities and speed controls on worksites are investigated, as well as the road users and workers' safety [109-112].

However, the techniques provided above mainly focused on the traffic delay occurring at the worksite, while the traffic conditions in the adjacent links are not considered. As the traffic flows through the network are dependent on each other, it is more realistic to optimise the worksite arrangements by considering its impact on the network performance. In this thesis, NTFM is applied as the platform from which the maintenance arrangements at the worksites are optimised so as to minimise the detrimental influence to the road users.

### 5.3.1 Methodology

The methodology developed in this chapter formulates a total cost objective function which is used to optimise lane closure options, work schedule, and traffic control strategies at the worksites. It is based on NTFM considering time-dependent traffic demand. The total cost function is composed of the agency cost and the road user cost.



### 5.3.1.1 Agency Cost

The agency cost is evaluated as the sum of the maintenance costs spent on individual worksites. The maintenance cost incurred on worksite  $i$  is formulated as:

$$C_i = c_0 + c_{k,r} \times l_i \quad (5-18)$$

Where

$c_0$  - the fixed cost for setting up a worksite, £ 1000

$k$ - the maintenance action implemented on worksite  $i$

$r$ - the road type of worksite  $i$

$c_{k,r}$ - the cost for maintenance action  $k$  associated with the road type  $r$  per lane kilometre, £ '000s

$l_i$ - the length of worksite  $i$ , lane kilometre

The corresponding maintenance duration spent on worksite  $i$  is represented by:

$$D_i = d_0 + d_{k,r} \times l_i \quad (5-19)$$

$d_0$ - the fixed set up time for a worksite, 1 hr

$d_{k,r}$ - the time required for maintenance action  $k$  associated with the road type  $r$  per lane kilometre, hrs

The unit cost for each combination of maintenance action and road type is described in Table 5-9 [101], which is obtained based on Table 5-1 by transforming the unit from days to hours.

Table 5-9: Maintenance duration and cost for each combination of maintenance action and road type (per lane kilometre)

M&R	Single carriageway (S2)		Dual 2 lane carriageway (D2AP)		Dual 3 lane Motorway (D3M)	
	Duration (hrs)	Cost (£ '000s)	Duration (hrs)	Cost (£ '000s)	Duration (hrs)	Cost (£ '000s)
Do nothing	0	0	0	0	0	0
Patching	24	15	18	15	16	13.33
Surface dressing	48	20	30	20	24	16.67
Resurfacing	96	60	84	82.5	80	90
Overlay	192	95	138	122.5	132	135

1. Costs are £'000s expressed in average 2012 price.

#### 5.3.1.2 Road User Cost

Road user cost is determined from the total travel duration spent in the highway network multiplied by the time value. For instance, the road user cost spent on section  $j$  is computed as:

$$T_j = t_j \times v \quad (5-20)$$

$t_j$  - total travel duration spent on road section  $j$

$v$  - time value of road user, the market price value of time for an average vehicle is £15.38 per hour, 2012 prices and values [61]

Moreover, the travel duration  $t_j$  consists of the journey time spent on section  $j$  and the travel delay which takes place at the junction downstream of section  $j$ , described as  $t_i = t_{i,j} + t_{i,d}$ . The evaluation of the journey time and travel delay on a road section is described in previous chapters.

### 5.3.1.3 Lane Closure Situations

In NTFM, three road types are investigated, i.e. S2, D2AP and D3M; the closure situation for each road type is described in the following sections.

#### 5.3.1.3.1 Single Carriageway

For a single carriageway road, a shuttle working site is employed to facilitate the traffic flows from opposing directions. The evaluation of traffic conditions on the shuttle working site is explained explicitly in Section 3.2.8.1.

As for the average speed through the worksite, it has been assumed that the travel speeds for light vehicles and heavy vehicles through the worksite are 43.9 km/hr and 38.2 km/hr, respectively, with 10% heavy vehicles [113]. In this thesis, the average speed at the worksite is selected as the travel speed for heavy vehicles.

#### 5.3.1.3.2 Dual 2 lane Carriageway

As for a dual 2 lane carriageway, a one lane closure and or a two lane closure can be used to implement the maintenance actions, which are depicted in Figure 5-15. For a one lane closure, only one of the service lanes in one direction is closed owing to maintenance activity, in this case maintenance has no impact on the traffic flow in the opposite direction. While for a two lane closure, both of the service lanes in one direction are closed, and the road at the worksite becomes effectively a single carriageway. At the direction with one service lane, the average speed is reduced to 80 km/hr from 112 km/hr.

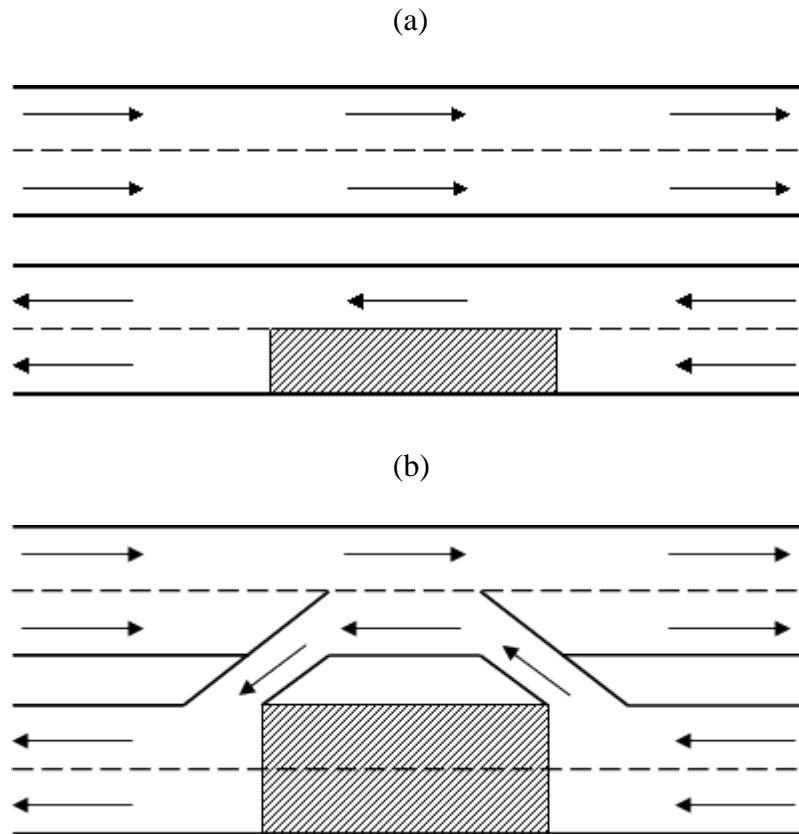


Figure 5-15: Dual two lane carriageway road with work: (a) one lane closure; (b) two lane closure and crossover

#### 5.3.1.3.2.1 One Lane Closure

Provided that one of the service lanes in one direction is closed due to maintenance its flow capacity is decreased to zero, and the flow capacity on the other lane in this direction is also affected. According to the delay modelling in QUADRO [58], the overall worksite capacity would be reduced to:

$$C_{D,1} = 0.85 \times C_{n,i} \quad (5-21)$$

Where

$C_{n,i}$ - normal capacity for a standard lane that belongs to road class  $i$

#### 5.3.1.3.2.2 Two Lane Closure

For a two lane closure, the number of service lanes in each direction is reduced to one. The resulting site capacities for both directions are calculated by Equation 5-21.

#### 5.3.1.3.3 Dual 3 Lane Motorway

As for D3M, the lane closure options include a one lane closure, a two lane closure, full carriageway closure and crossover. For a one lane closure, the average speed through the worksite remains 112 km/hr. When a two lane closure is implemented, the average speed is decreased to 80 km/hr.

If  $N$  lanes are in service through the worksite during maintenance, the worksite capacity would be formulated as Equation 3-46.

#### 5.3.1.4 Solution Algorithm

The decision variables of the total cost function include the lane closure options, arrangements of start time, and traffic controls for traffic flows, considering the variation of flows over time. In this section, the total cost is minimised by optimising these decision variables.

The application of a worksite is achieved by simply reducing the flow capacity as defined in Section 5.3.1.3 and reducing the link capacitance. The additional road user cost is evaluated by subtracting the road user cost incurred in the network under normal conditions from that spent in the network with the maintenance worksite. In this manner, the impact of the maintenance activity on the network performance is obtained, and the worksite arrangement that caused the least additional cost is chosen. One underlying assumption is that the speed reductions at the worksite due to the enforcement of road worker safety are not considered. As a result, the speed restriction at worksite is only influenced by the number of service lanes in this thesis.

### 5.3.2 Case Study

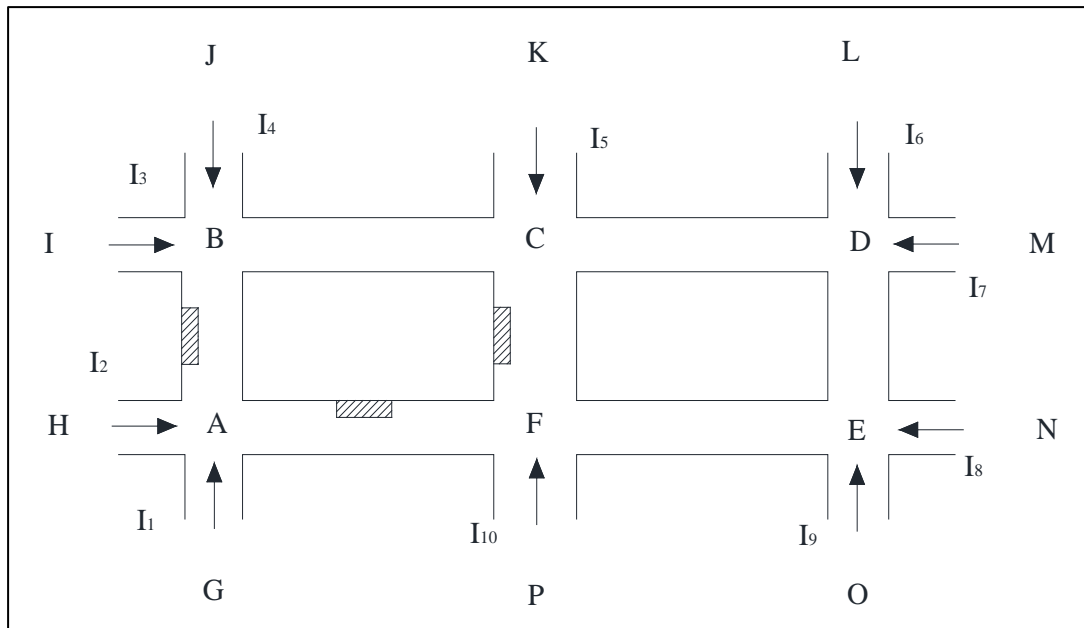


Figure 5-16: A simple road network with maintenance worksites

This road network is signal controlled; therefore the traffic interactions among the flows from different sources are removed. Road links *AB* and *DE* are D2AP, *CF* is a D3M road link; while the rest of the links belong to S2, which are used to connect the major roads. The lengths of links *AB*, *FC* and *DE* are 3.5 km, other links are 7 km. 3.5 km correspond to the capacitance of 500 pcu per lane. Three worksites are assigned to this network, i.e. *RN1*, *RN2* and *RN3*, on which patching is to be implemented to maintain the pavement surface. The average speed on S2 and other road types are defined as 50 and 112 km/hr, respectively. The flow capacity for each junction is fixed to 2000 pcu/hr, the green splits for the traffic that is going straight on and turning left, and the traffic turning right are defined as 0.5 and 0.3 respectively, and the junction turning ratios on each arm are chosen as 0.3, 0.4, and 0.3. The traffic inputs for each external node, i.e. *I1*, *I2*... *I10*, are identical and vary throughout the day, as shown in Table 5-10:

Table 5-10: Traffic inputs for external nodes

Time	7	8	9-15	16-18	19-21	22	23	0-5	6
Flow rate (pcu/hr)	1000	1000	500	1000	500	200	100	10	100

Based on the parameters described above, NTFM is used to evaluate the performance of the road network under normal conditions during a day, the results of which are given in Table 5-11:

Table 5-11: Daily network performance under normal conditions

Time	Inflow (pcu/hr)	Outflow (pcu/hr)	Queue (pcu)	Journey time (hrs)	Travel delay (hrs)	Total travel duration (hrs)	Agency cost (£)	Road user cost (£)	Total cost (£)
7-8	10000	10000	0	537.5	0	537.5		8266.75	8266.75
9-15	5000	5000	0	268.75	0	268.75		4133.375	4133.375
16-18	10000	10000	0	537.5	0	537.5		8266.75	8266.75
19-21	5000	5000	0	268.75	0	268.75		4133.375	4133.375
22	2000	2000	0	107.5	0	107.5		1653.35	1653.35
23	1000	1000	0	53.75	0	53.75		826.675	826.675
0-5	100	100	0	5.375	0	5.375		82.6675	82.6675
6	1000	1000	0	53.75	0	53.75		826.675	826.675
Total							0	86470.2	86470.2

Examining Table 5-11, it demonstrates that this road network is sufficient for delivering the required traffic flow and no traffic delay occurred. Subsequently, the performance of road network under each maintenance scenario is studied. The underlying assumption is that all the maintenance activities should be completed in one day, i.e. from 7:00 to 7:00 on the next day.

### 5.3.2.1 Road Network with RN1

Roadwork node *RNI* is assigned to the centre of link *AB* where one lane closure is applied, and the length of worksite is 0.35 km. This maintenance activity is required to be accomplished in one day, and so the start time of the maintenance can be varied from

7:00 to 23:00. Using Equations 5-18 and 5-19, the incurred maintenance cost and maintenance duration are £ 6250 and 8 hrs.

Initially, the start time of the maintenance is set to 9:00; therefore one lane is closed on *RNI* for maintenance between 9:00 and 17:00. During this period the flow capacity of *RNI* is reduced to 1530 pcu/hr. The resulting traffic conditions in the network are described in Table 5-12:

Table 5-12: Daily network performance with RN1

Time	Inflow (pcu/hr)	Outflow (pcu/hr)	Queue (pcu)	Journey time (hrs)	Travel delay (hrs)	Total travel duration (hrs)	Agency cost (£)	Road user cost (£)	Total cost (£)
7-8	10000	10000	0	537.5	0	537.5		8266.75	8266.75
9-15	5000	5000	0	269.375	0	269.375		4142.988	4142.988
16	10000	10000	0	538.75	0	538.75		8285.975	8285.975
17- 18	10000	10000	0	537.5	0	537.5		8266.75	8266.75
19- 21	5000	5000	0	268.75	0	268.75		4133.375	4133.375
22	2000	2000	0	107.5	0	107.5		1653.35	1653.35
23	1000	1000	0	53.75	0	53.75		826.675	826.675
0-5	100	100	0	5.375	0	5.375		82.6675	82.6675
6	1000	1000	0	53.75	0	53.75		826.675	826.675
Total							6250	86556.7	92806.7

According to Table 5-12, there are no traffic delays occurring on the network. This is because the residual flow capacity on *RNI* is still higher than the flow entering through link *AB* which is 1000 pcu/hr during the rush hour. Compared to the normal conditions the only difference is that the journey time during maintenance is slightly higher, since the average speed through the worksite is decreased to 80 km/hr from 112 km/hr so as to ensure the safety of road users and the maintenance workforce. The impact of *RNI* on the network performance is quite limited, whereas the additional road user cost can still



be minimised by varying the start time of maintenance. The daily network performance with various start times is compared in Table 5-13:

Table 5-13: Daily network performance with RN1 under various maintenance scenarios

Start time of maintenance	Agency cost (£)	Road user cost (£)	Total cost (£)
7	6250	86566.3	92816.3
8	6250	86556.7	92806.7
9	6250	86556.7	92806.7
10	6250	86566.3	92816.3
11	6250	86575.9	92825.9
12	6250	86575.9	92825.9
13	6250	86575.9	92825.9
14	6250	86575.9	92825.9
15	6250	86570.2	92820.2
16	6250	86562.5	92812.5
17	6250	86543.5	92793.5
18	6250	86524.4	92774.4
19	6250	86505.4	92755.4
20	6250	86496.0	92746.0
21	6250	86486.5	92736.5
22	6250	86477.1	92727.1
23	6250	86475.2	92725.2

Table 5-13 shows that the maintenance scenario with a start time of 23:00 led to the minimal total cost. This is as expected because the maintenance is conducted at nighttime during which the least number of vehicles were travelling through the network.

### 5.3.2.2 Road Network with RN2

Roadwork node *RN2* is located in the center of link *FC* and the length of worksite is set to 0.21875 mile. Using Equations 5-18 and 5-19, the required maintenance cost and

duration are £5667 and 7 hrs, respectively. As a result, the start time of maintenance is within 7:00 and 24:00 the next day. For instance, the start time of 9:00 is selected and one service lane is closed for maintenance. During maintenance the residual flow capacity of *RN2* is obtained as 3400 pcu/hr using Equation 3-46. NTFM is applied to evaluate the traffic conditions in the road network with *RN2*, and the same results are obtained as Table 5-11. This is due to the fact that there are still two lanes on *RN2*, where the average speed remains 112 km/hr and the flow capacity is much greater than the entry flow. In this case, the start time of maintenance work exerts no influence on the network performance. When a two lane closure is applied on *RN2*, its flow capacity is reduced to 1700 pcu/hr which is still greater than the maximum traffic flow entering link *FC* during a day. However, as there is only one lane open, the average speed through *RN2* is reduced to 50 mile/hr which resulted in additional road user costs. In this case the incurred additional road cost can be minimised by varying the start time of maintenance as Section 5.3.2.1.

### 5.3.2.3 Road Network with *RN3*

Roadwork node *RN3* is assigned to link *AF* and the geometry of *RN3* is identical to *RN1*. The required maintenance cost and duration for *RN3* are £6250 and 10 hrs, respectively. As link *AF* is single carriageway, signal control is implemented to assign gaps for each directional flow. The amber phase is set to 10% of the total cycle. Initially, the green phases for both directions are defined as 45%. In terms of Equations 3-44 and 3-45, the flow capacity for each direction is 810 pcu/hr. The start time of maintenance is 9:00, and the resulting traffic conditions in the network are illustrated in Table 5-14:

Table 5-14: Daily network performance with RN3

Time	Inflow (pcu/hr)	Outflow (pcu/hr)	Queue (pcu)	Journey time (hrs)	Travel delay (hrs)	Total travel duration (hrs)	Agency cost (£)	Road user cost (£)	Total cost (£)
7-8	10000	10000	0	537.5	0	537.5		8266.8	8266.8
9-15	5000	5000	0	272.5	0	272.5		4190.7	4190.7
16	10000	9624	377	519.6	188.5	708.1		10890.3	10890.3
17	1000	9412	967	519.3	672.0	1191.3		18321.7	18321.7
18	1000	9355	1614	526.5	1290.5	1817.0		27946.0	27946.0
19	5000	6609	0	384.5	296.7	681.2		10476.4	10476.4
20-	5000	5000	0	268.8	0	268.8		4133.4	4133.4
21									
22	2000	2000	0	107.5	0	107.5		1653.4	1653.4
23	1000	1000	0	53.8	0	53.8		826.7	826.7
0-5	100	100	0	5.4	0	5.4		82.7	82.7
6	1000	1000	0	53.8	0	53.8		826.7	826.7
Total							6250	125572	131822

From Table 5-14 it can be seen that the road user cost increased significantly. The delay cost to the road users is much greater than the corresponding maintenance cost; this is because the opposing flows have to switch to use the single available lane. As a consequence, traffic delays occurred on both lanes on link *AF*. In order to minimise traffic delay to the road users the effects of varying the start time of maintenance is investigated. The results are shown in Table 5-15:

Table 5-15: Daily network performance with RN3 under various maintenance scenarios

Start time of maintenance	Agency cost (£)	Road user cost (£)	Total cost (£)
7	6250	110606.0	116856.0
8	6250	110020.0	116270.0
9	6250	125572.0	131822.0
10	6250	139124.0	145374.0
11	6250	141194.0	147444.0
12	6250	141194.0	147444.0
13	6250	141160.0	147410.0
14	6250	141114.0	147364.0
15	6250	141058.0	147308.0
16	6250	141001.0	147251.0
17	6250	105976.0	112226.0
18	6250	91550.0	97800.0
19	6250	86682.3	92932.3
20	6250	86626.1	92876.1
21	6250	86580.3	92830.3

Table 5-15 indicates that the maintenance scenario with start time of 21:00 led to the least additional road user cost, as a small number of vehicles enter the network during the maintenance time. Apart from the start time of maintenance, the green splits for each direction can be optimized to minimise traffic delay. Given a start time for the maintenance of 9:00, the green phase for the traffic from junction A is varied from 30% to 60%, while the amber phase remains at 10%. The resulting network performance with different green splits on RN3 is illustrated in Figure 5-17:

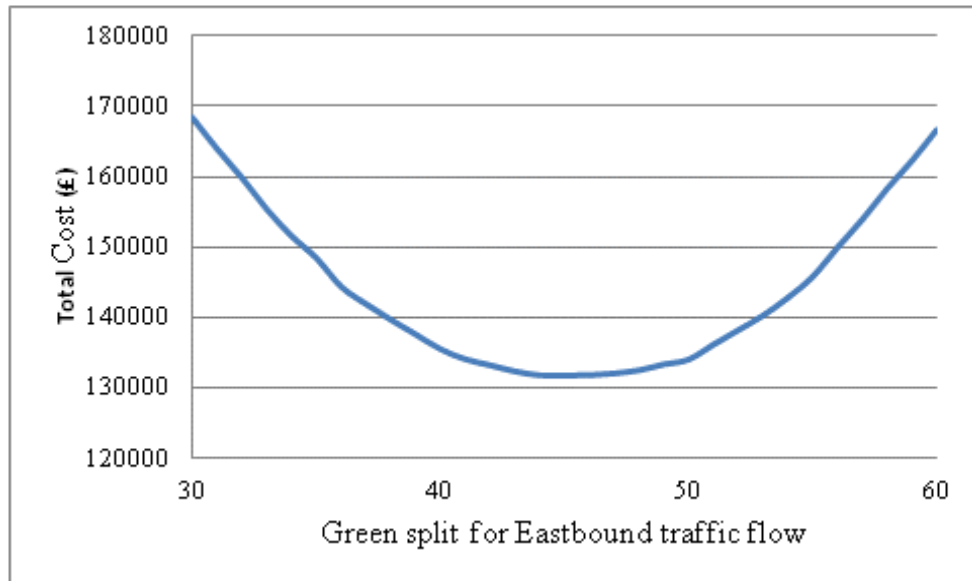


Figure 5-17: The network performance with various green phases

Figure 5-17 illustrates that the most optimal green split is located in the [40%, 50%] range, this is because the opposing flows through *RN3* are identical. It is therefore reasonable to treat the two directional flows equivalently. Moreover, the green split for each directional flow and start time of maintenance can be optimized jointly to find the best traffic control strategy.

In addition to the optimization of the green splits for directional flows, traffic rerouting strategy, proposed in Section 3.7, can be applied to alleviate the traffic congestion occurred during rush hour in the network. For instance, the traffic flow from junction *A* to junction *F* can be diverted to the alternative route *A-B-C-F* when a queue exists on link *AF*, and the traffic diversion proportion can be optimised to minimise the travel delay spent on link *AF*.

#### 5.3.2.4 Road Network with three Roadwork Nodes

In this section, three roadwork nodes are assigned to the road network concurrently. It further requires the maintenance activities to be carried out in one day. The start times of maintenance activities on *RN1*, *RN2* and *RN3* are optimized simultaneously so as to

relieve the effect of maintenance activities on the network performance. The optimal worksite arrangements are shown in Table 5-16:

Table 5-16: The optimal arrangements of worksites

Roadwork node	RN1	RN2	RN3
Start time of maintenance	23:00	7:00-24:00	21:00

As stated above, the maintenance work conducted on *RN2* has no influence to the road users when a one lane closure is applied. While for the other two worksites, maintenance activities are scheduled during nighttime so as to avoid the excessive traffic delays and the associated user costs due to heavy traffic volume. The costs under the optimal maintenance arrangements are summarized in Table 5-17:

Table 5-17: The minimal agency cost and road user cost

Agency cost (£)	Road user cost (£)	Total cost (£)
18166.7	86580.3	104747

The cost to the road users is slightly higher than that spent in the network under normal conditions, £86470.2, described in Table 5-11. The results indicate that a better maintenance schedule can eliminate, or at least reduce, the influence of maintenance activities to the road users.

#### 5.4 Maintenance Planning using the NTFM in the Long-term

By comparing to the optimisation method proposed in Section 5.2, road user costs are also taken into account to optimise the M&R strategy for road sections on a highway network, which is realised by applying NTFM to simulate the traffic characteristics on the network under both normal and maintenance conditions. In this section, in addition to the total costs, pavement deterioration is taken into account over the whole life cycle, for example, 60 years. The aim of this optimisation model is to minimise the total costs spent on road section *k* over the pavement life cycle, formulated as:

$$\text{Min}[\frac{T_k}{MTTF_{k,i}}M_{k,i} + \frac{T_k}{MTTF_{k,i}}U_{k,i}] \quad (5-22)$$

Where

$T_k$  - life cycle of road section  $k$ , years

$MTTF_{k,i}$  - mean time to failure, which represents the time that road section  $k$  deteriorates from the new condition to the serviceable level that required maintenance action  $i$ , years

$\frac{T_k}{MTTF_{k,i}}$  - the number of maintenance action  $i$  required through the life cycle. It is noted that  $\frac{T_k}{MTTF_{k,i}}$  is rounded to the adjacent larger integer number when it is a decimal number

$M_{k,i}$  - the cost of maintenance action  $i$  performed on road section  $k$ , depends on the length and type of road section  $k$ , £'000s

$U_{k,i}$  - additional cost to road users due to the implementation of maintenance action  $i$  on road section  $k$ , £'000s

The additional road user cost is evaluated as:

$$U_{k,i} = u_{k,i} - u_n \quad (5-23)$$

$u_n$  - road user cost spent on the network under normal conditions, £'000s

$u_{k,i}$  - road user cost spent on the network when maintenance action  $i$  is performed on road section  $k$ , £'000s

The underlying assumptions for this optimisation model are listed as:

1. Only one kind of maintenance action is implemented repeatedly for the investigated road section.
2. The traffic inputs for the network are constant throughout the life cycle.

3. Maintenance actions are performed at both daytime and nighttime without maintenance break, and the maintenance costs for daytime operation and nighttime operation are the same. It means that maintenance work starts from 7:00 until the end.
4. Worksite arrangements are the same when different maintenance actions are performed on the investigated road section.

The mean time to failure for each maintenance action associated with each road type is depicted in Table 5-18:

Table 5-18: Mean time to failure for maintenance actions

<i>Pavement condition state</i>	<i>Trigger of maintenance action</i>	<i>Pavement condition transition</i>	<i>Single carriageway (S2)</i>	<i>Dual 2 lane carriageway (D2AP)</i>	<i>3 lane Motorway (D3M)</i>
			Mean time to failure MTTF (years)		
0	Do nothing (0)	0-0	—	—	—
1	Patching (1)	0-1	2.5	3	3.5
2	Surface dressing (2)	1-2	1.5	2	2.5
3	Resurfacing (3)	2-3	2.5	3	3.5
4	Overlay (4)	3-4	1.5	2	2.5

Examining Table 5-18, pavement condition state 0 represents the best pavement condition, which requires no maintenance. For pavement condition state  $i$ , the corresponding maintenance action  $i$  can be applied to restore pavement to the initial condition. However, it also allows the pavement to deteriorate to the worse condition state  $i+1$ , where the remedial action would be maintenance action  $i+1$ . The mean time to failure for pavement condition state  $i$  is evaluated as:



$$MTTF_{k,i} = MTTF_{k,0\_1} + \dots + MTTF_{k,i-1\_i} \quad (5-24)$$

Where  $MTTF_{k,i-1\_i}$  represents the transition time from pavement condition state  $i-1$  to pavement condition state  $i$  for road section  $k$ . The values of mean time to failure for each pavement condition state are assumed in this chapter, which can be calibrated when historical data are available. The proposed optimisation method is applied to optimise the maintenance actions for road sections on an example network, which is described in Section 6.4.

### **5.5 Maintenance Planning by Balancing between Maintenance Costs and Road User Costs**

The roadwork performed on the network may result in severe traffic congestion and travel delays to road users. To minimise the additional road user costs, highways agencies usually restrict pavement maintenance activities to hours of off-peak traffic and nighttime. The adoption of nighttime maintenance shift is the best option in terms of traffic congestion and travel delays resulted in from daytime lane closures. However, nighttime operations have a greater potential for more severe accidents due to hazards of working at nighttime, and they cost more to highways agencies. When the additional road user costs are acceptable, daytime maintenance shift would be more preferable to highways agencies; as the safety of road workers and the quality of maintenance work are both enhanced. As a consequence, it is of vital importance to optimise the maintenance arrangements so as to make a balance between road user costs and maintenance costs. Several simplifying assumptions made in minimising the total cost spent on the network are listed below:

1. The same work rates for each maintenance action during both daytime and nighttime.

2. The unit maintenance cost at nighttime is twice as that at daytime. The unit maintenance cost for each maintenance action at daytime is shown in Table 5-9.
3. For a typical day, 7:00-23:00 is defined as daytime, while the rest is nighttime.
4. The minimum maintenance duration is 2 hours, and the maximum maintenance break is set to 2 hours.
5. The maximum number of maintenance time slots for each working day is defined as 2.
6. After the implementation of maintenance activity, the road condition returns to normal.
7. When the maintenance work is not performed consistently, the set up time and cost for worksite need to be taken into account which are described in Equations 5-18 and 5-19.

In terms of the assumptions listed above, the decision variables for each road section include start time of maintenance,  $st$ , maintenance duration 1,  $md_1$ , maintenance break,  $mb$ , and maintenance duration 2,  $md_2$ , for each working day. When the investigated road link belongs to single carriageway, the green splits for the opposing directions during maintenance are also optimised.

The constraints for these variables are formulated as:

$$st_e \leq st \leq st_l$$

$$md_{\min} \leq md_1 \leq md_{\max}$$

$$md_{\min} \leq md_2 \leq md_{\max}$$

$$mb_{\min} \leq mb \leq mb_{\max}$$

The first constraint restricts the start time of maintenance action for each day. The second and third constraints limit the duration for each maintenance time slot. The

fourth constraint is the threshold of the maintenance break, which aims to avoid heavy traffic at peak time. As for single carriageway, more constraints are considered:

$$g_{\min}^s \leq g_{s_1} \leq g_{\max}^s$$

$$g_{\min}^s \leq g_{s_2} \leq g_{\max}^s$$

These two constraints are used to define the range of green phases for each direction on the single carriageway during maintenance.  $g_{s_i}$  represents the green split for one direction during the maintenance duration  $i$ , the green split for the other direction is evaluated as  $(100-g_{s_i})\%$ -amber phase. Other constraints can be considered according to the requirements of highways agencies. In this thesis, a single-objective GA is utilised to optimise the decision variables for road sections on a network.

Based on the variables listed above, the number of working days for each road link under maintenance is calculated as:

$$wd = \frac{MD}{md_1 - 1 + md_2 - 1} \quad (5-25)$$

Where  $MD$  represents the desired maintenance duration.  $md_1-1$  is the time devoted to pavement maintenance by subtracting the maintenance set up time from the maintenance duration 1.  $wd$  is rounded to the adjacent integer number when it is a decimal number. As a result, the total cost spent on the network during the whole maintenance period is evaluated as:

$$C_T = wd \times C[st, md1, mb, md2] \quad (5-26)$$

Where  $C[st, md1, mb, md2]$  represents the total cost spent at each working day under the specific maintenance arrangements  $st$ ,  $md1$ ,  $mb$ , and  $md2$ . Green splits are taken into account when single carriageway is studied. The maintenance arrangement with the least  $C_T$  is recognised as the optimal one with the objective of minimising road user cost. The optimisation of maintenance arrangements for road sections on an example network is illustrated in Section 6.5.

## 5.6 Summary

For the network level pavement optimisation model, pavement age gain model is deployed to evaluate and forecast pavement condition, and NSGA-II is used to perform pavement optimisation. Moreover, each road section group is employed as a decision making unit for the optimisation model, from which the road sections are given the same treatment. In this manner, less computational effort is required.

The developed pavement optimisation model is applied to a simple road network. A set of GA parameters are compared to find the optimal combination that led to the best solution set. With the increase of simulation generations, the optimal Pareto frontier generated is approaching the real Pareto frontier. Afterwards, the resulting optimal Pareto frontier is normalised so as to establish the balance between the involved objectives. As it takes 15 minutes for NSGA-II to reach converge and achieve the optimal Pareto frontier for the example road network, the processing time for a more complex road network will be significantly increased.

Considering the stochastic nature of pavement deterioration, the deterministic pavement age gain model is transformed to a probabilistic one using the normal distribution. Some existing approaches are provided to generate random normal variables, among which central limit theorem is employed in the probabilistic model. The example road network is also investigated using the probabilistic pavement age gain model, the results indicated that the average probabilistic optimal Pareto frontier follows the same trend as the deterministic optimal Pareto frontier, whereas a little higher than the deterministic one.

When the optimal M&R strategy is obtained for a specific network, the total cost, i.e. maintenance cost and user cost, is minimised. At the project level, the arrangements of maintenance worksites can be optimised separately and simultaneously at the project

level, such as lane closure options, start time of the maintenance and traffic control measures, so as to relieve the traffic congestion occurred in the network. The case study showed that the disruptions caused to road users by maintenance activities can be reduced significantly by applying better maintenance arrangements. The optimal arrangement is to start the work at night in order to minimise the maintenance cost and user cost.

In addition, NTFM is utilised to evaluate road user costs spent on the network under different conditions. Afterwards, the resulting road user costs and the corresponding maintenance costs can be used to optimise the maintenance actions for road sections on a network in the long-term. Also the maintenance arrangements for road sections can be optimised using GAs so as to make a trade-off between maintenance costs and road user costs.

## 6 Model Application to the Loughborough- Nottingham Network

### 6.1 Loughborough-Nottingham Highway Network

Loughborough-Nottingham highway network consists of trunk roads and certain parts of the regional rural and urban road network. An overview of the highway network is shown in Figure 6-1. The central framework of the highway network is composed of trunk roads, which is recognised as the boundary of the highway network and highlighted in blue. The M1 motorway (*H-G-F*) serves regional and inter-regional traffic that is supposed to carry a large amount of traffic. To the south, Loughborough acts as a source hub for traffic entering and exiting it. The main routes that connect to Loughborough include the A512 (*A-H*), A6 (*A-J*) and A60 (*A-B*). Nottingham performs the same functionality as Loughborough with A52 (*F-D* and *C-D*) and some other trunk roads connected. There are three main routes from Loughborough to Nottingham, which are A60-A52 (*A-B-C-D*), A6-A453-A52 (*A-J-G-K-C-D*) and A512-M1-A52 (*A-H-G-F-D*).



Figure 6-1: Loughborough-Nottingham highway network

Except for a few one-way streets, all the other road links in the Loughborough-Nottingham network were modelled in two directions. The Loughborough-Nottingham network is simplified as the one in Figure 6-2. The black symbols, e.g. M1 and A52, denote the road class and road number of the road links; the green symbols represent the type and ID of junctions, for instance, the “diverge” junction in Loughborough is named *D2*; the red symbols are real local places, i.e. Ruddington. The traffic condition of a road link is also distinguished by colours: road link in black means no traffic delay occurred; road link in yellow indicates that only a few vehicles, under the threshold of 100 vehicles, are delayed; when there are more than 100 vehicles queuing on the link is highlighted in red. The two situations that suffer traffic congestion are termed minor queues and major queues, respectively.

For simplicity, only trunk roads and the roads that connect to major junctions are retained in the network, which resulted in a significant reduction of data requirement. Those rural and urban roads neglected are represented in the traffic model as minor

traffic inputs and outputs for each corresponding major link. There are 47 junctions modelled in the network, including 8 roundabouts (*R*), 3 signalized roundabouts (*SR*), 5 diverge junctions (*D*), 3 merge junctions (*MG*), 12 off-ramps and 12 on-ramps (both denoted by *S*), 2 signalized T-junctions (*ST*), and 2 signalized intersections (*SI*). The nodes on the border of the highway network are recognized as source nodes or buffer zones, which have the task of both receiving flow from the network and pushing it into the highway network. Also, they are assigned to infinite flow capacity so as to avoid queue propagation.





## 6.2 Model Calibration

The model calibration procedure enables the whole network to represent the traffic conditions with sufficient accuracy. There are two main parts involved; the calibration of the highway network and the application of real traffic data.

### 6.2.1 Network Calibration

There are 43 links in the modelled highway network, including Motorway, A road, B road, and some urban and rural roads. The flow capacity for each link and for each arm of the junctions is determined as defined in [58], and the capacitance for each link is obtained in terms of Equation 3-63, where the average headway is defined as 0.007 km, including vehicle length and the gap between two adjacent vehicles.

### 6.2.2 Available Traffic Flow Data

The available information in this highway network against which the model could be calibrated was a set of traffic counts which had been collected at various locations on trunk roads and at various junctions over the recent few years. Those traffic data are obtained from the Highways Agency and Nottingham County Council, and applied to the traffic flow model in the form of two-way hourly traffic flow. For instance, the northbound traffic flow data for *SR2* in Ruddington is depicted in Table 6-1, the two rows that are highlighted in red are not the real data obtained from Nottingham County Council, which is derived by linear extrapolation in terms of the two adjacent sets of traffic data.

Table 6-1: Northbound traffic flow data for SI2

<i>Northbound</i>	<i>traffic flow (pcu/hr)</i>				<i>proportion</i>		
	Left	Ahead	Right	Total	Left	Ahead	Right
07:00	82	710	38	830	0.098795	0.855422	0.045783
08:00	116	846	35	997	0.116349	0.848546	0.035105
09:00	102	595	48	745	0.136913	0.798658	0.06443
10:00	92	474	43	609	0.151067	0.778325	0.070608
11:00	82	354	38	474	0.172996	0.746835	0.080169
12:00	133	448	71	652	0.203988	0.687117	0.108896
13:00	121	401	49	571	0.211909	0.702277	0.085814
14:00	118	410	66	594	0.198653	0.690236	0.111111
15:00	116	419	84	619	0.187399	0.676898	0.135703
16:00	137	763	128	1028	0.133268	0.742218	0.124514
17:00	135	846	156	1137	0.118734	0.744063	0.137203
18:00	76	435	46	557	0.136445	0.780969	0.082585
19:00	47	218	32	297	0.158249	0.734007	0.107744
20:00	28	134	22	184	0.152174	0.728261	0.119565
21:00	24	114	18	156	0.153846	0.730769	0.115385
22:00	12	83	24	119	0.10084	0.697479	0.201681

However, some rural junctions in the network are not monitored, including *R4*, *R5*, *R6*, *R7* and *R8*; the information for them was estimated. Unfortunately, no information on the length of queue formed on individual links was available, so the performance of the simulation in modelling these could not be directly validated. However, the queues formed in the network were identified at several junctions.

### 6.2.3 Movement of Traffic Flow

This traffic model is supposed to identify the travel behaviour of two-way traffic flow through the highway network. The traffic flows analysed in this model are categorised into two groups; traffic from Loughborough to Nottingham and traffic from Nottingham to Loughborough. The sequence of the evaluation of the junctions in this network is

determined as: *SR1-SR2-SR3-R6-R7-R8-R2-R5-MG1-D5-MG3-D6-MG2-R4-R1-SI2-R3-SII-ST2-D2-ST1*, the on-ramps and off-ramps are evaluated at the same time as the junctions that they are connected to. After the modelling of traffic flow through the highway network, queue propagation is introduced to simulate the congestion through the network.

### **6.3 The Evaluation of Traffic Condition**

#### **6.3.1 Highway Network Performance under Normal Conditions**

The focus of this study is to predict the outflow and queue length for each junction/link in the highway network under normal conditions, and to identify the weak links/junctions that experienced severe traffic congestion. There are 16, one hour, time steps used to model the highway network, which represents the modelling duration from 7:00 am to 11:00 pm per day. On the basis of the results obtained from modelling the whole network, we found that *R1*, *R6* and *R7* suffered traffic congestion during the morning and afternoon peak periods, while other junctions can accommodate their entering flows without disturbance. The traffic conditions for the whole network over a day are explained in the following section and illustrated graphically in Figure 6-3 to Figure 6-11.

As shown in Figure 6-3, a relatively small amount of traffic presented in the highway network at the period 7:00-8:00, vehicles thereby could move smoothly without travel delay.

At the second period 8:00-9:00, there are more commuters in the network heading to work/school locations, which resulted in heavy traffic congestion at some large junctions. For instance, *R6*, recognised as a priority junction, assigns more transition capacity to the major routes that experienced higher traffic. According to the geometries

of the routes for *R6*, the two arms on A52 were identified as major routes, while the arm on the B5010 contributes the least traffic input for *R6*. At the 8:00-9:00 am time period, some vehicles on the B5010, destined for *R6*, have to queue. This is because the amount of the traffic passing by this arm exceeds the flow capacity of B5010, as they cannot find a gap to enter in *R6*. In addition, traffic congestion occurred in *R1* and *R7* due to the same reason as *R6*.

From 9:00-10:00, though there is less traffic than the last time period, it is still high. The traffic congestion in *R7* dissipated, since the entering traffic flow for *R7* is less than the corresponding exiting flow capacity; whereas, the clearance of the queue in *R7* further increased entering flow for *R6*, which made the traffic condition in *R6* more severe. As for *R1*, there is still a large amount of traffic intending to join and the capacitance on link *SI2-R1* is merely 125 vehicles, so it was filled and led to further queue propagation to upstream links which are the entry arms for *SI2* except for *R1-SI2*, since it has no contribution to the traffic in the reverse direction of the identical link.

From 10:00-11:00, *R6* and *R1* still suffered traffic congestion. For *R6*, the queue on the B5010 is cleared, while a new queue formed on the A6007. This is because there are stored vehicles in the queue on the B5010 at previous time step, the incoming traffic flow on the B5010 outweighs the amount of incoming traffic on the A6007. Therefore, some vehicles on the A6007 are delayed when competing traffic flow exceeds its flow capacity. The traffic condition in *R1* remains since the entering traffic flow for *R1* is still high.

For 11:00-12:00, because entering traffic for each junction is much less than the corresponding flow capacity, the queues formed at previous steps clear in the shortest possible time. In the following five hours entering traffic flow for the whole network keeps a low level, traffic thereby could leave the network without disturbance.

During the 17:00-20:00 period, recognised as another peak time during a day, the highway network also suffered severe traffic congestion as that during 8:00-10:00 due to the same reasons.

At the rest of the simulation steps, the highway network could deliver people and goods without interference since most of the commuters had arrived at their destinations, illustrated in Figure 6-11.

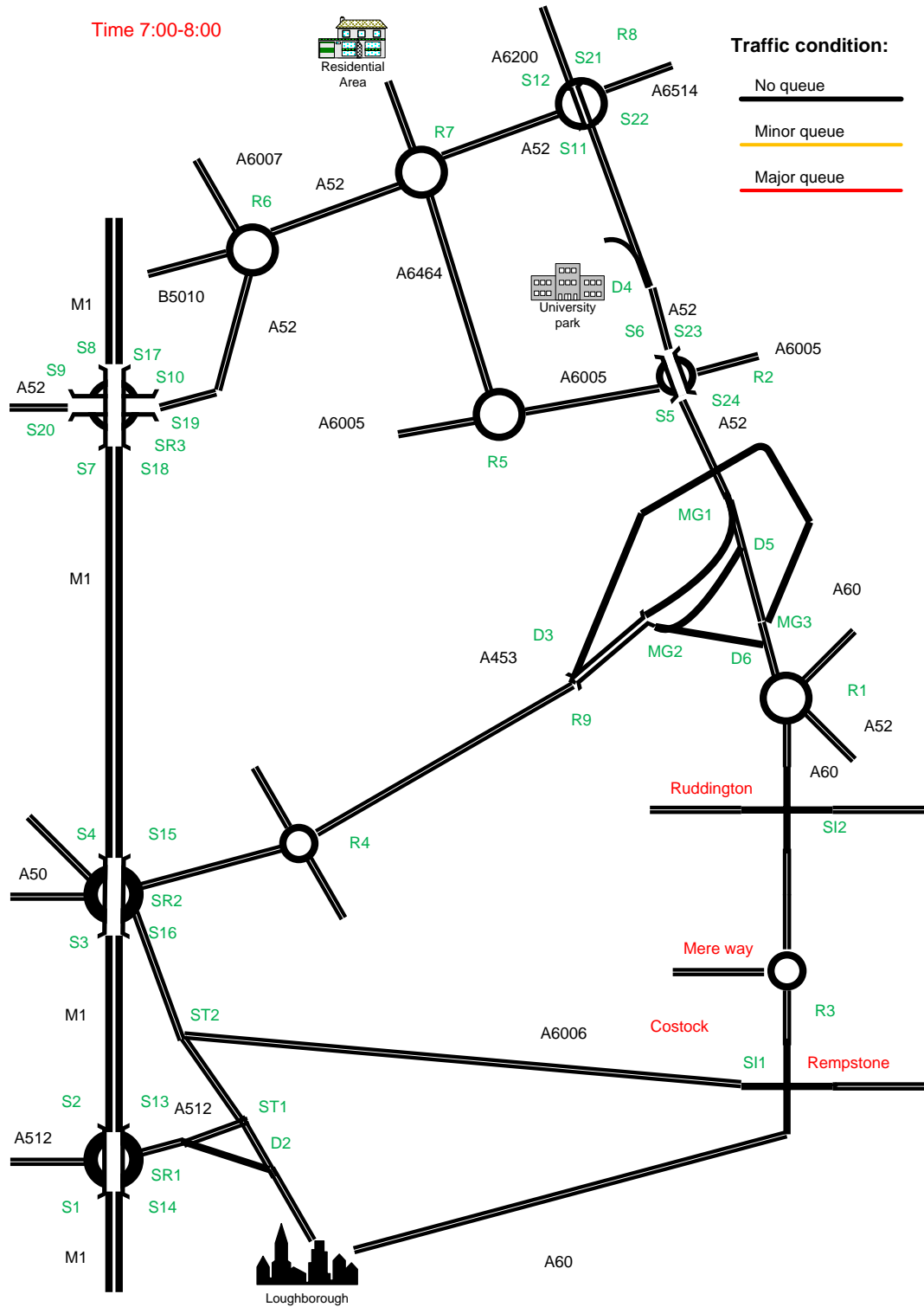


Figure 6-3: Traffic condition at 7:00-8:00

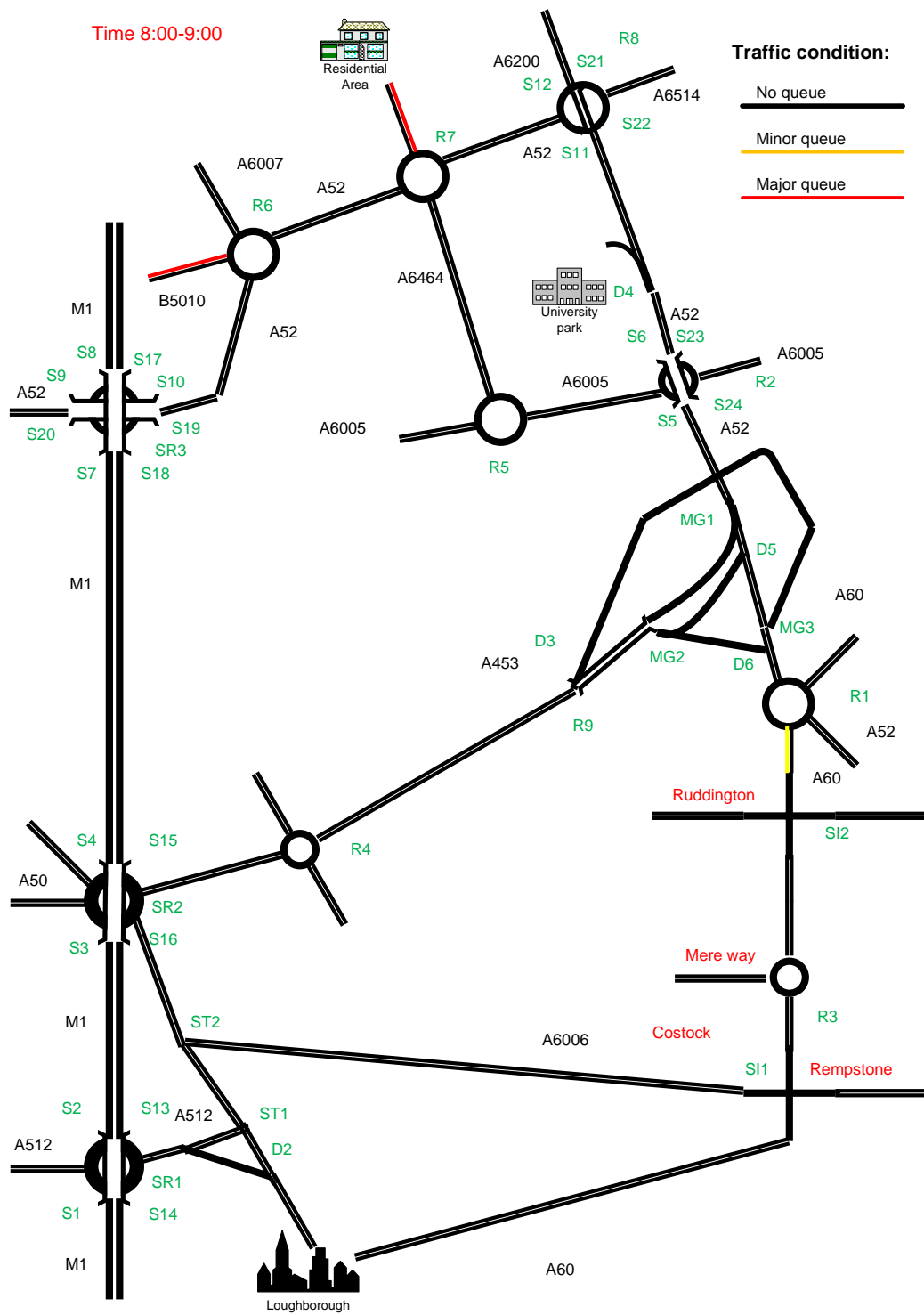


Figure 6-4: Traffic condition at 8:00-9:00



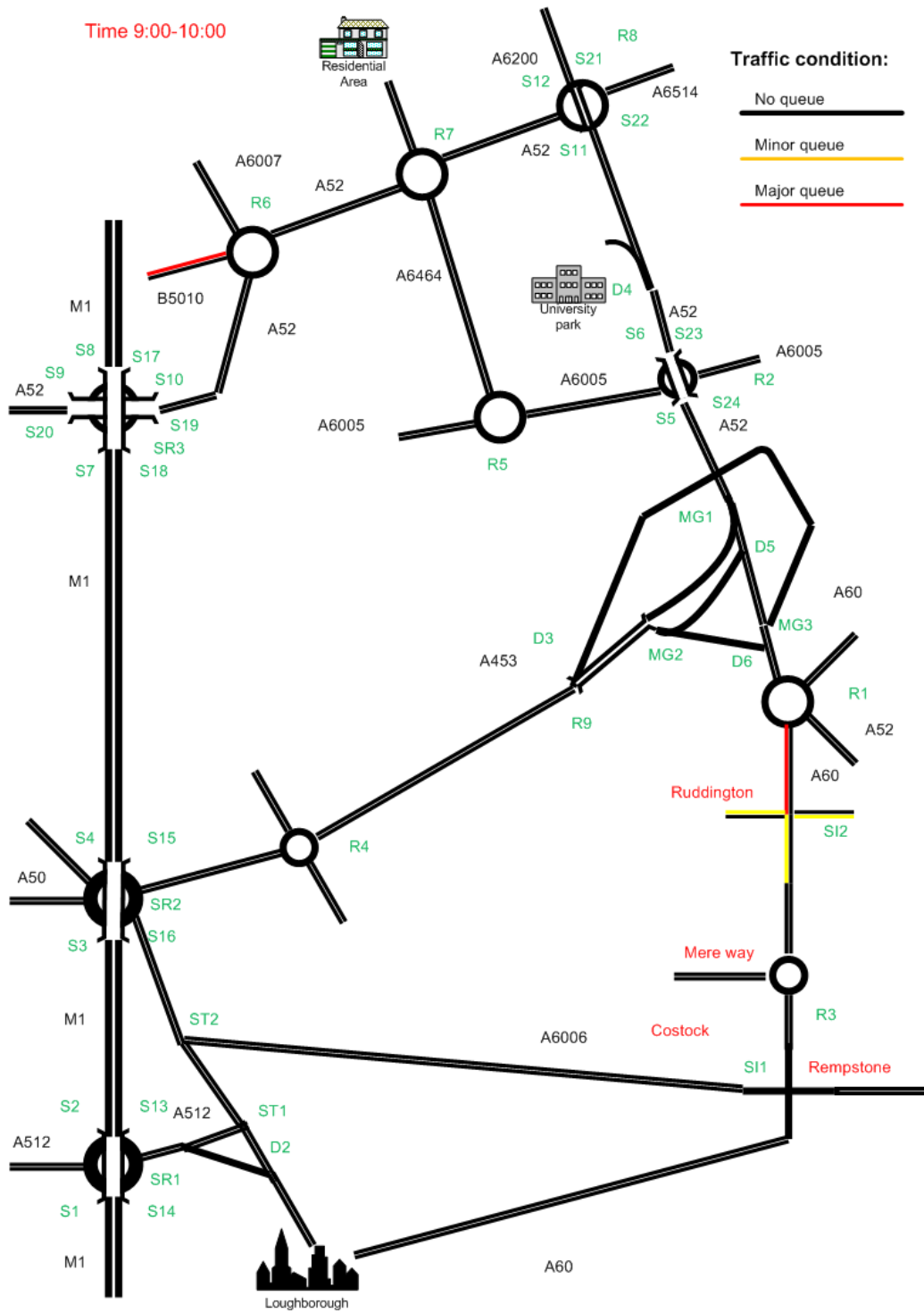


Figure 6-5: Traffic condition at 9:00-10:00

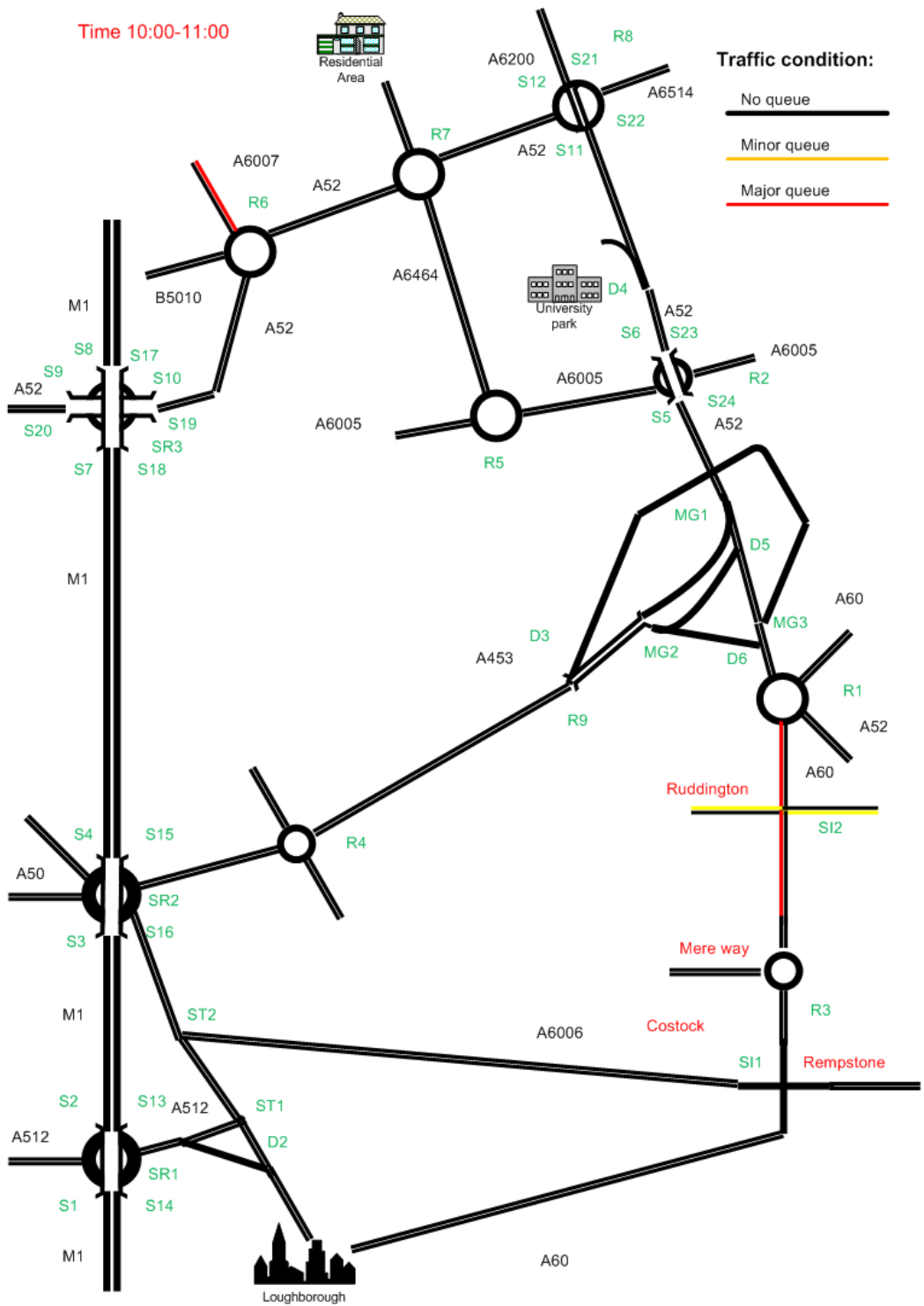


Figure 6-6: Traffic condition at 10:00-11:00

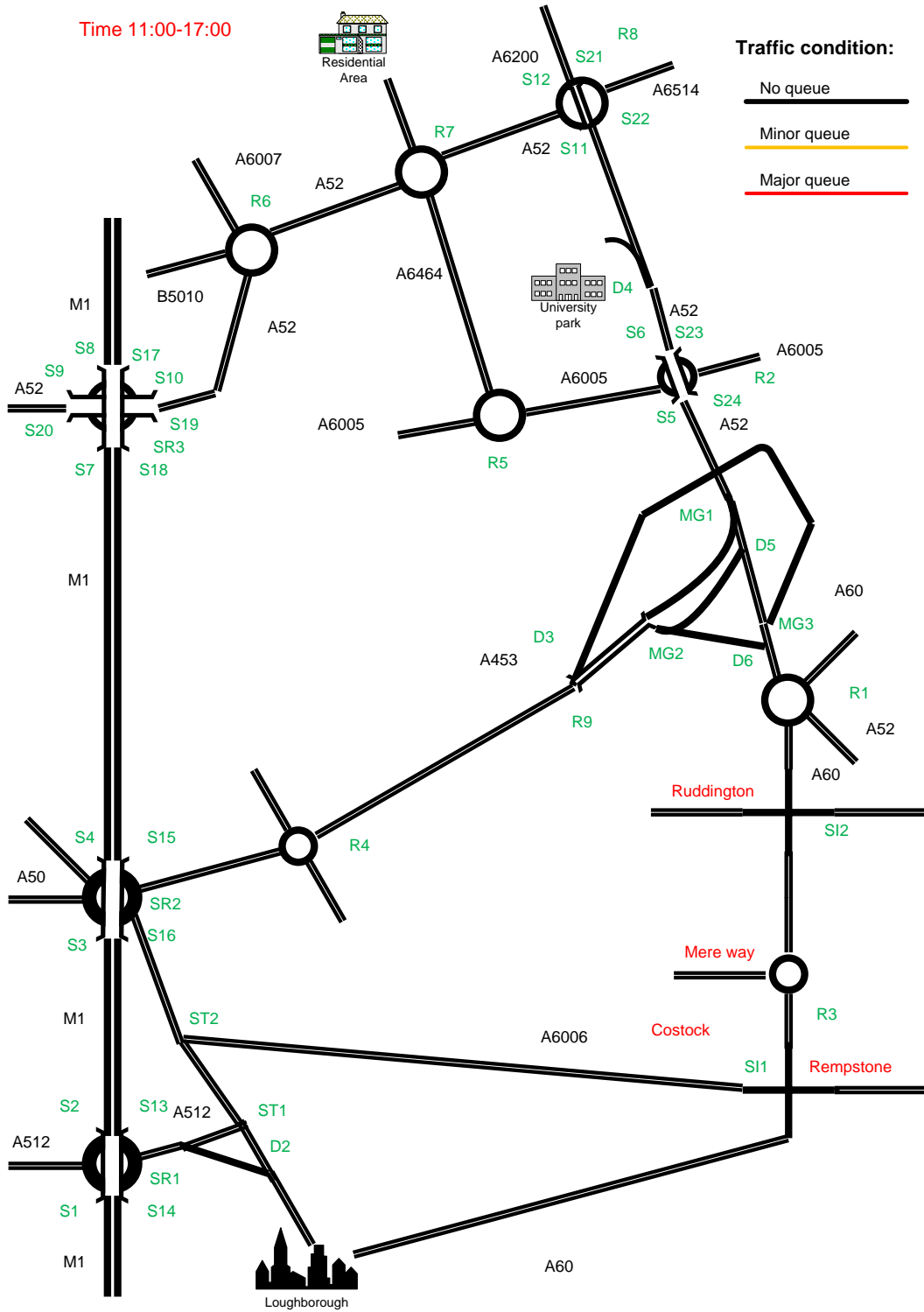
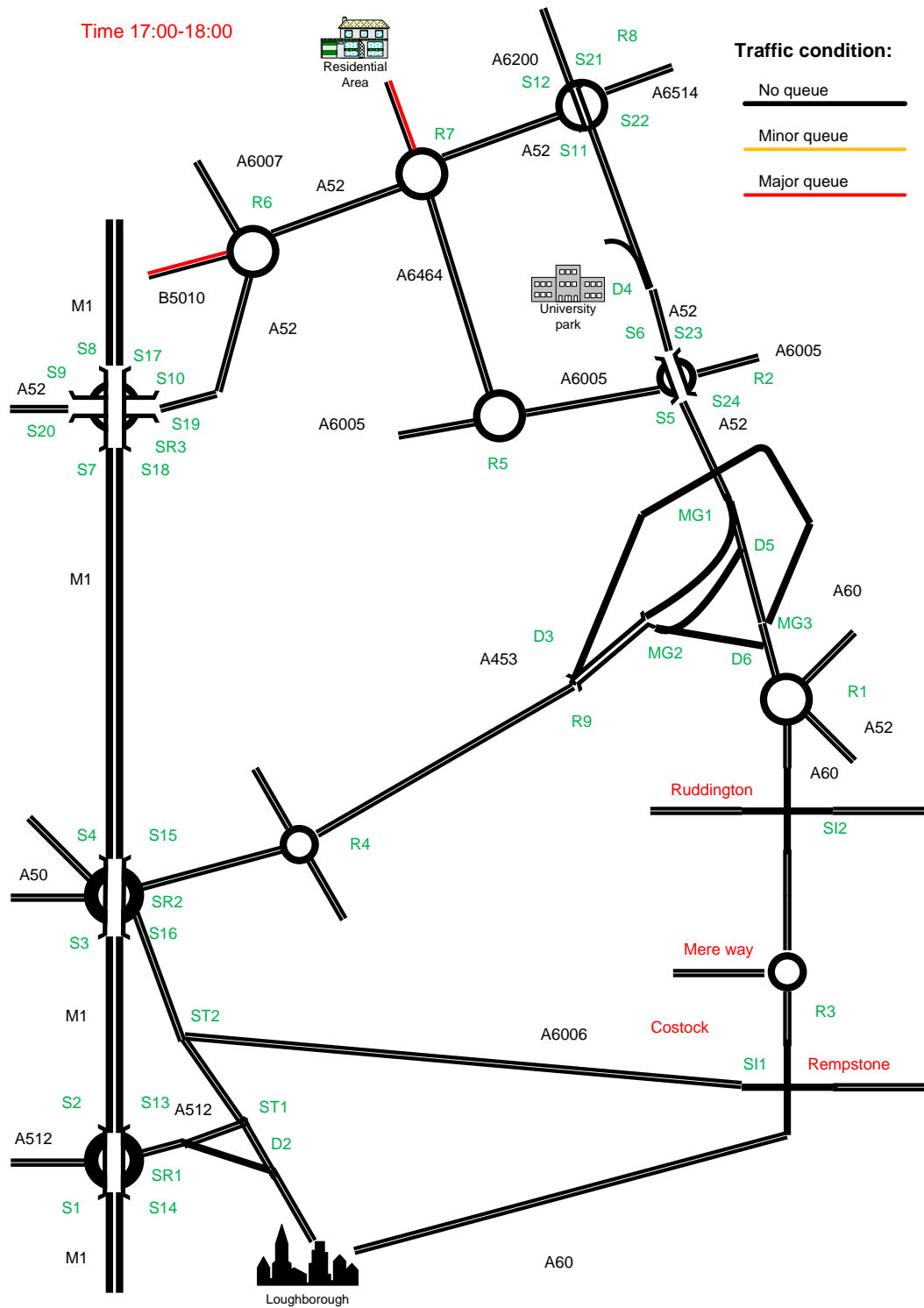


Figure 6-7: Traffic condition at 11:00-17:00



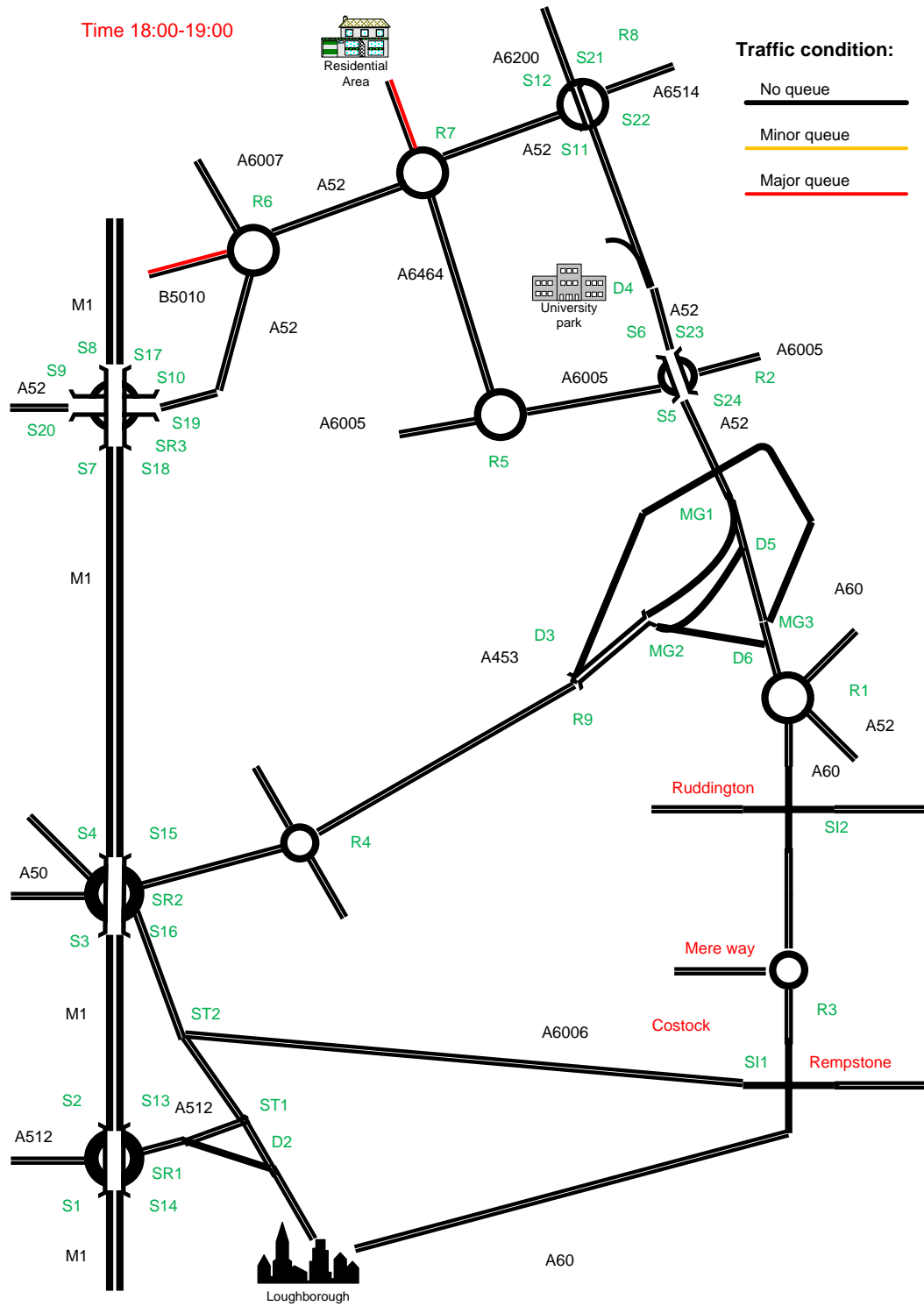


Figure 6-9: Traffic condition at 18:00-19:00

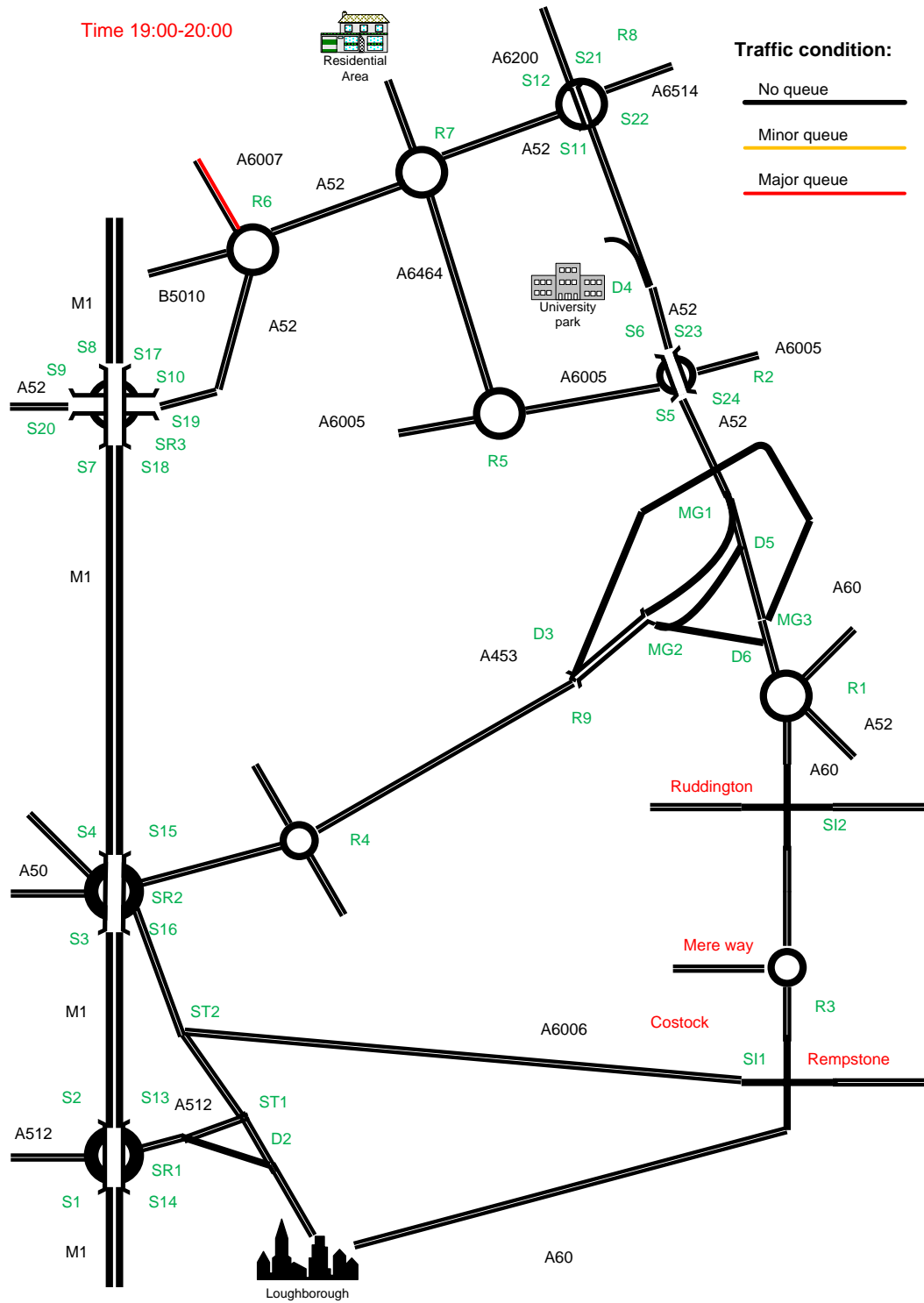


Figure 6-10: Traffic condition at 19:00-20:00

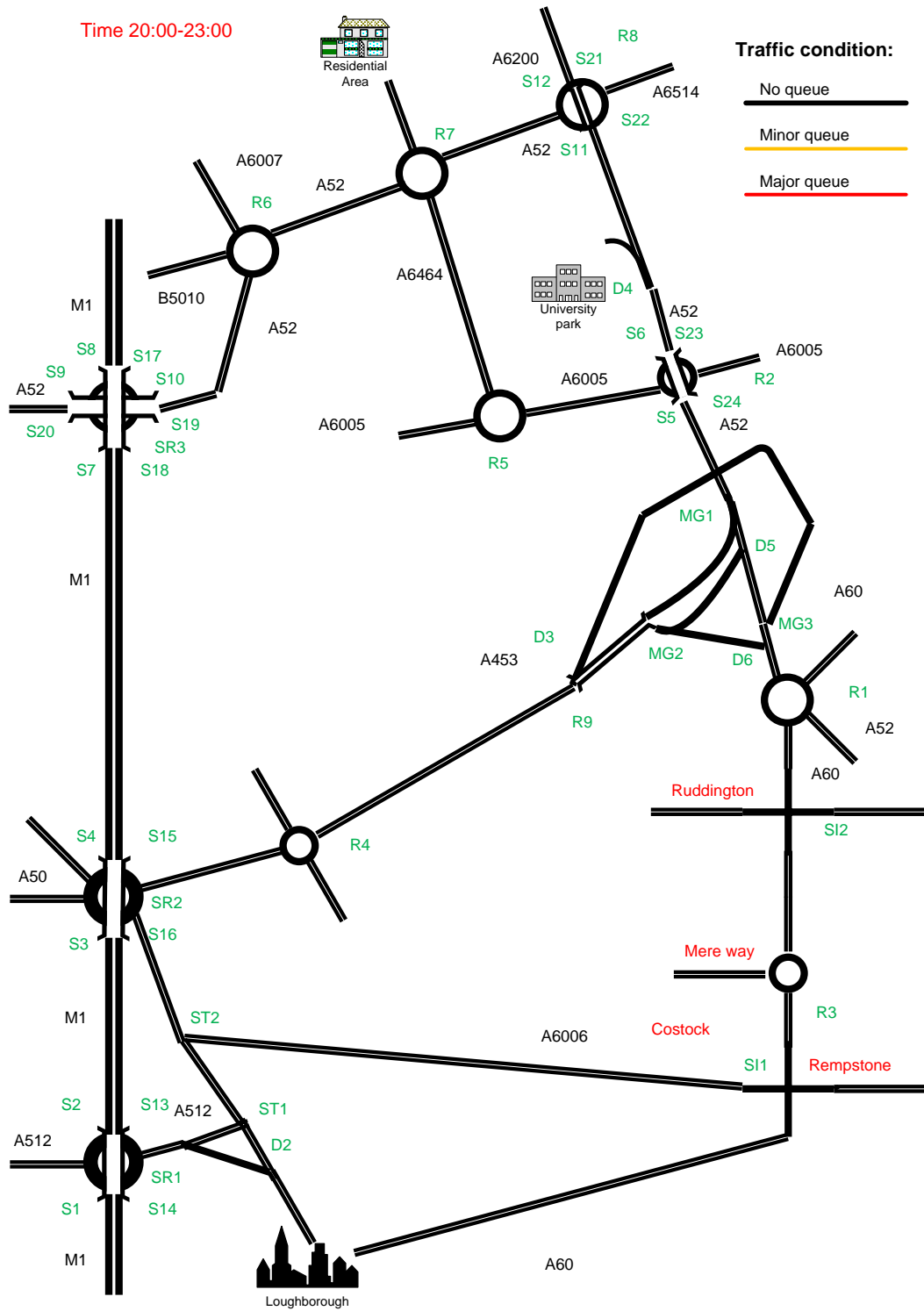


Figure 6-11: Traffic condition at 20:00-23:00

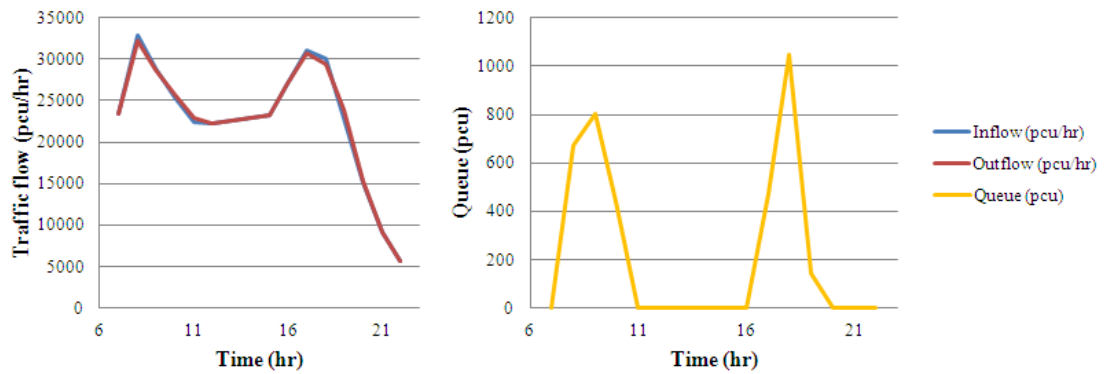


Figure 6-12: Traffic condition throughout a typical day

Additionally, the numerical performance of the network is expressed in Figure 6-12. The yellow line termed queue is calculated as the sum of all the queues formed in the network at each time step. It can be seen that the highway network experienced heavy traffic congestion during 8:00 to 11:00 and 17:00 to 20:00. This is because the rush hour periods fall into these two time slots. Moreover, at some times during the day, i.e. 10:00-12:00, the exiting traffic flow capacity is higher than the entering traffic flow and the queues formed at previous time periods clear.

### 6.3.2 The Implementation of Roadworks

In addition to the normal junctions, roadwork nodes (*RN*) have been introduced as appropriate to investigate the influence of maintenance activities to the traffic in the network. As such road links provided in the network depend on the local road classes, they need to be addressed differently according to the road class of the link it belongs to. In this case study three *RNs* are added to the network, which are *RNDI* (dual 2 lane carriageway), *RNSI* (single carriageway) and *RNMI* (motorway), as displayed in Figure 6-2.



### 6.3.2.1 Dual 2 Lane Carriageway

To investigate the effect of maintenance activities on a dual 2 lane carriageway, *RND1* is assigned to link *MG2-R4*. There are three cases considered for *RND1* in terms of the parameters defined in Section 3.2.8, as shown in Table 6-2.

Table 6-2: Road geometries for RND1

RND1	Case1 (No maintenance)	Case 2	Case 3
$L$ (km)	2.8	2.8	2.8
$x$ (km)	–	0.7	0.7
$m1$ (km)	–	0.5	1
$m2$ (km)	–	1.6	1.1
$n1$	2	1	1
$n2$	2	2	2

The layout of link *MG2-R4* is described in Figure 6-13, one service lane is left at the worksite *RND1* for conveying traffic owing to one lane closure.

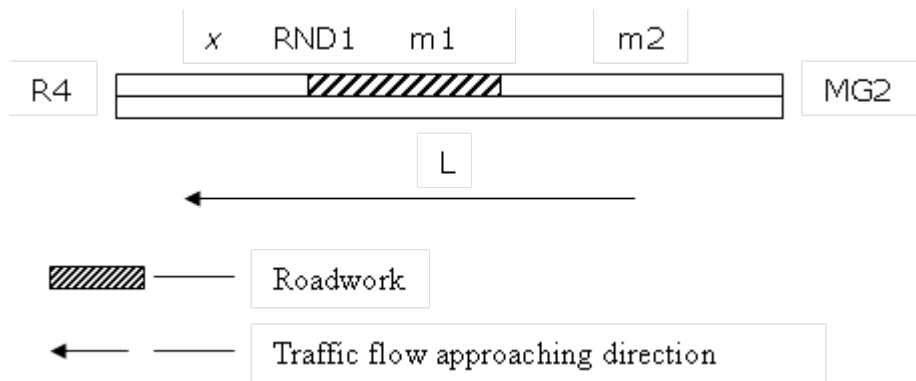


Figure 6-13: Layout of link *MG2-R4*

In Case 1, there is no maintenance implemented in the road link, which represents the normal road condition. While for Case 2 and Case 3, the locations for the worksite are the same, the only difference is that the length of the worksite in Case 3 is twice as it is in Case 2, which represents a choice in the maintenance scheduling. Under each road condition the flow capacity for the lane is assumed to be constant. The traffic

characteristics of the network through the day in each case are illustrated in Figure 6-14, the first one refers to the transportability of highway network and the other one presents the travel delay occurred in the network.

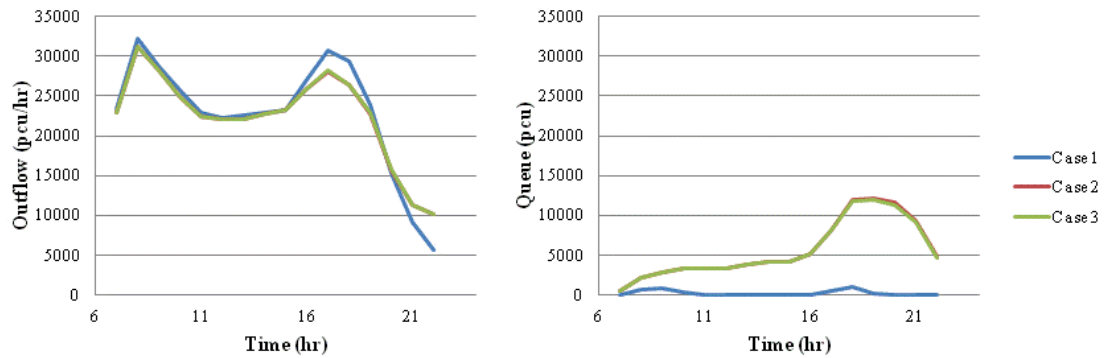


Figure 6-14: Highway network with RND1: (a) Exiting traffic flow through the network; (b) Total queue within the network

As only one lane open for the worksite on link *MG2-R4* for Case 2 and Case 3, more traffic are delayed in queues than in Case 1, especially during the afternoon peak period. For Cases 2 and 3, the road capacity on *RND1* is significantly reduced from 3600 pcu/hr to 1530 pcu/hr. As a consequence, the queue that formed on link *MG2-R4* exceeded its corresponding road capacitance and propagated back to its source links, which caused queues and delays on other directional flows at source junctions resulting in excessive traffic congestion. Also, it is reasonable that the traffic conditions in Case 2 and Case 3 almost coincide since the difference of the link capacitance for Case 2 and Case 3 is quite small, the discrepancy in the resulting queue propagation is very slight. As a consequence, the variation of the length of worksite on a dual 2 lane carriageway appears to have limited influence to network performance.

### 6.3.2.2 Single Carriageway

For this type of road, two opposing traffic flows have to share the only lane in service when maintenance operation is implemented. Road link *SR2-R4* is analysed and the

worksite on it is denoted by *RNS1* as shown in Figure 6-15, the traffic flows from *SR2* and *R4* can only pass through the worksite during the corresponding green splits, as for the rest time traffic flows have to wait at the back of the worksite so as to leave gaps for the opposing flows.

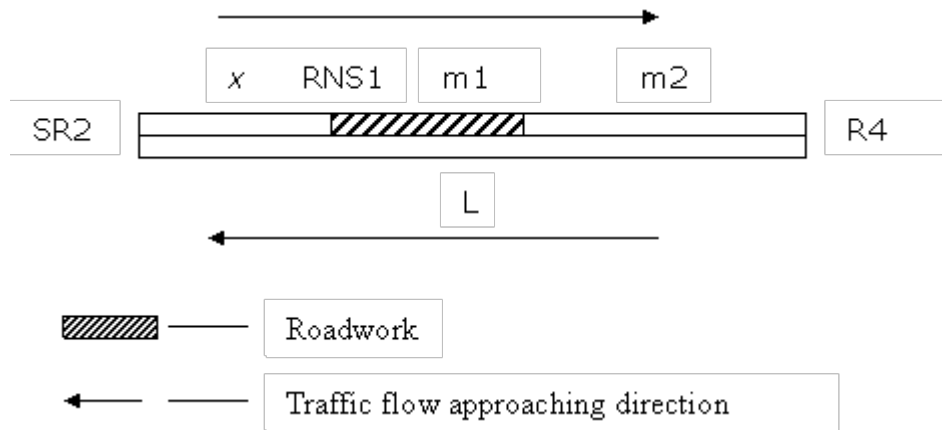


Figure 6-15: Layout of link SR2-R4

Three maintenance scenarios are investigated at *RNS1*, described in Table 6-3.

Table 6-3: Road geometries for RNS1

RNS1	Case1 (No maintenance)	Case 2	Case 3
$L (km)$	7	7	7
$x (km)$	–	3.2	3.2
$m1(km)$	–	0.6	1.2
$m2(km)$	–	3.2	2.6
$n1$	1	1	1
$n2$	1	1	1
$G$	–	45%	40%
$A_m$	–	10%	20%

When *RNS1* is undergoing repair, signal control is employed to assign gaps for each directional flow. For Case 2, the green splits for both directions are defined as 45% with the amber phase taking 10% of the whole cycle. While for Case 3, the length of worksite is the double of it in Case 2. As amber phase increases with the length of

worksite, the amber phase in Case 3 is set to 20% while green splits for both directions are 40%.

The traffic conditions of the whole network under normal conditions and that with the repair at *RNS1* are compared in Figure 6-16,

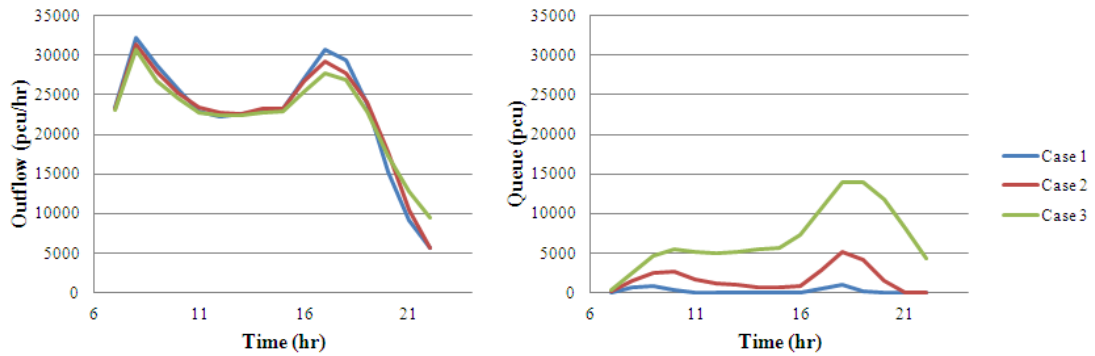


Figure 6-16: Highway network with RNS1: (a) Exiting traffic flow through the network; (b) Total queue within the network

Figure 6-16 reveals that the travel delay occurring in the network during maintenance is much greater than when the network is operating under normal conditions. This is contributed to the fact that both the two directional traffic flows are restricted by the flow capacity of the worksite and signal control. In addition, it shows that the traffic delay experienced in the network in Case 3 is more than twice of that in Case 2, as the residual flow capacity for each direction at *RNS1* in Case 3 is evaluated as 720 pcu/hr which is much smaller than the residual flow capacity in Case 2, 810 pcu/hr, also the link capacitance on link *SR2-R4* is further reduced in Case 3, resulting more vehicles spilled back to source links that led to much severer traffic congestion at upstream junctions.

### 6.3.2.3 Motorway

Link *S4-S7* is a typical 3 lane motorway, on which maintenance is assumed to be performed at *RNM1*. The parameters of this link are defined as:  $x=4$  km,  $m1=1.12$  km,  $m2=3.2$  km,  $L=8.32$  km. The layout of this link is shown in Figure 6-17,

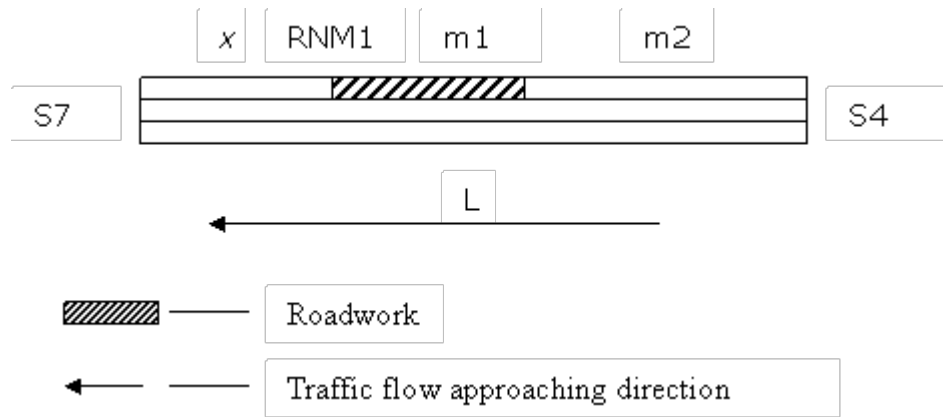


Figure 6-17: The layout of link S4-S7

There are two cases considered for *RNMI* that are depicted in Table 6-4, Case 1 represents the normal road condition. The performance of overall traffic network under each condition, illustrated in Figure 6-18, is considered. The two situations investigate the effects of different number of lane closures for the repair work.

Table 6-4: Road geometries for *RNMI*

RNMI	Road conditions	
	Case 1	Case 2
Lanes in service	3	2

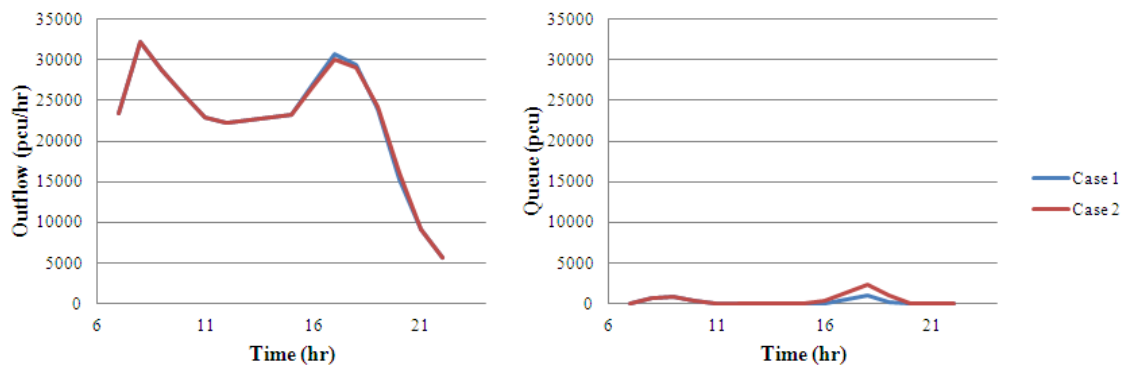


Figure 6-18: Highway network with *RNMI*: (a) Exiting traffic flow through the network; (b) Total queue within the network

According to Figure 6-18, travel delays only occurred in rush hour. Due to maintenance work, the site capacity on *RNMI* is decreased to 3400 pcu/hr from 6000 pcu/hr. However, Case 1 and Case 2 possess similar travel behaviour; this is because the

residual flow capacity of Motorway is relatively large with comparison to entering traffic flow at most of the time during the day, which could accommodate the loss of one service lane. During the period 16:00-20:00, the aggregate queue length of the network with maintenance is slightly higher than that under normal conditions; it means that additional delay has occurred due to insufficient transportability on *RNMI*.

Moreover, when the entering flow through link *S4-S7* is much smaller than the corresponding flow capacity, two service lanes can be repaired simultaneously to minimise maintenance duration. Nevertheless, in this case the entering flow on link *S4-S7* is greater than the residual flow capacity on the worksite during the afternoon peak, which would result in severe traffic congestion when a two lane closure is applied; as a result, two lane closure becomes infeasible for *RNMI*.

### 6.3.3 Discussion

Based on the results in Sections 6.3.1 and 6.3.2, it is reasonable that travel delays experienced on the network during maintenance are more severe than that under normal conditions since the flow capacity of the link undergoing repair is reduced due to lane closure. Moreover, it was found that the network with maintenance at *RNSI* suffered the most severe traffic congestion. As the flow capacity on *RNSI* is not only affected by the reduction of service lanes but also restricted by signal control. The delayed traffic filled link *SR2-R4* and spilled back to upstream junctions, i.e. *SR2* and *R4*, which further led to the blockage of motorway and normal road. For the network with repair work at *RNI*, the flow capacity on *RNI* is also significantly affected by the maintenance work, and the amount by which the entry flow exceeded the corresponding flow capacity is relatively large. As expected it resulted in the formation of queue on the current link which propagated back to the source links. For the network with maintenance at *RNMI*, owing to the length and width of link *SR2-SR3*, the delayed traffic is relatively small and the

queue only occurred leading up to *RNMI*. More than one lane can be maintained concurrently when the entering flow is relatively small in comparison to the residual flow capacity.

From the above results it confirms that maintenance work has a great impact on the network performance, especially when the normal flow along a road is very much higher than the worksite capacity, i.e. at morning and afternoon peaks. Thus, when scheduling pavement maintenance, the expected flow rates and flow capacities should be considered to determine how it can be achieved to cause less disruption. Furthermore, the length of the worksite with shuttle working plays an important role in traffic control, as it is correlated with the amber phase which influences the flow capacity on the worksite directly. By comparing the resulting travel delay incurred by maintenance work, the best maintenance options can be selected.

#### **6.4 Maintenance Planning for Road Sections in the Long-term**

In this section, the optimisation method developed in Section 5.4 is used to evaluate road user costs and maintenance costs over the pavement life cycle, and the assumptions listed in Section 5.4 are also adopted. To investigate the effect of road user cost on the selection of maintenance actions, some road links on the network are analysed separately in the following sections, which are road links *MG2-R4*, *SR2-R4*, *S4-S7*, *SII-ST2*, and *S2-S3*. The life cycle for each road link is defined as 60 years.

##### **6.4.1 Dual Carriageway (MG2-R4)**

For each maintenance action that conducted on road link *MG2-R4*, one lane closure is applied. The length of link *MG2-R4* is defined as 1.75 miles, and the duration and cost for each maintenance action are calculated according to Table 5-9. For instance, the duration for maintenance action 1 is calculated as:

$$1.75 \times 1.6 \times 2 \times 18 = 101 \text{hrs}$$

As road link *MG2-R4* is a 2 lane carriageway, both service lanes need to be maintained.

The results for all the maintenance actions are described in Table 6-5:

Table 6-5: Maintenance costs and durations for road link *MG2-R4*

<i>Road link MG2-R4</i>					
Maintenance action	0	1	2	3	4
Length (km)	2.8	2.8	2.8	2.8	2.8
Duration (hrs)	0	101	168	471	773
Cost (£'000s)	0	84	112	462	686
Mean time to failure (years)	—	3	5	8	10
Number of maintenance actions	—	20	12	8	6

In addition, the mean time to failure and the number of maintenance actions for each maintenance action are also provided in Table 6-5.

The next step is to calculate the additional road user costs when road link *MG2-R4* is undergoing each maintenance action. For maintenance action 1, the maintenance duration would be 4 days and 5 hrs, which starts from 7:00 on the first day. Thus, NTFM is employed to model the traffic conditions on the network over 5 days. The traffic conditions on road link *MG2-R4* through the day under maintenance action 1 are shown in Table 6-6:



Table 6-6: Traffic conditions on road link MG2-R4 under maintenance action 1

<i>Time</i>	<i>Inflow</i> ( <i>pcu/hr</i> )	<i>Flow</i> <i>capacity</i> ( <i>pcu/hr</i> )	<i>Outflow</i> ( <i>pcu/hr</i> )	<i>Queue (pcu)</i>	<i>Queue</i> <i>propagation</i> ( <i>pcu</i> )
7:00	2033	1530	1530	457	46
8:00	2603	1530	1530	457	1073
9:00	3123	1530	1530	457	1593
10:00	3785	1530	1530	457	2255
11:00	3900	1530	1530	457	2370
12:00	4116	1530	1530	457	2586
13:00	3955	1530	1530	457	2425
14:00	4030	1530	1530	457	2500
15:00	4234	1530	1530	457	2704
16:00	4103	1530	1530	457	2573
17:00	4235	1530	1530	457	2705
18:00	4441	1530	1530	457	2911
19:00	4157	1530	1530	457	2627
20:00	3941	1530	1530	457	2411
21:00	3721	1530	1530	457	2191
22:00	3598	1530	1530	457	2068
23:00	2750	1530	1530	457	1220
24:00	1251	1530	1530	178	0
1:00	31	1530	209	0	0
2:00-7:00	31	1530	31	0	0

Table 6-6 indicates that road link *MG2-R4* experienced severe traffic congestion at the daytime. When maintenance action is performed on road link *MG2-R4*, its flow capacity is decreased to 1530 pcu/hr from 3600 pcu/hr using Equation 3-46. Thus, the inflows at daytime are greater than the residual flow capacity during maintenance. Also it is assumed that the link capacitance on link *MG2-R4* during maintenance is a definite value, 457 pcu. When the delayed vehicles filled the link they propagated back to their source links, which makes the traffic condition even worse. Those vehicles attempted to

enter into the link at the next time step. The traffic conditions on the network under both normal and maintenance conditions are compared in Figure 6-14.

According to the simulation results, the travel delay taken place on the network under normal conditions is obtained as 2,785.5 hrs, which is contributed from the congestion links on the network throughout the day that are illustrated in Figure 6-3 to Figure 6-11. The queue length for each congested link through the day is shown in Table 6-7:

Table 6-7: Queue length for the congested links on the network

<i>Link</i>	<i>R6</i>		<i>R7</i>		<i>R1</i>		<i>SI2</i>		<i>Total queue (pcu)</i>
	B5010	A6007	Residential Area	SI2- R1	Ruddington	R3- SI2	Westbound arm		
8:00	320	0	264	89	0	0	0	673	
9:00	620	0	0	125	17	39	3	804	
10:00	0	162	0	125	44	102	8	441	
11:00	0	0	0	0	0	0	0	0	
17:00	240	0	231	0	0	0	0	471	
18:00	540	0	506	0	0	0	0	1046	
19:00	0	143	0	0	0	0	0	143	
20:00	0	0	0	0	0	0	0	0	

and the travel delays spent on the congested links are calculated in terms of Equation 3-59, described as:

Table 6-8: Travel delay for the congested links on the network

<i>Link</i>	<i>R6</i>	<i>R7</i>	<i>R1</i>	<i>SI2</i>				<i>Total</i>
	B5010	A6007	Residential Area	SI2-R1	Ruddington	R3-SI2	Westbound arm	<i>delay (hrs)</i>
8:00	160	0	132	44.5	0	0	0	336.5
9:00	470	0	30.1	107	8.5	19.5	1.5	636.6
10:00	165.6	81	0	125	30.5	70.6	5.5	478.2
11:00	0	10.4	0	4.8	0.8	1.14	0.06	17.2
17:00	120	0	115.5	0	0	0	0	235.5
18:00	390	0	368.5	0	0	0	0	758.5
19:00	135	71.5	108.5	0	0	0	0	315.0
20:00	0	8.01	0	0	0	0	0	8.01

Afterwards, it is transformed to road user cost by multiplying time value, £15.38, obtained as £42,840.8. As for the network under maintenance 1, the travel delay for each of the first four days is achieved as 89,617.65 hrs, which corresponds to £1,378,319.5. While for the 5<sup>th</sup> day, one lane closure is applied during 7:00 and 12:00, the resulting travel delay is obtained as 15,685.6 hrs that costs £241,244.5. The additional road user cost is computed as:

$$U_{MG2-R4,1} = 1378319.5 \times 4 + 241244.5 - 42840.8 \times 5 = 5540318.5$$

In this manner, the additional costs to road users that resulted from the rest maintenance actions are evaluated, which are depicted in Table 6-9:

Table 6-9: Additional road user costs due to maintenance actions on MG2-R4

<i>Road link MG2-R4</i>				
Maintenance action	1	2	3	4
Maintenance duration	4 days and 5 hrs	7 days	19 days and 15 hrs	32 days and 5 hrs
Additional travel delay (hrs)	360,228.75	607,825.12	1,732,744.85	2,791,529.23
Additional road user cost (£'000s)	5,540	9,348	26,650	42,934

By combining Table 6-5 and Table 6-9, the total cost for each maintenance action throughout the life cycle is described in Table 6-10:

Table 6-10: Total cost spent on the network when road link MG2-R4 is repaired

<i>Road link MG2-R4</i>				
Maintenance action	1	2	3	4
Number of maintenance actions	20	12	8	6
Maintenance cost (£'000s)	84	112	462	686
Additional road user cost (£'000s)	5,540	9,348	26,650	42,934
Total maintenance cost (£'000s)	1,680	1,344	3,696	4,116
Total addition road user cost (£'000s)	110,800	112,176	213,200	257,604
Total cost (£'000s)	112,480	113,520	216,896	261,720

Table 6-10 shows that the additional costs to road users due to maintenance activities far exceed the corresponding maintenance costs. Thus, it is realistic to take into account road user costs when optimising M&R strategy for a road network. Also maintenance action 1 is recognised the most optimal one which led into the least total cost over the pavement life cycle.

#### 6.4.2 Single Carriageway (SR2-R4)

As road link *SR2-R4* is a single carriageway road, one lane closure and shuttle working is applied to implement maintenance activities. The geometries of link *SR2-R4* are shown in Table 6-11:

Table 6-11: Maintenance costs and durations for road link SR2-R4

<i>Road link SR2-R4</i>					
Maintenance action	0	1	2	3	4
Length (km)	7	7	7	7	7
G	—	45%	45%	45%	45%
Am	—	10%	10%	10%	10%
Duration (hrs)	0	168	336	672	1,344
Cost (£'000s)	0	105	140	420	665
Mean time to failure (years)	—	2.5	4	6.5	8
Number of maintenance actions	—	24	15	10	8

As the investigated road link is a single carriageway, the maintenance action performed would influence the traffic flows from both directions. There is no traffic congestion taken place under normal conditions. However, during maintenance the flow capacity for each direction on this link would be reduced to 810 pcu/hr using Equation 3-45. The road user cost spent on the network under normal conditions is also £42,840.8. By using the method provided in Section 6.4.1, the additional delays and costs to road users that incurred by each maintenance action are evaluated, which are depicted in Table 6-12:

Table 6-12: Additional road user costs due to maintenance actions on SR2-R4

<i>Road link SR2-R4</i>				
Maintenance action	1	2	3	4
Maintenance duration	7 days	14 days	28 days	56 days
Additional travel delay (hrs)	159,025.4	318,050.7	636,101.5	1,272,203
Additional road user cost (£'000s)	2,446	4,892	9,783	19,566

Based on Table 6-11 and Table 6-12, the total costs resulted from the maintenance actions that applied on road link *SR2-R4* over the life cycle are illustrated in Table 6-13:

Table 6-13: Total cost spent on the network when road link SR2-R4 is repaired

<i>Road link SR2-R4</i>				
Maintenance action	1	2	3	4
Number of maintenance actions	24	15	10	8
Maintenance cost (£'000s)	105	140	420	665
Additional road user cost (£'000s)	2,446	4,892	9,783	19,566
Total maintenance cost (£'000s)	2,520	2,100	4,200	5,320
Total addition road user cost (£'000s)	58,704	73,380	97,830	156,528
Total cost (£'000s)	61,224	75,480	102,030	161,848

Table 6-13 also indicates that the road user costs resulted from maintenance actions are significantly greater than the corresponding maintenance costs. Moreover, maintenance action 1 is determined as the optimal one as the least travel delay is experienced when maintenance action 1 is performed on road link *SR2-R4*.

#### 6.4.3 Motorway (S4-S7)

Road link *S4-S7* belongs to 3 lane motorway, both one lane closure and two lane closure can be used to perform maintenance activities. In this chapter, only one lane closure is utilised. The geometries of link *S4-S7* are shown in Table 6-14:

Table 6-14: Maintenance costs and durations for road link S4-S7

<i>Road link S4-S7</i>					
Maintenance action	0	1	2	3	4
Length (km)	8.32	8.32	8.32	8.32	8.32
Duration (hrs)	0	400	600	2,000	3,300
Cost (£'000s)	0	333.25	416.75	2,250	3,375
Mean time to failure (years)	—	3.5	6	9.5	12
Number of maintenance actions	—	18	10	7	5

As stated above, the road user cost spent on the network under normal conditions is calculated as £42,840.8. Based on the method described in Section 6.4.1, the additional travel delays and road user costs that resulted from each maintenance action are evaluated, which are depicted in Table 6-15:

Table 6-15: Additional road user costs due to maintenance actions on S4-S7

<i>Road link S4-S7</i>				
Maintenance action	1	2	3	4
Maintenance duration	16 days and 16 hrs	25 days	83 days and 8 hrs	137 days and 12hrs
Additional travel delay (hrs)	55,199.51	81,175.75	269,530.49	446995.32
Additional road user cost (£'000s)	849	1,248	4,145	6,875

Based on the parameters listed in Table 6-14 and Table 6-15, the total costs spent on the network due to the maintenance activities on link S4-S7 are shown in Table 6-16:

Table 6-16: Total cost spent on the network when road link S4-S7 is repaired

<i>Road link S4-S7</i>				
Maintenance action	1	2	3	4
Number of maintenance actions	18	10	7	5
Maintenance cost (£'000s)	333.25	416.75	2,250	3,375
Additional road user cost (£'000s)	849	1248	4,145	6,875
Total maintenance cost (£'000s)	5,999	4,168	15,750	16,875
Total addition road user cost (£'000s)	15,282	12,480	29,015	34,375
Total cost (£'000s)	21,281	16,648	44,765	51,250

It is identified that maintenance action 2 is the best selection for road link S4-S7 with both the minimum maintenance cost and additional user cost, which is resulted from the

less frequency of this maintenance action during the life cycle comparing to maintenance action 1.

#### 6.4.4 Dual Carriageway (SI1-ST2)

Road link *SI1-ST2* is a 2 lane carriageway, the length of which is 5.25 mile. As mentioned before, one lane closure is utilised to perform maintenance activities. The geometries of link *SI1-ST2* are described in Table 6-17:

Table 6-17: Maintenance costs and durations for road link SI1-ST2

<i>Road link SI1-ST2</i>					
Maintenance action	0	1	2	3	4
Length (miles)	8.4	8.4	8.4	8.4	8.4
Duration (hrs)	0	303	504	1,412	2,319
Cost (£'000s)	0	252	336	1,386	2,058
Mean time to failure (years)	—	3	5	8	10
Number of maintenance actions	—	20	12	8	6

The road user cost spent on the network under normal conditions is determined as £42840.8. In terms of the method provided in Section 6.4.1, the additional delays and costs to road users caused by each maintenance action are evaluated, which are depicted in Table 6-18:

Table 6-18: Additional road user costs due to maintenance actions on SI1-ST2

<i>Road link SI1-ST2</i>				
Maintenance action	1	2	3	4
Maintenance duration	12 days and 15 hrs	21 days	58 days and 20 hrs	96 days and 15 hrs
Additional travel delay (hrs)	0	0	0	0
Additional road user cost (£'000s)	0	0	0	0



Based on the results of NTFM, the road user costs spent on the network under maintenance scenarios are also calculated as £42,840.8. It is due to the fact that the residual flow capacity on link *SII-ST2* during maintenance is still larger than the entering flow. As a consequence, no additional cost incurred to road users. The total costs spent on the network due to the maintenance actions performed on link *SII-ST2* are illustrated in Table 6-19:

Table 6-19: Total cost spent on the network when road link *SII-ST2* is repaired

<i>Road link SII-ST2</i>				
Maintenance action	1	2	3	4
Number of maintenance actions	20	12	8	6
Maintenance cost (£'000s)	252	336	1,386	2,058
Additional road user cost (£'000s)	0	0	0	0
Total maintenance cost (£'000s)	5,040	4,032	11,088	12,348
Total addition road user cost (£'000s)	0	0	0	0
Total cost (£'000s)	5,040	4,032	11,088	12,348

As no additional road user costs resulted from maintenance actions. The optimal maintenance action is only influenced by the maintenance cost. As a result, maintenance action 2 is selected as the optimal one.

#### 6.4.5 Motorway (S2-S3)

Road link S2-S3 is 3 lane motorway on which one lane closure is utilised to perform maintenance actions. The geometries of link S2-S3 are shown in Table 6-20:

Table 6-20: Maintenance costs and durations for road link S2-S3

<i>Road link S2-S3</i>					
Maintenance action	0	1	2	3	4
Length (km)	3.2	3.2	3.2	3.2	3.2
Duration (hrs)	0	154	231	768	1,268
Cost (£'000s)	0	128	160	864	1,296
Mean time to failure (years)	—	3.5	6	9.5	12
Number of maintenance actions	—	18	10	7	5

The road user cost spent on the network under normal conditions is calculated as £42840.8. According to the method described in Section 6.4.1, the additional travel delays and road user costs spent on the network under various maintenance activities are achieved, illustrated in Table 6-21:

Table 6-21: Additional road user costs due to maintenance actions on S2-S3

<i>Road link S2-S3</i>				
Maintenance action	1	2	3	4
Maintenance duration	6 days and 10 hrs	9 days and 15 hrs	32 days	52 days and 20 hrs
Additional travel delay (hrs)	17,034.14	28,126.8	90,005.75	149,072.04
Additional road user cost (£'000s)	262	433	1,384	2,293

Based on the results provided in Table 6-20 and Table 6-21, the total costs spent on the network under different maintenance actions are compared in Table 6-22:

Table 6-22: Total cost spent on the network when road link S2-S3 is repaired

<i>Road link S2-S3</i>				
Maintenance action	1	2	3	4
Number of maintenance actions	18	10	7	5
Maintenance cost (£'000s)	128	160	864	1,296
Additional road user cost (£'000s)	262	433	1,384	2,293
Total maintenance cost (£'000s)	2,304	1,600	6,048	6,480
Total addition road user cost (£'000s)	4,716	4,330	9,688	11,465
Total cost (£'000s)	7,020	5,930	15,736	17,945

In comparison to the optimal maintenance action for road link *S4-S7*, the same maintenance action is achieved. This is because both two road links are 3 lane motorway and similar traffic conditions are experienced.

#### 6.4.6 Discussion

The results obtained in this section show that the less extensive maintenance actions, i.e. patching and surface dressing, are more favourable to pavement engineers. These maintenance actions result in fewer disruptions to road users as they can be performed in a relatively short time duration. Also, they require less funding and resources during the pavement life cycle comparing to major maintenance actions.

We also found that the additional cost to road users is much greater than the corresponding maintenance cost when the residual flow capacity of the investigated road link under maintenance is less than the inflow. In this case, the selection of the optimal maintenance action is dominated by the additional road user cost. There were situations when maintenance cost plays the main role in optimising maintenance actions, as shown in Section 6.4.4. Moreover, the mean time to failure for each pavement

condition state is crucial to the optimisation of M&R strategy, as it is directly proportional to both maintenance cost and road user cost.

When pavement deterioration is taken into account, a comparison between different maintenance actions is possible and total costs for the whole life cycle can be calculated. Total costs are calculated in terms of maintenance cost and road user cost, which are obtained using the NTFM. As expected, maintenance actions that can be performed more frequently and need shorter periods of maintenance are cheaper.

## **6.5 Maintenance Planning for the Road Sections in the Short-term**

In terms of the results provided in Section 6.4, we found that the additional road user costs that resulted from maintenance activities can be much greater than the corresponding maintenance costs. The assumptions proposed in Section 5.5 are applied to the maintenance planning in this section.

To investigate the impact of maintenance arrangements on road users, a dual carriageway road, link *MG2-R4*, is studied on which maintenance action 1, patching, is implemented. Also, one lane closure is applied to conduct maintenance activities. Three possible maintenance arrangements are assigned to road link *MG2-R4*: the first arrangement is to start the maintenance at 7:00 am until it is finished, which has already been studied in Section 6.4.1; the second one is to perform maintenance activities at off-peak time, i.e. 11:00-16:00 and 19:00-23:00; the third one is to performance maintenance at nighttime, 23:00-7:00 on the next day.

### **6.5.1 Possible Maintenance Arrangements on Dual Carriageway (MG2-R4)**

Three maintenance arrangements are considered in this section. For maintenance arrangement 1, the total maintenance duration required by maintenance action 1 would

be 101 hours, and the total cost is £84,000. The evaluation of total cost spent on the network under maintenance arrangement 1 is described in Section 6.4.1.

For maintenance arrangement 2, the maintenance duration for each day is 9 hrs, which requires 15 working days with the consideration of maintenance set up time according to Equation 5-19. In the first 14 working days, maintenance activities are performed at 11:00-16:00 and 19:00-23:00. As for the last working day, maintenance slot falls to 11:00-15:00. For each of the first 14 working days, the travel delay and cost to road users are calculated as 7,768.54 hours and £119,480 by using NTFM. For the 15<sup>th</sup> working day, the travel delay and road user cost are obtained as 5,098 hours and £78,407.6. In terms of the method proposed in Section 6.4.1, the additional travel delay and cost to road users on the network under maintenance arrangement 2 are evaluated as 72,075 hours and £1,108,516.

For maintenance arrangement 3, nighttime operation is applied. The maintenance work starts from 23:00 to 7:00 on the next day, which requires 15 days to complete the project in terms of Equation 5-19. For the first 14 days, the road section is maintained at 23:00-7:00. While for the last day, maintenance work is performed at 23:00-3:00. Afterwards, NTFM is applied to model the traffic characteristics on the network under maintenance arrangement 3. The results show that the same traffic conditions are experienced on road link *MG2-R4* as that on the network under normal conditions. This is because maintenance activities are performed at nighttime when limited traffic existed on the network.

The results obtained on the network under three maintenance arrangements are compared in Table 6-23:

Table 6-23: Maintenance cost and road user cost spent on the network under various maintenance arrangements when maintenance action 1 is performed

<i>Maintenance arrangement</i>	<i>1</i>	<i>2</i>	<i>3</i>
Additional travel delay (hrs)	360,228.75	72,075	0
Maintenance cost (£'000s)	84	113	183
Additional road user cost (£'000s)	5,540	1,109	0
Total cost (£'000s)	5,624	1,222	183

Examining Table 6-23, we found that maintenance arrangement 1 resulted in the most additional cost to road users, as maintenance activities are performed at peak time during the daytime. However, maintenance arrangement 1 requires the least funding from highway agencies. As for the maintenance arrangement 2, the road is only maintained at off-peak time, which caused less disruption to road users. The third maintenance arrangement is the most costly option for highways agencies which has no influence to road users. The maintenance cost for maintenance arrangement 2 is calculated as £ 113,000 owing to the maintenance set up cost in line to Equation 5-18. Therefore, in this section using the results obtained from NTFM it is indicated how maintenance work can be arranged in the daytime so as to minimise the additional road user cost. Different types of road will be considered, in order to demonstrate effective maintenance planning depending on the road type.

#### 6.5.2 Optimisation of Maintenance Arrangements on Dual Carriageway (MG2-R4)

This section investigates the maintenance arrangements on road link *MG2-R4* at daytime with the aim of minimising road user costs. As in Section 6.5.1, maintenance action 1 is performed. Another assumption is that the maintenance arrangements for each day during the whole maintenance period are the same. In terms of the assumptions listed in Section 5.5, the decision variables for road link *MG2-R4* include start time of maintenance, *st*, maintenance duration 1, *md<sub>1</sub>*, maintenance break, *mb*, and

maintenance duration 2,  $md_2$ , for each working day. The constraints for these variables are formulated as:

$$7 \leq st \leq 12$$

$$2 \leq md_1 \leq 6$$

$$2 \leq md_2 \leq 6$$

$$1 \leq mb \leq 2$$

$$st + md_1 + mb + md_2 \leq 23$$

The first constraint restricts the start time of the maintenance action from 7:00 to 12:00. The second and third constraints limit the duration for each maintenance time slot with the lower bound of 2 hours and upper bound of 6 hours. The fourth constraint is the threshold of maintenance break. The fourth constraint means that the maintenance work for each day must be completed by 23:00.

Based on the variables listed above, the number of working days for road link *MG2-R4* under maintenance action1 is calculated by Equation 5-25. The desired maintenance duration for road link *MG2-R4* under maintenance action 1 is evaluated as 101 hours. Afterwards, the total cost under each maintenance arrangement is achieved using Equation 5-26.

In this section, the involved decision variables are recognised as four strings of genes that formed a chromosome in the optimisation of maintenance arrangements using a binary coded single-objective GAs. The whole chromosome is represented as:

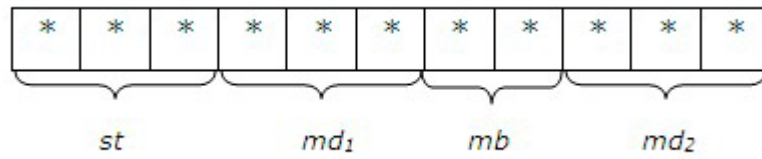


Figure 6-19: A typical genotype of chromosome for road link MG2-R4

Where “\*” represents a binary variable, i.e. 0 or 1.

In the optimisation control, the size of parent pool is selected as 100, and the crossover rate and mutation rate are adopted as 0.8 and 0.01, the maximum generation is 100. As the aim of this study is to find the traffic control that incurred the least travel delay, the objective function is defined as the total road user cost spent during the whole maintenance period, as shown in Equation 5-26. After the simulation, the best chromosome for all the generations are is described as 111-101-11-011, and then it is decoded to integer numbers where the left end of the string is recognised as the first allele. For example, the fourth string 011 is decoded as:

$$\frac{0 \times 2^0 + 1 \times 2^1 + 1 \times 2^2}{2^3 - 1} \times (6 - 2) + 2 = 5$$

According to the results obtained from GA, the optimal maintenance arrangements are listed as:

- st* 12:00
- md<sub>1</sub>* 12:00-16:00
- mb* 16:00-18:00
- md<sub>2</sub>* 18:00-23:00
- wd* 15 days

Consequently, the additional road user cost spent on the network during maintenance under the optimal maintenance arrangements is obtained as £1,009,098 when 15 working days are required, and the corresponding maintenance cost is £114,000. By



comparing to the results for the network under maintenance arrangement 2 that described in Table 6-23, the additional cost to road users is greatly reduced by optimising maintenance arrangements, and similar maintenance cost incurred.

### 6.5.3 Optimisation of Maintenance Arrangements on Single Carriageway (SR2-R4)

For road link *SR2-R4*, maintenance action 1 is also applied to restore pavement condition which requires 168 working hours. In addition to the decision variables studied in Section 6.5.2, the green splits for both directions on the road link during maintenance can also be optimised so as to convey the opposing traffic flows through the link. Consequently, six variables are considered in the optimisation of maintenance arrangements on road link *SR2-R4*, which are start time of maintenance,  $st$ , maintenance duration 1,  $md_1$ , green split for eastbound traffic during maintenance duration 1,  $gs_1$ , maintenance break,  $mb$ , maintenance duration 2,  $md_2$ , and green split for eastbound traffic during maintenance duration 1,  $gs_2$ . The green split for westbound traffic on road link *SR2-R4* during maintenance duration  $i$  is evaluated  $(100-gs_i)\%$ -amber phase. Amber phase is defined as 10% of the cycle. The constraints for the decision variables are described as:

$$7 \leq st \leq 12$$

$$2 \leq md_1 \leq 6$$

$$2 \leq md_2 \leq 6$$

$$30 \leq gs_1 \leq 60$$

$$30 \leq gs_2 \leq 60$$

$$1 \leq mb \leq 2$$

$$st + md_1 + mb + md_2 \leq 23$$

The fourth and fifth constraints restrict the green split for eastbound traffic on road link *SR2-R4* which varies from 30% to 60% of the total cycle. The rest is the same as the

constraints provided in Section 6.5.2. The chromosome for road link *SR2-R4* is formulated as:

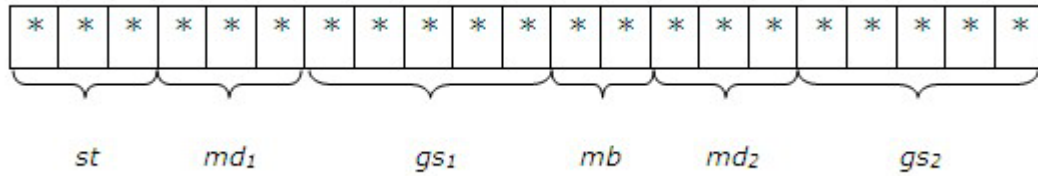


Figure 6-20: A typical genotype of chromosome for road link SR2-R4

The GA parameters are set as follows:

Population 100

Maximum generation 500

Crossover rate 0.8

Mutation rate 0.01

According to the simulation results, the optimal chromosome is evaluated as 011-111-11110-11-101-11110. Then the chromosome is decoded to integer numbers, described as:

*st* 11:00

*md<sub>1</sub>* 11:00-17:00

*gs<sub>1</sub>* 44%

*mb* 17:00-19:00

*md<sub>2</sub>* 19:00-23:00

*gs<sub>2</sub>* 44%

*wd* 21 days

Based on the optimal maintenance arrangements, 21 working days are required and the resulting additional road user cost is evaluated as £66,177.3, and the corresponding maintenance cost is £147,000. We found that the additional cost to road users is even less than the half of the maintenance cost. Thus it is acceptable for highways agencies to

perform maintenance activities at daytime under the achieved optimal maintenance arrangement.

#### 6.5.4 Discussion

This section investigates the additional travel delays and costs to road users incurred in the network under various maintenance arrangements, where start time of maintenance, the allocation of maintenance time slot, maintenance break, and traffic control measures are all taken into account. The results indicate that nighttime operation led to the least disruption to road users, and the travel delays to road users resulted from daytime operation can be greatly minimised by optimising maintenance arrangements. However, when the additional road user cost due to daytime maintenance operation is acceptable, highways agencies would prefer daytime operation in order to guarantee the safety of road workers and minimise the possibility of the occurrence of more severe accidents.

### **6.6 Summary**

This Chapter presents the network level traffic flow tool, NTFM, to calculate the traffic flows in an integrated road network. In addition to other junction sub-models, a roadwork node sub-model has been embedded in NTFM to define the geometry of a maintenance worksite and to demonstrate the effect of maintenance work on the traffic conditions in the network. Further, signal control is employed to manage the traffic in worksites with shuttle working.

A case study has been performed to investigate the traffic conditions in the Loughborough-Nottingham highway network under different road conditions. Traffic movement, queue formation and dissipation in the network are identified. The results demonstrated that the performance of the network is significantly worse when maintenance is implemented, i.e. more queues and delay times.

Also the road user costs due to maintenance can be evaluated based on NTFM which is used to optimise maintenance actions for road sections along with the corresponding maintenance cost during the pavement life cycle, where pavement deterioration process is considered. The results show that the maintenance actions that can be performed more frequently and require shorter maintenance durations are more acceptable to highways agencies.

Moreover, the maintenance arrangements for the road sections under maintenance can be optimised using GA to facilitate the traffic through the worksite based on NTFM so as to cause fewer disturbances to road users, where the trade-off between the maintenance cost and road user cost must be taken into account.

## 7 Conclusions and Future Work

### 7.1 Summary

Following an extensive literature review on the existing traffic flow models, the need for the development of a macroscopic traffic model capable of modelling an integrated network of motorway and urban roads was identified. The majority of the traffic models used for the prediction of traffic movement is only sufficient for signalized road networks, where the traffic interactions experienced at the priority junctions cannot be captured. For instance, the traffic on the minor road at a T-junction is not only characterised by the corresponding flow capacity but also limited by the incoming flow on the major road. The techniques presented for priority junctions mainly focused on investigating the traffic on single intersection and are not accurate for modelling directional flow, therefore they are not capable of identifying traffic characteristic at the network level. Considering the traffic interaction at both signalized and priority junctions, this thesis describes a network level traffic flow model (NTFM) the purpose of which is to provide a method for predicting traffic flows and travel delays through the road network.

NTFM consists of a larger number of junction types than other existing traffic models, such as signalized junctions, priority junctions, and maintenance worksite. The model produces journey times spent on the road links and travel delays spent at the junctions. To cope with two-way traffic flow in the road network, an iterative simulation method is utilised to generate the evolution of dependent traffic flows and queues. In addition a traffic rerouting strategy has been introduced to NTFM to model the real travel behaviour, i.e. adjusting pre-defined route to compensate travel delay, when traffic congestion takes place in the network.

Application examples of NTFM have been investigated. Initially relatively simple road networks were studied. The first case study demonstrated that NTFM is capable of identifying the relationship between traffic flows and capturing traffic phenomena such as queue dynamics and the second one indicated that the traffic congestion experienced in the network can be decreased or eliminated by taking alternative routes.

The development of an optimisation framework based on NSGA-II was completed so as to create a technique which identifies the best pavement maintenance scheduling with the optimal allocation of available budgets and resources, consequently the least maintenance cost is required. The NSGA-II software has proven its ability on a wide range of real engineering problems. One advantage of using GAs is the ability to achieve an optimal or near-optimal solution for any engineering problem considered rather than developing the optimisation approach in terms of the specified problem. For the optimisation model maintenance cost and total remaining pavement life have been selected to indicate the performance of M&R strategy. Limitations of budgets and resources, restrictions of the number of maintenance actions have been applied as possible constraints for the optimisation framework.

The developed optimisation framework was applied to allocate maintenance actions for simple case studies during a planning period of 20 years. This case study demonstrated the capability of the optimisation model to maintain pavement condition. The results showed that the optimal M&R strategy can reduce maintenance cost and prolong the remaining pavement life of individual road section.

In addition to the maintenance costs, the cost to road users can also be minimised by varying the maintenance arrangements at the worksite. The case study in Section 5.3.2 demonstrated that the start time of maintenance activities and traffic controls at the worksite have a substantial impact on network performance.

Moreover, NTFM has been applied to model traffic characteristics on the Loughborough-Nottingham highway network that is composed of both urban and motorway roads. Maintenance activities were defined in this network in order to evaluate effects of alternative maintenance strategies, which is realised by introducing a reduced flow rate on links of the network. Afterwards, NTFM has been applied to optimise maintenance actions for road sections on a highway network with the consideration of pavement degradation process during the pavement life cycle. As for the maintenance actions performed on the road sections, the maintenance arrangements can also be optimised using the NTFM in order to improve transportability and cause fewer disruptions to road users. The results indicated that the network efficiency is substantially influenced by undertaking maintenance work, especially during the morning and afternoon rush hour; and the travel delays to road users can be significantly reduced by optimising maintenance arrangements, where a binary single-objective GA is utilised to perform maintenance arrangement optimisation.

## **7.2 Sensitivity of Main Assumptions**

### **7.2.1 Junction model**

1. PCU (passenger car unit) is used to describe traffic flow and the number of cars on the network. Therefore, other types of vehicles are transformed to the corresponding PCU. For example, a truck is defined as 3 PCU. If vehicle types are considered in NTFM, longer headways are assigned to heavy vehicles. For instance, the headway for cars are 7 m, while for heavy vehicles headway can be defined as 8 m for each PCU.
2. Two-lane arms are used to model most of the signalised intersections. The left-hand lane is used for going straight on and turning left and the right-hand lane

for turning right. For some signalised intersections, three-lane arms are used. In this case the left-hand lane is occupied by traffic turning left, the middle lane is used for going straight on and the right-hand lane for turning right. However, the traffic going straight and turning left experiences the same green phase in the model with three lanes. Thus the two kinds of models for signalised intersections give a similar performance under the same traffic controls and the adequacy between the NTFM and a real-world network is marginally influenced by the assumption of number of lanes. The similar assumption is applied to some other types of junctions, for example, signalised T-junction and signalised roundabout.

3. For signalised intersections, the green phases for both lanes in the two-lane model are defined as 50% and 30% of the total cycle, respectively. These numbers have been derived while observing a representative junction on the road network locally. For a better representation such phases could be derived using traffic data on each junction on the network and the numerical values of network performance from the model would be different. However, the conclusions of the overall performance would, most likely, stay the same, since the difference between default green phases and the values from the real life is very small, for example, the same link would be identified as the most congested one.
4. For an urban roundabout, it is assumed that vehicles are not allowed to queue on the roundabout, while on a motorway roundabout vehicles can be delayed directly on the roundabout due to the larger size of the motorway roundabout than the urban one. Such traffic behaviour is commonly observed on the roundabouts close to a busy motorway.
5. In the on-ramp model, the conflicting flow for traffic from the slip road is determined as the flow on the left-hand lane on the major road, which is the case



in reality. In the off-ramp model, it is assumed that only the traffic leaving the major road takes the left-hand lane, since most vehicles that are carrying on other lanes on the major road. In the merge and diverge models, the traffic flows on slip roads are given the same priority as these junctions are symmetrically constructed and do not create conflicting flows on the main road. All the models described in this section model the traffic on motorways, as commonly observed on the highway networks.

6. Flow rates and turning ratios at junctions are defined at each arm of the junction and they vary with each time step, i.e. each hour, as generally observed on the network. The model is very sensitive to the flow rates and turning ratios, and such data influences the network performance directly. For example, due to different flow rates a network without any queue can turn into a heavily congested network. Therefore, the validity of this input data is very important, as discussed in validation section.

### 7.2.2 Queue model

7. A queue can occur at every node on the network, i.e. at junctions and worksites, since at these points there is a change in traffic conditions in terms of disturbance on the link. The headway is defined as 0.007 km when a queue is present, i.e. cars leave 0.002 km gap between them in a queue if the average car length is 0.005 km. With the increase of the headway the length of queue increases slightly. For example, the length of a queue with 10 vehicles is increased to 0.08 km from 0.07 km when headway is defined as 0.008km. Also, the capacitance on each link is determined in terms of the headway in the queue. As before, the effect of the length of headway is limited. For instance, the

capacitance on a link with the length of 0.56 km is decreased to 70 pcu from 80 pcu when headway is defined as 0.008 km.

### 7.2.3 Pavement deterioration model

8. For the probabilistic pavement deterioration model, normal distribution is applied to characterise the expected life for new pavement and the age gain resulting from maintenance actions. This distribution was chosen as it is one of the most commonly used probability distributions in engineering problems. Other lifetime distributions, such as polygon distribution and Weibull distribution, can also be used to describe the stochastic nature of pavements. However, the same optimal M&R strategy, i.e. the one with the objectives of maximizing pavement condition and minimizing maintenance cost studied in Section 5.2.5, will be achieved as the majority of the samples generated from a different distribution would fall into a narrow interval approaching the mean value.

### 7.2.4 Road user cost

9. The road user cost is defined as the time cost, including the journey time cost and the delay time cost, whereas the vehicle operating cost is not taken into account. For the normal condition, i.e. when there are no congestions, vehicle operating cost is proportional to the journey time cost and it can be neglected, since in this case the selection of maintenance actions is dominated by maintenance cost. While at peak hours and at road works, with the increase of congestions, the vehicle operating cost also increases significantly. As a consequence, vehicle operating cost should be included as part of the road user cost in the future.

### 7.2.5 Maintenance implementation

10. The maintenance set-up costs and removal costs are not considered when the road link needs to be repaired, since they are very small and most likely would not affect the optimal solution. Such costs need to be taken into account when the maintenance works on a link are performed with maintenance breaks, and they are relatively larger in comparison to the maintenance cost.
11. For the optimisation framework, the traffic inputs for the network during the planning period, i.e. 10 years, are assumed to be constant. For the illustration such assumption is important in order to obtain results reasonably quickly, 30 minutes. In reality, the traffic flow rates on the network would increase slightly and it would have some effect on the traffic conditions on the network.

## 7.3 Conclusions

The developed NTFM provides a modelling framework which enables any road network to be modelled by constructing the network topology using the embedded junctions connected by road sections. The traffic inputs for NTFM include the entry flows and turning ratios at the involved junctions, and the geometries of junctions and road sections. NTFM is developed to be applicable for any road network under consideration.

A roadwork node sub-model has been introduced in the NTFM to study the effect of maintenance work on network performance, which specifies the location of worksite, the length of the worksite, and lane closure options.

Traffic rerouting strategy enables NTFM to model the driver behaviour, i.e. when excessive traffic congestion exists on a highway network, traffic will take alternative routes and therefore traffic rerouting takes place.

Pavement age gain model is applied to evaluate and predict the evolution of pavement condition during the planning period, in which pavement age gain is deployed as the pavement improvement indicator. Moreover, in order to capture the uncertainty inherent in pavement deterioration process, the deterministic pavement age gain model can be transformed to a probabilistic one by using a normal distribution for the pavement age gain.

NSGA-II is utilized in the optimisation framework to achieve the global optimal M&R strategy subject to the specified objectives and constraints. Application of the NSGA-II provided a number of good solutions which are critical in the scheduling of pavement maintenance.

Road user cost and maintenance cost are taken into account concurrently so as to achieve the optimal maintenance actions for the road sections during maintenance. Maintenance arrangements, i.e. start time of maintenance, maintenance duration, maintenance break, and traffic control measures, can be optimised based on NTFM according to the network performance under various maintenance scenarios. In comparison to other worksite optimisation techniques, not only the traffic delays at the worksites are taken into account by NTFM, but the traffic flows through the adjacent links are also considered.

#### **7.4 Future Model Validation**

In order to apply the NTFM to model a highway network and obtain a close to reality representative of the network, the following steps need to be fulfilled:

1. The different types of junctions and links on the network need to be identified.
2. The flow capacity at each junction, and the length, flow capacity and capacitance of each link need to be identified.

3. Traffic flow rates, turning ratios and traffic signals at each entry arm of the junctions need to be derived and calibrated according to traffic data throughout a typical day.
4. The average travel speed and the speed of queue dissipation for vehicles on each type of road need to be defined.
5. The headway between delayed vehicles needs to be specified.
6. The expected lives of pavement for each type of maintenance actions associated with road classes need to be verified.

## **7.5 Future Work**

Future work would involve an application of the developed decision making system to a variety of highway networks. It is anticipated that the analysis of more complex problems would require more computational effort in terms of the size of road network and the complexity of objective functions, which may result in a processing time issue. Therefore, future work should be devoted to improving the efficiency of the analysis to minimise CPU processing time.

Only the maintenance activities on road links are taken into account by NTFM, thus the impact of maintenance activities at the junctions should be investigated in the future. More regulations and rules, such as speed reductions due to the enforcement of worker safety at worksites and the evaluation of vehicle operating cost, could be applied to NTFM. Further sub-models for the junction can be derived e.g. a three-lane signalised intersection and a three-lane roundabout.

GAs were identified as very robust and flexible techniques for the optimisation of pavement maintenance arrangements. In this thesis, NSGA-II is applied as the optimisation method which has proven itself to be sufficient enough for constrained multi-objective problems. However, more constraints can be added to the optimisation

model, such as limitation of the use of surface dressing on motorways, and forcing structural maintenance on heavily trafficked roads (at least once during the life cycle). Therefore, other advanced GAs and hybrid heuristic techniques can be implemented, as they have the feature of retaining the advantages of GAs but prominently improve their searching ability and computational efficiency in achieving the global optimal.

Pavement age gain model is a very useful technique for evaluating pavement condition, whereas the pavement surface defects, usually recognised as the trigger of pavement actions, are not investigated. In addition, the maintenance actions provided by ADEPT/RSTA [63] can be applied to the pavement age gain model. Deploying other existing pavement deterioration models, i.e. HDM-4 and ARRB, and individual pavement distress models, such as cracking model and rutting model, can provide an alternative solution for the problem in terms of the decision making criteria. When an alternative pavement deterioration model is used, it is still possible to take advantage of NTFM model.

## References

1. *Transport Statistics Great Britain: 2011 Edition*, 2011: UK Department for Transport.
2. (AIA), A.I.A., *Annual Local Authority Road Maintenance Survey 2011*.
3. Wang, F., Zhang, Z., Machemehl, R. B., *Decision-Making Problem for Managing Pavement Maintenance and Rehabilitation Projects*. Transportation Research Record: Journal of the Transportation Research Board, 2003. 1853(-1): p. 21-28.
4. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 13 Section 1 Part 5: Speeds on links 2004*, Department for Transport
5. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 14 Section 1 Part 5: The Traffic input to QUADRO*, 2004, Department for Transport
6. Lighthill, M.J. and G.B. Whitham, *On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads*. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1955. 229(1178): p. 317-345.
7. Richards, P.I., *Shock Waves on the Highway*. Operations Research, 1956. 4(1): p. 42-51.
8. Cremer, M. and J. Ludwig, *A fast simulation model for traffic flow on the basis of boolean operations*. Mathematics and Computers in Simulation, 1986. 28(4): p. 297-303.
9. Nagel, K. and M. Schreckenberg, *A cellular automaton model for freeway traffic*. Journal de Physique I, 1992. 2(12): p. 2221-2229.
10. Daganzo, C.F., *The Cell Transmission Model - a Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory*. Transportation Research Part B-Methodological, 1994. 28(4): p. 269-287.
11. Payne, H.J., *Models of freeway traffic and control*. . 1971. Simulation council proceeding 1: ch6: p. 51-61.
12. Wagner, C., Hoffmann, C., Sollacher, R., Wagenhuber, J., Schürmann, B., *Second-order continuum traffic flow model*. Physical Review E, 1996. 54(5): p. 5073.
13. Rothery, R.W., *Car-following models*. In Traffic flow theory: A state of the art report - revised monograph on traffic flow theory, N. H. Gartner, C. Messer and A. K. Rathi, Eds., Oak Ridge, Tennessee, Oak Ridge National Laboratory., 1997.
14. Daganzo, C.F., *The Cell Transmission Model .2. Network Traffic*. Transportation Research Part B-Methodological, 1995. 29(2): p. 79-93.
15. Lo, H.K., *A novel traffic signal control formulation*. Transportation Research Part A: Policy and Practice, 1999. 33(6): p. 433-448.

16. Lo, H. K., Chang, E., Chan, Y. C., *Dynamic network traffic control*. Transportation Research Part A: Policy and Practice, 2001. 35(8): p. 721-744.
17. Messmer, A. and M. Papageorgiou, *METANET: A macroscopic simulation program for motorway networks*. Traffic Engineering and Control, 1990. 31(9): p. 466-470.
18. Breton, P., Hegyi, A., De Schutter, B., Hellendoorn, H., *Shock wave elimination/reduction by optimal coordination of variable speed limits*. in *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference*. 2002.
19. Hegyi, A., Bart De, S., Hellendoorn, J., *Optimal coordination of variable speed limits to suppress shock waves*. Intelligent Transportation Systems, IEEE Transactions on, 2005. 6(1): p. 102-112.
20. Deflorio, F.P., *Evaluation of a reactive dynamic route guidance strategy*. Transportation Research Part C: Emerging Technologies, 2003. 11(5): p. 375-388.
21. Karimi, A., Hegyi, A., De Schutter, B., Hellendoorn, H., Middelham, F., *Integration of dynamic route guidance and freeway ramp metering using model predictive control*. in *American Control Conference, 2004. Proceedings of the 2004*. 2004.
22. Van den Berg, M., Hegyi, A., De Schutter, B., *Integrated traffic control for mixed urban and freeway networks: A model predictive control approach*. European Journal of Transportation and Infrastructure Research, 2007. 3(3): p. 223-250.
23. Kashani, H.R. and G.N. Saridis, *Intelligent control for urban traffic systems*. Automatica, 1983. 19(2): p. 191-197.
24. Troutbeck, R.J. and W. Brilon, *Unsignalized intersection theory*. 1997.
25. Brilon, W., Koenig, R., Troutbeck, R. J., *Useful estimation procedures for critical gaps*. Transportation Research Part A: Policy and Practice, 1999. 33(3-4): p. 161-186.
26. Tian, Z. Z., Vandehey, M., Robinson, B. W., Kittelson, W., Kyte, M., Troutbeck, R., Brilon, W., Wu, N., *Implementing the maximum likelihood methodology to measure a driver's critical gap*. Transportation Research Part a-Policy and Practice, 1999. 33(3-4): p. 187-197.
27. Hewitt, R.H., *Measuring Critical Gap*. TRANSPORTATION SCIENCE, 1983. 17(1): p. 87-109.
28. Siegloch, W., *Die Leistungsermittlung an Knotenpunkten ohne Lichtsignalsteuerung g (Capacity determination at intersections without traffic signals)*. Schriftenreihe Strabenbau und strabenverkehrstechnik, 1973. 154.



29. Cowan, R.J., *Useful headway models*. Transportation Research, 1975. 9(6): p. 371-375.
30. Plank, A.W. and E.A. Catchpole, *A general capacity formula for uncontrolled intersection*. Traffic Eng. Control, 1984. 25: p. 327-329.
31. Kimber, R.M., *Gap-Acceptance and Empiricism in Capacity Prediction*. TRANSPORTATION SCIENCE, 1989. 23(2): p. 100.
32. Troutbeck, R.J. and S. Kako, *Limited priority merge at unsignalized intersections*. Transportation Research Part A: Policy and Practice, 1999. 33(3-4): p. 291-304.
33. Bunker, J. and R. Troutbeck, *Prediction of minor stream delays at a limited priority freeway merge*. Transportation Research Part B-Methodological, 2003. 37(8): p. 719-735.
34. Tanner, J.C., *A Theoretical Analysis of Delays at an Uncontrolled Intersection*. Biometrika, 1962. 49(1/2): p. 163-170.
35. Dirk, H., *Queue length and waiting-time distributions at priority intersections*. Transportation Research Part B: Methodological, 1991. 25(4): p. 163-174.
36. Yeo, G.F. and B. Weesakul, *Delays to Road Traffic at an Intersection*. Journal of Applied Probability, 1964. 1(2): p. 297-310.
37. Kremser, H., *Ein Zusammengesetztes Wartezeitproblem Bei Poissonschen Verkehrsströmen (A Complex Problem of Delay with Poisson-Distributed Traffic Flows)*. Österreichisches Ingenieur-Archiv, 1962. 16.
38. Kremser, H., *Wartezeiten Und Warteschlangen Bei Einfadlung Eines Poissonprozesses in Einen Anderen Solchen Prozess (Delays and Queues with One Poisson Process Merging into Another One)*. Österreichisches Ingenieur-Archiv, 1964. 18.
39. Daganzo, C.F., *Traffic Delay at Unsignalized Intersections: Clarification of Some Issues*. TRANSPORTATION SCIENCE, 1977. 11(2): p. 180.
40. Poeschl, F.J., *Die Nicht Signalgesteuerte Nebenstrassenzufahrt Als Verallgemeinertes M/G/1- Warteschlangensystem (The Unsignalized Minor Street Entry as a Generalized M/G/1 Queueing System)*. Zeitschrift für Operations Research, 1983. 27B.
41. Ning, W., *A universal procedure for capacity determination at unsignalized (priority-controlled) intersections*. Transportation Research Part B: Methodological, 2001. 35(6): p. 593-623.
42. Ruskin, H.J. and R. Wang, *Modeling Traffic Flow at an Urban Unsignalized Intersection*, in *Proceedings of the International Conference on Computational Science-Part I2002*, Springer-Verlag. p. 381-390.

43. Tracz, M. and S. Gondek, *Use of simulation to analysis of impedance impact at unsignalized intersections*. Transportation Research Circular, 2000(E-C018): p. 471-484.
44. Robinson, B. W., Tian, Z. Z., Kittelson, W., Vandehey, M., Kyte, M., Brilon, W., Wu, N., Troutbeck, R., *Extensions of theoretical capacity models to account for special conditions*. Transportation Research Part A: Policy and Practice, 1999. 33(3-4): p. 217-236.
45. Tian, Z. Z., Troutbeck, R., Kyte, M., Brilon, W., Vandehey, M., Kittelson, W., Robinson, B. W., *A further investigation on critical gap and follow-up time* Transportation Research Circular E-C108: 4th International Symposium on Highway Capacity 2000: p. 397-408.
46. Messmer, A. and M. Papageorgiou, *Route Diversion Control in Motorway Networks Via Nonlinear Optimization*. Ieee Transactions on Control Systems Technology, 1995. 3(1): p. 144-154.
47. Wardrop, J.G., *SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH*. Proc. Inst. Civ. Engrs II, 1952: p. 325-378.
48. Wie, B., Tobin, R. L., Bernstein, D., Friesz, T. L., *A comparison of system optimum and user equilibrium dynamic traffic assignments with schedule delays*. Transportation Research Part C: Emerging Technologies, 1995. 3(6): p. 389-411.
49. Papageorgiou, M., *An Integrated Control Approach for Traffic Corridors*. Transportation Research Part C-Emerging Technologies, 1995. 3(1): p. 19-30.
50. Iftar, A., *A decentralized routing controller for congested highways*. Proceedings of the 34th Ieee Conference on Decision and Control, Vols 1-4, 1995: p. 4089-4094.
51. Iftar, A., *A linear programming based decentralized routing controller for congested highways*. Automatica, 1999. 35(2): p. 279-292.
52. Chang, G. L., Ho, P. K., Wei, C. H., *A dynamic system-optimum control model for commuting traffic corridors*. Transportation Research Part C: Emerging Technologies, 1993. 1(1): p. 3-22.
53. Wu, J. and G.L. Chang, *An integrated optimal control and algorithm for commuting corridors*. International Transactions in Operational Research, 1999. 6(1): p. 39-55.
54. Liu, Y. and G.L. Chang, *An arterial signal optimization model for intersections experiencing queue spillback and lane blockage*. Transportation Research Part C-Emerging Technologies, 2011. 19(1): p. 130-144.
55. Liu, Y., Yu, J., Chang, G. L., Rahwanji, S., *A Lane-group Based Macroscopic Model for Signalized Intersections Account for Shared Lanes and Blockages*.

- Proceedings of the 11th International Ieee Conference on Intelligent Transportation Systems, 2008: p. 639-644.
56. Liu, Y., Chang, G. L., Yu, J., *An Integrated Control Model for Freeway Corridor Under Nonrecurrent Congestion*. Vehicular Technology, IEEE Transactions on, 2011. 60(4): p. 1404-1418.
  57. Zhang, L. and D. Levinson, *Optimal freeway ramp control without origin-destination information*. Transportation Research Part B-Methodological, 2004. 38(10): p. 869-887.
  58. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 14: Economic Assessment of Road Maintenance*, 2004, Department for Transport
  59. Gazis, D.C. and R.B. Potts, *The oversaturated intersection*. Proceedings of the 2nd International Symposium on Traffic Theory 1963, London, United Kingdom
  60. Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., *Store-and-forward based methods for the signal control problem in large-scale congested urban road networks*. Transportation Research Part C: Emerging Technologies, 2009. 17(2): p. 163-174.
  61. DfT, *WebTAG: Unit 3.5.6*, D.f. Transport, Editor 2011.
  62. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 7: Pavement Design and Maintenance* 1994: Department for Transport
  63. ADEPT/RSTA, *SERVICE LIFE OF SURFACE TREATMENTS BY ADEPT/RSTA*. 2011.
  64. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 7 Section 4 Part 1: HD31/94*, 1995: Department for Transport
  65. Cook, W.D. and R.L. Lytton, *Recent developments and potential future directions in ranking and optimisation procedures for pavement management*, in *In Proc. Second North American Conf. on Managing Pavements* 1987: Toronto, Canada. p. 2. 135-2.155.
  66. Robinson, R., Danielson, U., Snaith, M., *Road Maintenance Management: Concepts and Systems* 1998: Anthony Rowe Ltd, Chippenham, Wiltshire.
  67. A., H. and Hassab-alla, *Decision support model for prioritising pavement maintenance activities LP GIS approach*, I.i.f.a.s.a.e.s. enschede, Editor 2001: Netherland.
  68. Zimmerman, K.A., *Pavement management methodologies to select projects and recommend preservation treatments*, ed. T.R.B.N.S. 222.1995.
  69. Gendreau, M. and P. Soriano, *Airport pavement management systems: an appraisal of existing methodologies*. Transportation Research Part A: Policy and Practice, 1998. 32(3): p. 197-214.

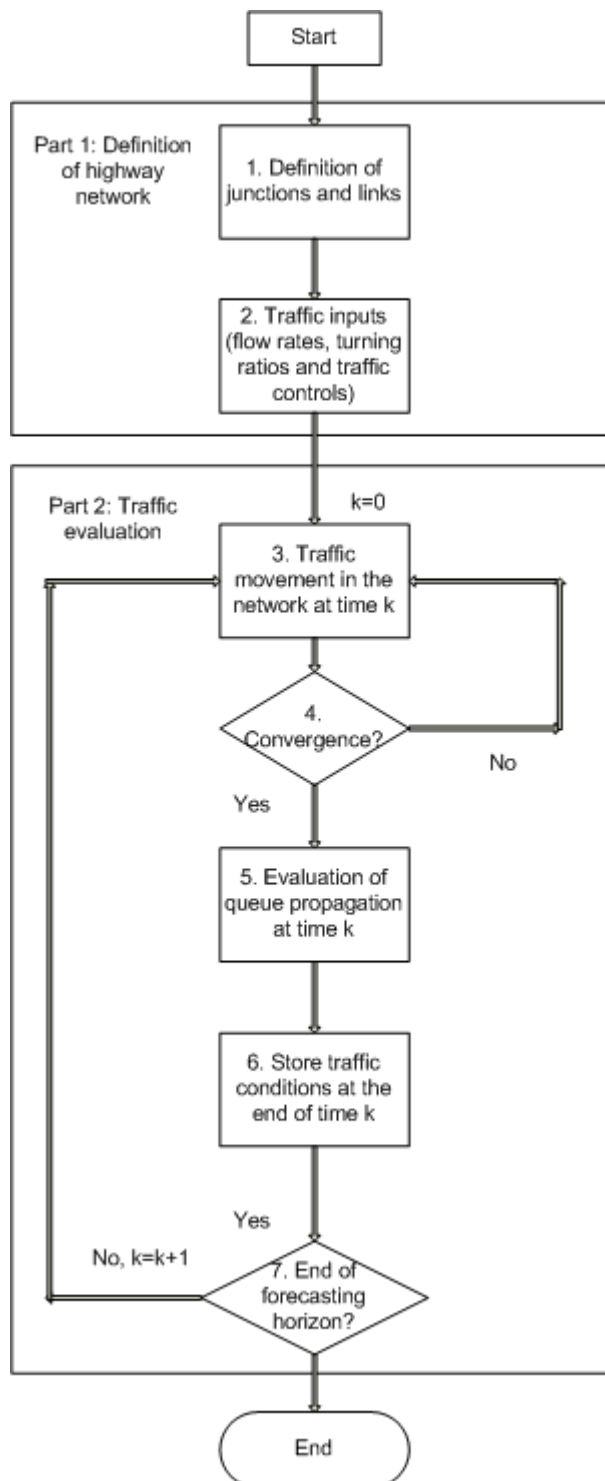
70. Golabi, K., Kulkarni, R., Way, G., *A Statewide Pavement Management System*. Interfaces, 1982. 12(6).
71. Ferreira, A., Picado-Santos, L., Antunes, A., *A Segment-linked Optimization Model for Deterministic Pavement Management Systems*. International Journal of Pavement Engineering, 2002. 3(2): p. 95-105.
72. Mbwana, J. and M. Turnquist, *Optimization Modeling for Enhanced Network-Level Pavement Management System*. Transportation Research Record: Journal of the Transportation Research Board, 1996. 1524(-1): p. 76-85.
73. Ferreira, A., Antunes, A., Picado-Santos, L., *Probabilistic Segment-linked Pavement Management Optimization Model*. Journal of Transportation Engineering, 2002. 128(6): p. 568-577.
74. Wu, Z. and G.W. Flintsch, *Pavement Preservation Optimization Considering Multiple Objectives and Budget Variability*. Journal of Transportation Engineering, 2009. 135(5): p. 305-315.
75. Cohon, J.L., *Multiobjective Programming and Planning* ed. N.Y. Academic 1978.
76. Holland, J.H., *Adaptation in Natural and Artificial Systems* 1975, University of Michigan Press, Ann Arbor, Mich
77. Zadeh, L.A., *Optimality and Non-Scalar-Valued Performance Criteria*. {IEEE} Transactions on Automatic Control, 1963. 8: p. 59-60.
78. Messac, A., Ismail-Yahaya, A., Mattson, C. A., *The normalized normal constraint method for generating the Pareto frontier*. Structural and Multidisciplinary Optimization, 2003. 25(2): p. 86-98.
79. Marler, R.T. and J. Arora, *The weighted sum method for multi-objective optimization: new insights*. Structural and Multidisciplinary Optimization, 2010. 41(6): p. 853-862.
80. Morcoux, G. and Z. Lounis, *Maintenance optimization of infrastructure networks using genetic algorithms*. Automation in Construction, 2005. 14(1): p. 129-142.
81. Goldberg, D.E., *Genetic algorithms in search, optimization and machine learning* 1989: Addison-Wesley Publishing Company, Reading, Mass.
82. Bosurgi, G. and F. Trifirò, *A model based on artificial neural networks and genetic algorithms for pavement maintenance management*. International Journal of Pavement Engineering, 2005. 6(3): p. 201-209.
83. Chan, W.T., Fwa, T. F., Tan, C.Y., *Road-Maintenance Planning Using Genetic Algorithms. I: Formulation*. Journal of Transportation Engineering, 1994. 120(5): p. 693-709.

84. Fwa, T.F., Tan, C.Y., Chan, W.T., *Road-Maintenance Planning Using Genetic Algorithms. II: Analysis*. Journal of Transportation Engineering, 1994. 120(5): p. 710-722.
85. Abaza, K.A. and S.A. Ashur, *Optimum microscopic pavement management model using constrained integer linear programming*. International Journal of Pavement Engineering, 2009. 10(3): p. 149-160.
86. Fwa, T.F., Chan, W.T., Hoque, K. Z., *Multiobjective Optimization for Pavement Maintenance Programming*. Journal of Transportation Engineering, 2000. 126(5): p. 367-374.
87. MBWANA, J.R., *A FRAMEWORK FOR DEVELOPING STOCHASTIC MULTI-OBJECTIVE PAVEMENT MANAGEMENT SYSTEMS*, in *First Road Transportation Technology Transfer Conference in Africa*2001. p. 350-363.
88. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., *A fast and elitist multiobjective genetic algorithm: NSGA-II*. Evolutionary Computation, IEEE Transactions on, 2002. 6(2): p. 182-197.
89. Konak, A., Coit, D., Smith, A., *Multi-objective optimization using genetic algorithms: A tutorial*. Special Issue - Genetic Algorithms and Reliability, 2006. 91(9): p. 992-1007.
90. Bandaru, S., Tulshyan, R., Deb, K., *Modified SBX and adaptive mutation for real world single objective optimization*. in *Evolutionary Computation (CEC), 2011 IEEE Congress on*. 2011.
91. Tao, H.H., Liao, G. S., Wang, L., et al. *Integer coded genetic algorithm design of staggered sampling MTI*. in *Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on*. 2003.
92. Coley, D.A., *An Introduction to Genetic Algorithms for Scientists and Engineers*1999, World Scientific Publishing Co. Pte. Ltd., Singapore.
93. Mitchell, M., *An Introduction to Genetic Algorithms*1996: MIT Press.
94. Zalzala, A.M.S., Fleming, P. J., *Genetic algorithms in engineering systems*1997: London: Institution of Electrical Engineers.
95. Ivo F. S., Sibylle, M., Petros, K., *Multiobjective optimization using evolutionary algorithms*, in *Center for Turbulence Research, Proceedings of the Summer program 2000*2000. p. 63-74.
96. Schaffer, J.D. *Multiple objective optimization with vector evaluated genetic algorithms*. in *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*. 1985. Lawrence Erlbaum.
97. Andersson, J., *A survey of multiobjective optimization in engineering design*, in *Technical Report LiTH-IKP-R-1097*2000: Department of Mechanical Engineering, Linkping University, Linkping, Sweden.

98. Srinivas, N. and K. Deb, *Multiojective Optimization Using Nondominated Sorting in Genetic Algorithms*. *Evolutionary Computation*, 1994. 2(3): p. 221-248.
99. Goldberg, D. and J. Richardson. *Genetic algorithms with sharing for multimodal function optimization*. in *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*. 1987. Lawrence Erlbaum Associates, Inc.
100. Zitzler, E., Deb, K., Thiele, L., *Comparison of Multiobjective Evolutionary Algorithms: Empirical Results*. *Evolutionary Computation*, 2000. 8(2): p. 173-195.
101. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 5 Section 1 Part 3: TA 46/97*, 1997, Department for Transport
102. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 13: Economic Assessment of Road Schemes*, 2004, Department for Transport
103. Chien, S., Tang, Y., Schonfeld, P., *Optimizing Work Zones for Two-Lane Highway Maintenance Projects*. *Journal of Transportation Engineering*, 2002. 128(2): p. 145-155.
104. Schonfeld, P. and S. Chien, *Optimal Work Zone Lengths for Two-Lane Highways*. *Journal of Transportation Engineering*, 1999. 125(1): p. 21-29.
105. Chien, S. and P. Schonfeld, *Optimal Work Zone Lengths for Four-Lane Highways*. *Journal of Transportation Engineering*, 2001. 127(2): p. 124-131.
106. Fwa, T.F., Cheu, R. L., Muntasir, A, *Scheduling of pavement maintenance to minimize traffic delays*. *Transportation Research Record 1650*, 1998: p. 28-35.
107. Jiang, X. and H. Adeli, *Freeway Work Zone Traffic Delay and Cost Optimization Model*. *Journal of Transportation Engineering*, 2003. 129(3): p. 230-241.
108. Chien, C.-H. and P. Schonfeld, *Work Zone Lengths for a Four-Lane Road with an Alternate Route*. *Journal of Transportation Engineering-Asce*, 2005. 131(10): p. 780-789.
109. MEmmott, J.L. and C.L. Dudek, *Queue and User Cost Evaluation of Work Zones (QUEWZ)*. *Transportation Research Record 1086*, 1984. 979: p. 12-19.
110. Krammes, R.A. and G.O. Lopez, *UPDATED CAPACITY VALUES FOR SHORT-TERM FREEWAY WORK ZONE LANE CLOSURES*. *Transportation Research Record 1086*, 1994. 1442: p. 49-56.
111. Richards, S.H. and C.L. Dudek, *IMPLEMENTATION OF WORK-ZONE SPEED CONTROL MEASURES*. *Transportation Research Record 1086*, 1986. 1086: p. 36-42.

112. Shepard, F.D. and B.H. Cottrell Jr, *BENEFITS AND SAFETY IMPACT OF NIGHT WORK-ZONE ACTIVITIES*. Transportation Research Record 1086, 1986. 1086: p. 31-36.
113. DfT, *Design Manual for Roads and Bridges (DMRB) Volume 14 Section 1 Part 5: The Traffic Input to QUADRO* 2004, Department for Transport

## Appendix A: Flow Diagram of NTFM





The NTFM, programmed in Visual C++ and presented in the flow chart, is processed as follows:

#### Part 1: Definition of highway network

1. The first step for the NTFM is to define the geometries of junctions and links on the network. A class is constructed for each junction studied in the NTFM to describe its type and flow capacity, and to store its inflows, outflows and queues. Links are specified in terms of the road type, the number of lanes, and their length. The relationship between junctions needs to be specified according to the topology of the highway network. As junctions are connected by two-way links, the two-way traffic flows are investigated in the NTFM by using iterative algorithm.
2. The junctions are defined with the initialisation of traffic flow rates, turning ratios at each entry arm, the measured traffic controls, and the traffic conditions at the beginning of each time step,  $k$ , i.e. the length of the queue.

#### Part 2: Traffic evaluation

3. NTFM is used to model the network at time step  $k$  iteratively until the convergence has been achieved. The inflows for each junction are updated at each iteration according to the outflows on the corresponding upstream links at the previous iteration. The traffic characteristics at each junction are evaluated based on the geometry, the initial traffic condition and the inflows at each iteration; the outputs include the outflows and the queues for each junction, and the components of queue are identified, i.e. the number of vehicles for each potential direction.
4. The convergence is reached when the traffic outputs for each junction on the network are identical at the last 6 iterations of the process.

5. The queue propagation is performed when some links on the network are completely filled with vehicles, and a number of the delayed vehicles that exceed the corresponding link capacitance spill back to upstream links.
6. After the evaluation of the traffic condition on the network at time  $k$ , the inflows, the outflows and the queues for each junction and link on the network are kept in vectors and are used in the next time step.
7. As for the next time step  $k+1$ , the traffic conditions at the end of time  $k$  are transferred as the initial traffic conditions at the beginning of time  $k+1$ . Steps 3-7 are repeated until the end of the planning horizon.

## **Appendix B: Practical Implementation and Computational Time of the Decision Making System**

For a simple road network, i.e. the case study network provided in Section 3.6 (4 nodes and 12 links), it takes about 1 second for NTFM to model the traffic on the network for an Intel Core 2 PC. As for a medium-sized network, i.e. Loughborough-Nottingham highway network (47 nodes and 51 links), it takes around 10 seconds to model the traffic through the network during a typical day. Consequently, it is expected that it may take about 30 minutes for the NTFM to simulate the traffic on a large-sized network during a relatively longer period, i.e. a network composed of more than 100 nodes and 100 links, and the planning period is defined as 1 month.

As for the optimisation framework, it takes about 15 minutes to optimise the M&R strategy for a simple road network, i.e. the example road network that was investigated in Section 5.2.4 (4 links), when the maximum generation is defined as 5000. The total optimisation time is increased to 2 hours when 20000 generations are required, which indicates that the computational effort increases exponentially with the increase of the number of GA parameters. As a result, it would take a couple of months to optimise the M&R strategy for Loughborough-Nottingham highway network (51 links) in a longer term, i.e. 20 years.