



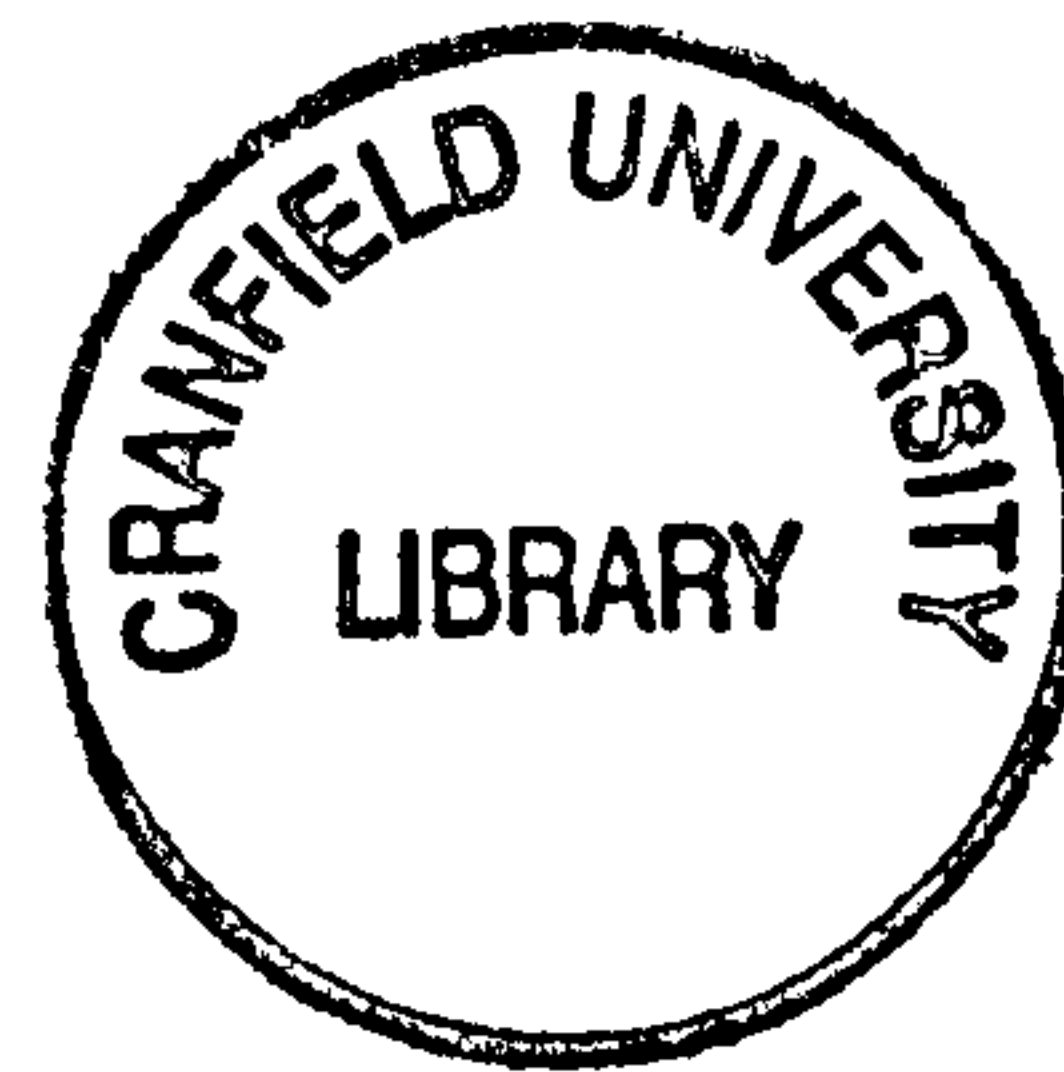
Cranfield University
School of Management

Adel S. A. Abd EL-Maksoad

**TIME-DEPENDENT ROAD PRICING:
MODELLING AND EVALUATION**

Cranfield Centre for Logistics and Transportation

PhD Thesis



Cranfield University
School of Management
Cranfield Centre for Logistics and Transportation

PhD Thesis

Adel S. A. Abd EL-Maksoad

**TIME-DEPENDENT ROAD PRICING:
MODELLING AND EVALUATION**

Supervisor: Ian G. Black

(September 1995)

This thesis is submitted for the degree of Doctor of Philosophy

Abstract

Road pricing has an established history in the literature of transport economics, and its use as a theoretical and practical means of traffic restraint and management has attracted considerable interest for more than half a century. The theory of road pricing asserts that the optimal price should be the one that reflects the full cost of making an additional trip. Evidently, the magnitude of such a congestion toll varies over time and space. However, a review of some practical applications of road pricing in different countries reveals that no policy has as yet been implemented which aims to charge road users directly in relation to the congestion they actually cause and the time delay they impose on others. Therefore, the aim of this study is to model such a time-varying pricing scheme, termed: Time-Dependent Road Pricing, *TDRP*, and evaluate its various impacts on a single bottleneck as well as a traffic network.

The *TDRP* function is derived based on the solution of the time-dependent queues and delays problem at traffic junctions. The derived function is demonstrated to lead to a very unstable user equilibrium for a single bottleneck. Therefore, two different approaches are adopted to modify this function: first, by considering the schedule delay changes imposed by vehicles on one another; and second, by using the day-to-day adjustment process. The former approach is demonstrated to eliminate queues completely and thus lead to system optimal *SO* for a single traffic bottleneck. Besides, the simulation solution demonstrates that *TDRP*, modified by the second approach, could lead to a stable equilibrium, and although it does not lead to *SO* it results in a very substantial reduction in queuing delay and travel time.

To evaluate the stability of the results and the different impacts of *TDRP* on a traffic network, a traffic assignment model is developed. This model embraces route choices, departure time choices and the *TDRP* function, and it has the ability to evaluate the road network under different charging systems. Before evaluating the different impacts of *TDRP* on a traffic network, the importance of the phenomenon of interaction between nodes and its impacts on the value of *TDRP* are discussed. A general solution under specific traffic conditions as well as different *TDRP* scenarios are suggested.

A set of numerical simulation experiments using the assignment model and a typical traffic network for urban areas is conducted. The results demonstrate that although *TDRP* does not eliminate the queues completely, it leads to a very substantial saving in travel time and queuing delay for all movements throughout the network under different levels of congestion. On the other hand, exempting some nodes (or links) from the charges, would lead to a very substantial fall in the benefit obtained. The comparative analysis demonstrates that *TDRP* is a superior charging system compared with other charging systems. It is also concluded that *TDRP* does not represent the optimal charging system for a traffic network since other charging methods could lead to a better performance under very high levels of charge.

Finally, the sensitivity of the results to work start time flexibility and the shadow values of the schedule delay function is investigated, and at the end, directions for further research are proposed.

Acknowledgement

I would like to take this opportunity to express my gratitude and sincere thanks to my supervisor Ian G. Black who encouraged me to start this study and whose valuable comments, attention and interest have contributed most to the development of this work.

I am also indebted to professor M. Kuwahara, Institute of Industrial Science, Tokyo, Japan, for developing the network traffic simulation model and for his invaluable assistance in adding the necessary modifications to the model in order to suit the objectives of this study.

I would like to acknowledge the invaluable advice and useful suggestions received from professor M. J. Smith, York University.

Also I wish to express my sincere appreciation to my parents, brothers and sisters for their encouragement and support during all the stages of my education. Besides, I must express my considerable debt to my wife, son, Mazen, and daughter, Rana, for their warmth, support and above all endurance during my graduate studies at Cranfield.

Finally, this research was supported by the Overseas Development Administration (ODA) and the British Council, UK, and Ishikawa Fellowship, Japan, their contributions are gratefully acknowledged.

Table of Contents

	<i>Page</i>
1. INTRODUCTION	1
1.1 Urban Traffic Congestion	1
1.2 Different Policies of Traffic Restraint	1
1.3 Road Pricing: Basic Concept and Rationale for Further Development .	6
1.4 Time-Dependent Road Pricing (<i>TDRP</i>)	7
1.5 Setting the Study Objectives	8
1.6 Structure of the Study	9
 2. ROAD PRICING BACKGROUND	 10
2.1 History of Road Pricing	10
2.1.1 Previous Studies	10
2.1.2 Economics of Road Pricing	11
2.1.3 Is Road Pricing fair?	11
2.2 Paying for Roads	12
2.2.1 Why pay for roads?	13
2.2.2 Who will pay the road user charges?	13
2.2.3 Who will benefit from road user charges?	13
2.3 The Costs of Roads	14
2.3.1 The costs of providing the roads	14
2.3.2 The costs arising out of journeys	15
2.4 Pricing Strategies (Policies)	17
2.4.1 Arbitrary pricing	17
2.4.2 Average cost pricing	17
2.4.3 Marginal cost pricing	17
2.4.4 Monopoly pricing	19
2.5 Theory of Road Pricing and its Drawback	20
2.5.1 Theory of road pricing	20
2.5.2 Determining the size of charge in practice	21
2.5.3 Drawback of the theory and the line of development	22
2.6 Practical Issues Involved in the Implementation of RP	22
2.6.1 Methods of charging	22
2.6.2 Exemptions for certain vehicles or persons	25
2.6.3 Temporal pattern of charges	26
2.6.4 Spatial charging structures	26
2.6.5 Other aspects	27
2.7 Singapore, Hong Kong and Norway Experiences with RP	27
2.8 Recent UK Studies Concerning RP (RP systems proposed for Cambridge) .	30
2.9 Summary	31
 3. THE APPROACH FOR MODELLING <i>TDRP</i>	 33
3.1 Introduction	33
3.2 Theory of queuing and delay	33
3.2.1 Characteristics of queuing process	33
3.2.2 Mathematical solution of queuing problems	35
3.2.2.1 Steady-State Solutions	35
3.2.2.2 Deterministic Solutions	36

	<i>Page</i>
3.3 Time-Dependent queues and delays at road junctions	37
3.3.1 Time-dependent demand-capacity interaction at traffic junctions	38
3.3.2 The basic queuing problem	38
3.3.3 The average queue length	39
3.3.4 The average vehicular delay	40
3.4 Solutions of time-dependent queuing problem	40
3.4.1 Queuing probabilities	41
3.4.1.1 Known initial queue length	41
3.4.1.2 General probability distribution for initial queue length	41
3.4.2 Average queue length	41
3.4.3 Average Delays	43
3.5 Co-ordinate Transformation solution for the time-dependent queuing problem	44
3.5.1 queue length (L) as a function of demand, capacity and time	44
3.5.2 The co-ordinate transformation technique applied to vehicle delays	46
3.5.3 Modification for accuracy	49
3.6 Modelling the time-dependent delay imposed by each vehicle on other vehicles	49
3.7 Summary	54
4. <u>A REVIEW OF MODELS FOR TEMPORAL DISTRIBUTION OF PEAK PERIOD DEMAND FOR A BOTTLENECK</u>	55
4.1 Introduction	55
4.2 General description of the dynamic models of peak period	55
4.2.1 Deterministic User Equilibrium (<i>DUE</i>) Approach	57
4.2.2 Stochastic User Equilibrium (<i>SUE</i>) Approach	62
4.2.3 System Optimal (<i>SO</i>) Approach	67
4.3 The day-to-day evolution of the departure-time decisions	68
4.4 The stability of <i>DUE</i> under Time-Dependent Road Pricing (<i>TDRP</i>) ..	70
4.4.1 Modifying <i>TDRP</i> to include the schedule delay changes to other vehicle	73
4.4.2 Modifying <i>TDRP</i> using the day-to-day adjustment process	75
4.5 Summary	75
5. <u>SIMULATION SOLUTION FOR A SINGLE TRAFFIC BOTTLENECK</u>	76
5.1 Introduction	76
5.2 The dynamic simulation model	76
5.2.1 Structure of the simulation model	76
5.2.2 Deterministic departure-time selection process	78
5.2.3 Stochastic departure-time selection process	78
5.2.4 Other features of the simulation model	79
5.3 Simulation experiments (A numerical example)	79
5.3.1 Base values of the model parameters	80
5.3.2 Convergence and stability of the base case results	80
5.3.3 Base case departure time choices and Generalized Cost results	83
5.3.4 The different impacts of <i>TDRP</i> on the base case results	85
5.4 Summary and concluding remarks	92
6. <u>NETWORK TRAFFIC SIMULATION MODEL</u>	93
6.1 Introduction	93
6.2 General description of the simulation model	93

	<i>Page</i>
6.3 Principles of the model	94
6.3.1 Overall Structure	94
6.3.2 Time Variation/Time-Intervals	94
6.3.3 Packets/Vehicles Generations	94
6.3.4 Link Performance	94
6.3.4.1 Link Flow-Density Relationship	97
6.3.4.2 Car-Flow Behaviour	99
6.3.5 Route Choice Model (Dial's Model)	100
6.3.6 Propagation of Packets/Vehicles from one link to another	102
6.3.7 Model input/output	103
6.3.8 Other features of the simulation model	104
6.4 Expanding the original model to include <i>TDRP</i>	104
6.4.1 Overall Structure of the expanded model	104
6.4.2 Conditions for building up queues	104
6.4.3 Time-Dependent Road Pricing Model	105
6.4.4 Generalized Cost Function	107
6.4.5 Departure time choices model	108
6.4.6 Checking for convergence	108
6.4.7 Final output	109
6.5 Interaction between nodes and its effect on the value of <i>TDRP</i>	110
6.5.1 Interaction between nodes (case with no merging/no diverging flow)	110
6.5.2 Introducing merging and diverging flows at nodes	114
6.5.3 Procedures for adapting the charges at each individual node	115
6.5.4 Suggested algorithm for a typical network	119
6.5.5 Summary and implications for practical applications	120
7. <u>EVALUATING THE DIFFERENT EFFECTS OF <i>TDRP</i> ON A TRAFFIC NETWORK</u>	122
7.1 Introduction	122
7.2 Description of the case study used for evaluation	122
7.3 The base case simulation results	125
7.4 The effects of congestion on the base case results	127
7.5 The different effects of <i>TDRP</i> on the travel conditions	131
7.5.1 The different effects of <i>TDRP</i> under scenario (I)	132
7.5.2 Comparison between the different <i>TDRP</i> scenarios	142
7.6 Summary	146
8. <u>COMPARATIVE AND SENSITIVITY ANALYSIS</u>	147
8.1 Introduction	147
8.2 Comparative Analysis	147
8.2.1 Comparing <i>TDRP</i> with Distance-Based Charging System	149
8.2.2 Comparing <i>TDRP</i> with Time-Based Charging System	151
8.2.3 Comparing <i>TDRP</i> with Delay-Based Charging System	154
8.2.4 Comparing <i>TDRP</i> with Flat Charging System	157
8.2.5 <i>TDRP</i> vs. alternative charging systems under different demand levels ..	158
8.3 Sensitivity Analysis	162
8.3.1 The effects of changing the work-time flexibility on the results	162
8.3.2 The effects of changing the values of the schedule delay parameters	165
8.4 Summary	172

	<i>Page</i>
9. <u>CONCLUSIONS AND FURTHER RESEARCH</u>	173
9.1 Summary and Conclusions	173
9.2 Limitations and further research	177
REFERENCES	179
APPENDEX (A): Analytical Approximation for <i>TDRP</i> Model	188
APPENDEX (B): Network Traffic Data	191

CHAPTER ONE

1. Introduction

1.1 Urban Traffic Congestion.

1.2 Different Policies of Traffic Restraint.

1.3 Road Pricing: Basic Concept and Rationale for Further Development.

1.4 Time-Dependent Road Pricing (*TDRP*).

1.5 Setting the Study Objectives.

1.6 Structure of the Study.

1.1 Urban Traffic Congestion

Urban traffic congestion has long been recognised as a major problem in the developed world, but the rapid explosion of urbanisation in the third world now makes it a global issue. In fact, the dramatic increase in road traffic in urban areas is leading to serious problems of congestion arising from limitations in capacity of roads and from effects of street parking on the movement of vehicles (OECD, 1973). On the other hand, economists believe that the immediate cause of traffic congestion in urban areas is the failure to share the full costs of urban transport, which in turn arises from the failure of road vehicles to pay the full cost in the form of congestion imposed on other road vehicles (Reynolds, 1961).

Whatever the cause is, traffic congestion in town and city centres is considered as an economic problem of increasing concern. Both employees and employers suffer from long, unreliable journeys to work. Over-crowded roads seriously sometimes fatally, hamper the emergency services and substantially increase the cost of deliveries to offices, shops and restaurants (Hewitt, 1989). This consequently leads to considerable social and economic loss in the form of additional operating expenses, loss of time, reduced safety, undesirable environmental effects (e.g. air and noise pollution) and a lower standard of comfort and convenience for travellers.

One way of attempting to solve the problem of urban traffic congestion is to build more roads to increase the capacity of the existing roads, but there are practical limits due to the topography, existing buildings, and economic considerations (Pretty, 1988). Buchanan (1952) has reported that in order to decide how much resources should be devoted to highways and streets, society must choose between providing a structure which is too large in off-peak periods and one which is too small in peak periods. It seems certain that if enough resources were to be devoted to highways construction to reduce congestion to acceptable proportions in peak traffic periods, over-investment in highways would be present. A highway system of compromise size would appear preferable. This would mean that some highway resources still wasted in off-peak periods. However, many American cities have tried to achieve this approach by building vast road works the like of which no other countries could afford, and yet they have not rid their cities of congestion (Roth and Thomson, 1963). Thus, the reduction of congestion by road expansion can never represent an adequate solution to the problem of traffic congestion in major urban centres. That is why policies aimed at diverting and restraining traffic are of increasing concern of transportation planners and economists.

Over the years, however, as the congestion problem became more widespread in cities throughout the world, several ways of solving urban traffic problems by imposing some restraints on road traffic were considered. These different policies (or measures) of traffic restraint are discussed in some detail in the next section.

1.2 Different Policies of Traffic Restraint

The aim of restraint policies, as defined by the Department of Environment (1976), is simply to discourage the use of private cars in circumstances where they impose a high social cost on the community and where alternatives are available. On this basis, they work by changing the observed behaviour of travellers in making modal choice, routing, destination, and other travel decisions including the decision to make the trip at all (OECD, 1973). In general, the common objectives of traffic restraint policies include the following: –

- to improve traffic efficiency;
- to improve accessibility;
- to improve the environment;
- to improve the distribution of profits; and
- to improve the utilisation of resources.

Pretty (1988) has classified the traffic restraint policies, designed to ease urban traffic and to make the best use of available resources, into four main categories as shown below and illustrated in Figure (1.1): –

1. Restraint on ownership.
2. Increasing running cost (User Taxes).
3. Increasing parking charges.
4. Control of moving vehicles.

Each of the above policies is briefly considered in turn.

1.2.1 Restraint on ownership

An indirect but effective restraint on private car usage is to inhibit vehicle ownership through high import duties, purchase taxes or registration and annual licensing fees. These can be effective in reducing the severity of transport problems by raising ownership costs and making cars systematically less available (Armstrong-Wright, 1988). A primary disadvantage of vehicle ownership restraints is that they are insensitive to location and time of use. They also encourage the use of old cars, which are likely to be less safe and breakdown more frequently, thereby adding to congestion level. The latter could be resolved by using special incentives to replace older cars or inspection procedures can be implemented. Similarly, high penalties for traffic infringements and obstructions may suppress journeys in congested areas by older, less reliable vehicles and thus lower any adverse impact on safety and traffic flow.

As car ownership and congestion have grown, a few countries have introduced taxes to control the increase in car ownership and thereby the likelihood of congestion rather than congestion itself. However, these are blunt instruments as once the owner has paid the tax there is no reason to avoid the use of congested roads. On the contrary, there may be an incentive to make full use of the road system in order to get the value from the high taxes paid (The Chartered Institute of Transport, 1990).

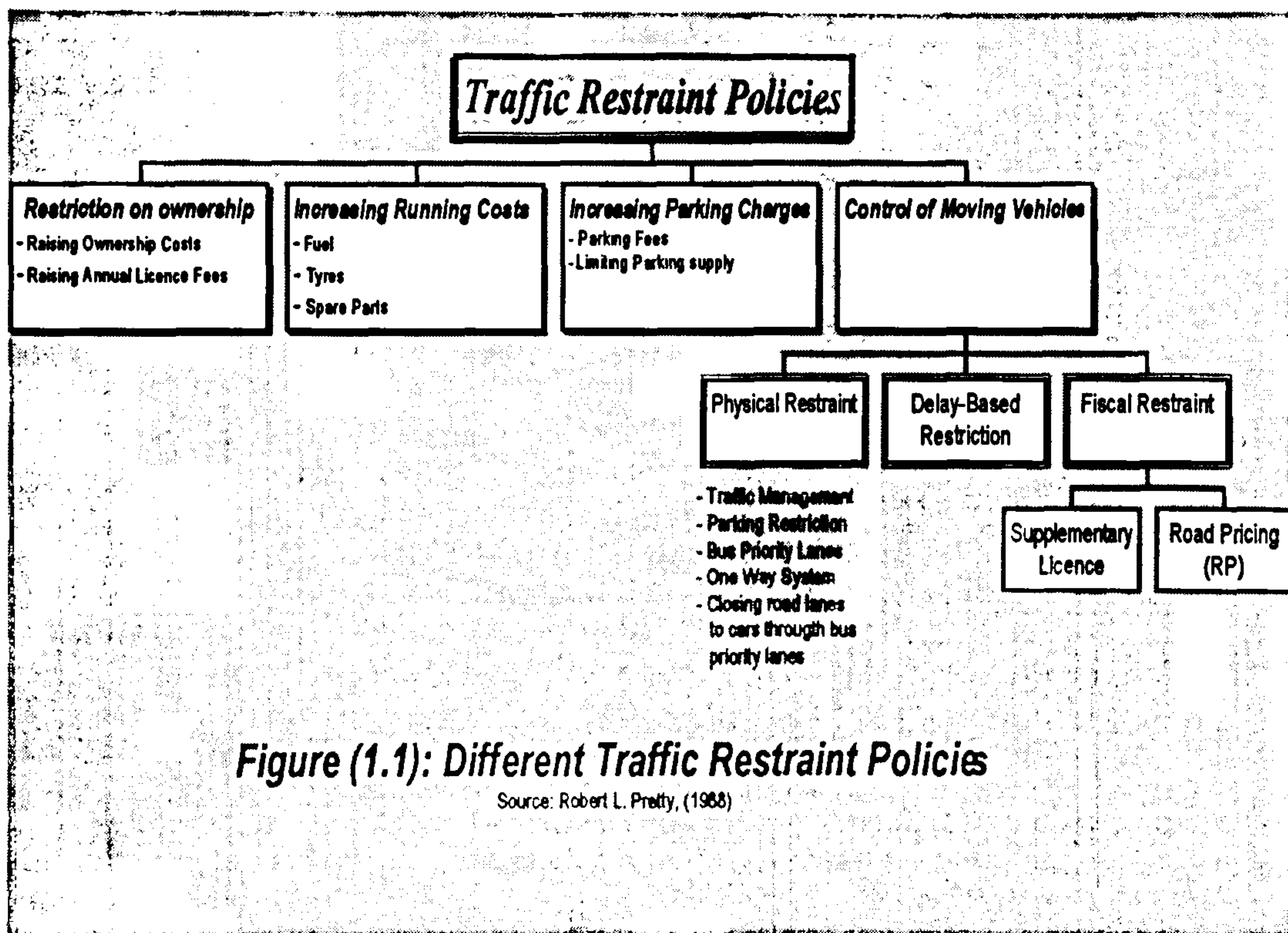
While restraints on ownership may not provide the optimum or equitable solution to congestion, they can usually be applied relatively quickly within the framework of existing legislation and thus can provide an urgent stop-gap while more effective measures are being devised and implemented. This approach has been adopted by Hong Kong to deal urgently with serious congestion (Hong Kong Government, 1985).

1.2.2 Increase Running Cost (User Taxes)

Vehicle use may be restrained through user taxes imposed on motor fuel, tyres, spare parts, etc., thus adding to the running costs in relation to the distance travelled (World Bank, 1975; Churchill et al., 1972). While potentially restraining total use, these taxes will not affect where or when vehicles are used. In particular, they do not differentiate between use during peak and off-peak periods in congested or non congested areas.

They also have some side effects which mitigate against their use. For example, although a tax on tyres may provide a restraint directly related to the amount of vehicle use, it can give rise to the use of unsafe tyres. Similarly, a tax on spare parts may lead to poor maintenance, breakdowns and accidents, and hence tend to increase rather than reduce congestion. To be effective, these particular taxes would need to be associated with comprehensive vehicle inspection programmes which may prove to be difficult and costly to introduce (Armstrong-Wright, 1986).

Like restraint on ownership, user taxes are insensitive for restraining vehicles in congested areas (May, 1983). Fuel tax, for example, is related to road use as the more a car is driven the more fuel (tax) it will pay. Moreover, heavier vehicles use more fuel and fuel consumption rates are higher in congested conditions. Therefore, fuel tax payments are related to wear-and-tear on the road system and congestion. However, the Chartered Institute of Transport (1990) pointed out that this relation is very weak and insufficient as a form of restraining congestion.



1.2.3 Increase Parking Charges

Another effective measure of traffic restraint, to achieve better use of limited road space, is the possibility of increasing parking charges or limiting the supply of parking spaces. This has been used as a measure of restraining private motor vehicle trips since at least the 1950s. This restraint can be achieved by prohibiting or charging for on-street parking. Accordingly, this could actually increase the capacity of streets for moving vehicle traffic. On the one hand, if the highway authority is also responsible for the supply and control of off-street parking, then the amount of parking in central areas and the charges can be determined so as to restrain trips to these areas. On the other hand, other than property taxes, there can be no control over privately owned parking areas for private use, only areas for public use, so that authority will still be well short of total control. However, within the areas for public use, the charges levied can be set comparatively low for an initial short period, and progressively high for larger periods. Thus, commuters, who are long-term parker, would pay substantial charges and be discouraged from parking in the area. On the contrary, people with business appointments and shoppers who need short-term parking are encouraged; they substantially benefit from both lower charges and greater availability of parking places because of more rapid turnover.

In practice, parking restraints are inclined to give rise to substantial increases in illegal parking and therefore need to be accompanied by strict enforcement of regulations. High parking charges may also encourage additional trips during the peak period (for example, people being driven to work in the morning by other family members, who then return home and repeat the journey in the evening). In addition, parking control may create unproductive trips, for example, when drivers circulate in the Central Business District (CBD) waiting for passengers or looking for parking places. These previous side effects may add substantially to congestion if they are allowed to develop. Moreover, parking control is a blunt tool for tackling congestion problems since it has limited effects on through traffic, which is often a primary cause of congestion in city centres.

As a result of the above discussion, parking control may be more effective as a restraint policy when it forms part of a more comprehensive demand management scheme as in the case of Singapore (see Watson and Holland, 1978).

1.2.4 Control of Moving Vehicles

With little success possible for the restraint of parked vehicles, attention can be turned to the control of moving vehicles. This restraint can be subdivided into three main categories: –

- (i) Physical Restraint.
- (ii) Delay-Based Restraint.
- (iii) Fiscal Restriction.
 - (a) Area Licensing Schemes (ALS).
 - (b) Road Pricing (RP).

Each of these subdivisions is discussed in turn below.

1.2.4.1 Physical Restraint

This can be interpreted in a number of ways; such as regulatory control, traffic management, closing road to traffic through pedestrianization, closing road lanes to cars through bus priority lanes and one way streets.

Two types of regulatory control are worthy of mention; the first involves the allocation permits based on need and the exclusion of vehicles without permits. This system was proposed for London by Lane and Hodgkinson (1976). The basis for permit issue can take several forms (May, 1986); they can be limited to easily identified users such as the disabled or doctors, or they can be allocated implicitly to those making more efficient use of their vehicles, as in the exemption of vehicles with four or more occupants from area licensing charges in Singapore (see Watson and Holland, 1978). The second type is altogether more simple; it is the "Odds and Evens" system that operates in Athens and Lagos in which odd-numbered vehicles are permitted to enter the controlled area during the working day on certain days of the week, and even-numbered vehicles on the others.

In general, the most serious side effect of physical restraint is that they transfer motor vehicles to other streets, thereby, they could become congested. Furthermore, in particular, they do nothing to counteract people's willingness or inability to appreciate the congestion that their trip-making imposes on others (Button, 1982).

Some success in Europe with such policy has been reported by May (1983).

1.2.4.2 Delay-Based Restraint

Congestion, on its own, imposes restraint through the penalties of extra travel time and uncertainty. May (1986) suggested that organised delay would provide more certain and equitable form of restraint, because all users would be equally penalised. Time penalties have been used in freeway access control schemes to relocate queues and hence increase the efficiency of vehicle movements (Everall, 1972). The same principle has been suggested by Coventry and Dickinson (1977) using technique of "gating", for arterial and has also been used effectively in Southampton Bitterne Road Scheme to improve efficiency of person movements (see Department of the Environment, 1960).

Another way of applying this policy of restraint, as addressed by Pretty (1988), is to draw a cordon around a central city area and access can be provided according to priority system headed by buses. Admission to the central city area for other motor vehicles (i.e. rather than buses) depends on the capacity of the street system as well as the capacity and occupancy of parking facilities available in the area. In some cases, private cars with greater number of occupants may be given priority over

other vehicles. Through traffic, of course, has the option of travelling around the cordon and so avoiding the central city. However, it may be impossible to provide enough parking facilities for terminating traffic, so that private car users may be forced to park outside the cordon and enter the central city area by bus. In effect, the system is making the road users pay for a commodity that everyone has in equal quantities: time.

An experiment along these lines was tried in Nottingham, England, in 1970s, but it failed and was abandoned as a result of the lack of storage space needed for delayed vehicles and violation of the controlling signals.

Pretty (1988) reported that the inherent difficulty with this restraint policy is that the places where congestion occurs may cause blockages of routes and increase journey times for travellers who were not planning to pass through or into the central city area anyway.

1.2.4.3 Fiscal Restraint

Of all the forms of restraint on moving vehicles, fiscal restraints have received the most detailed study. There are two main forms of charging motorists directly for the use of roads. One, usually called Supplementary Licensing, has been implemented in Singapore. The other is Road Pricing which is the subject of this study. Each of these two forms is discussed below.

- **Supplementary Licensing (SL)**

Supplementary Licensing (SL) is also known as Area Licensing or Cordon Charges. The system of supplementary licensing implemented in Singapore requires car drivers with low-occupancy vehicles to pay an extra daily charge if they want to drive into the central area during weekday peak-hours. A special sticker has to be displayed in the windscreen and cars are checked as they drive past enforcement points. If a car fails to display the appropriate sticker, its licence number is noted and a fine posted to the owner. The system encourages greater use of public transport and shared private cars, and discourages unnecessary journeys during rush periods.

For the system to be successful, public transport has to be adequate and suitable for handling motorists and passengers diverted from private cars. Also, alternative routes of adequate capacity have to be available for traffic that is diverted from the congested area. Moreover, to facilitate area licensing, the road layout has to permit the isolation of the restricted area. Hewitt (1989) reported that the main problem with SL lies in defining the boundaries and in the 'boundary effects'. The boundary effects could be worst when SL is only checked at the boundary itself, as in Singapore. It would, however, be possible to enforce SL schemes throughout the area by considering driving a car anywhere within the licensing area without possessing and displaying an up-to-date SL an offence.

More recently, Oldridge (1990) criticised the charging structures of SL as being inflexible and inequitable in imposing the same charge on short and long journeys, and disruptive in that they can lead to congestion on boundary routes.

The SL system applied in Singapore is briefly discussed in section 2.7 in the next chapter. For more detail see Watson and Holland (1978), and May (1975, 1978).

- **Road Pricing (RP)**

Road Pricing is essentially a form of traffic restraint, but it differs from other measures such as physical restraints, fuel tax, parking control, or supplementary licensing by directly affecting the root cause of congestion: the use of the congested roads themselves. It also has a number of obviously attractive features as a means of traffic restraint policy. It could, in theory, levy charges according to the practical level of congestion on each road link or sub-area; discriminate between different classes

of vehicles, vary charges at different times of day; levy parking charges as well as charges for movement on the road (DoE, 1976).

Of course, road pricing will have to rely on advances in electronic technology in order to meet all its requirements. As with Electronic Road Pricing (ERP), for example, motoring outside congested areas or busy times of the day can be made cheaper than under alternative policies. This is particularly true in Hong Kong where the main alternative to ERP is continuation of very high car-ownership taxation, which is insensitive to vehicle use (Dawson and Catling, 1986).

Studies of traffic restraint in the UK over the last thirty years have almost all concluded that selective road pricing is the most effective and efficient way of managing urban traffic congestion. Most studies have shown that there are also significant environmental benefits from its introduction and that there are further advantages to be gained from implementing road pricing as a part of a wide package of environmental and public transport improvement proposal (The Chartered Institute of Transport, 1990).

For these reasons and many others, this study is primarily concerned with road pricing and the implications of electronic road pricing in particular. Thus, the theory of road pricing as well as its historical background are reviewed in some detail in the following chapter. Meanwhile, the basic concept of road pricing and rationale for further developments are considered below.

1.3 RP: Basic Concept and Rationale for Further Development

The essence of congestion problem is that each individual road user imposes certain costs or disadvantages on others who are using the road at the same time. Moreover, the unrestricted use of congested roads is inefficient because users usually base their decisions to use the road on their own private costs, neglecting these costs they impose on others. Road pricing, however, has the potential to improve the efficiency of the road network and to reduce the congestion which accompanies inefficient use. The basic concept of RP is to influence the decision of the road users before they use the road network in the most congested times and places by charging them a part or all the external costs they impose on others. The charges incurred by drivers will depend on how they use their vehicles such that drivers are encouraged to minimise (reduce) the use of their vehicles in congested conditions.

In practice, the possible reactions of motorists to charges, as addressed by Petty (1988), include the following: –

- to make the trip and pay;
- to not make the trip;
- to make the trip by different route;
- to change the time of travel given that the demand is not steady state;
- to change origin or destination;
- to form a car pool; and
- to change mode away from private car travel (i.e. to switch to use public transport).

Thus, any motorist who thought his loss from being diverted would be greater than the size of the road charge would pay the price. In contrast, the loss to any motorist who refused to pay could be considered less than the gain due to his absence. The charge (or price) paid also shows how valuable road space is, and therefore the strength of the case for expanding it.

For a system to be reasonably effective, Armstrong-Wright (1986) stated that charges to individual users would need to be related to the amount of use of congested roads and the degree of congestion. It is thus necessary to vary the charges for different locations, times and vehicles. Charges in all these cases may prove difficult to calculate.

It has been reported by Small et al (1989) that most economists believe that congestion pricing should be targeted on congested roads at peak hours only. But, this pricing system (i.e. the system of peak/off-peak road pricing differential) takes very little account of the enormous variation in congestion costs caused by vehicles over different parts of the road system. This is because costs caused by vehicles on each others vary greatly from one time and place to another throughout the road system and by the time of day. In other words, it can be recognised that this system is inequitable between motorists, since the same charges are levied on car use during the peak period regardless the fluctuations in demand during that period in different locations.

On the other hand, professor R. J. Smeed (1964) suggested that the most efficient price system might appear to be the one in which price varied with cost on every road at every moment of the day. He also added that this system presupposes that road users are able and willing to take account of such a complicated system which could be impractical. Buchanan (1952) also pointed out that a comprehensive and highly differentiated system of charge would be required to secure the ideal pricing structure. Then, he added that such a system would be completely unworkable from an administrative point of view and would be uneconomic besides.

In practice, the use of Electronic Road Pricing (ERP) systems offers the best solution from the point of view of efficiency, flexibility, and ensure an equitable congestion charging policy. One of the most attractive feature, according its supporters, is the power it gives its controllers to adjust charges over time. Thus, it is possible to impose different charges on drivers depending on the time and the particular road they use. However, more sophisticated electronic charging systems are already reality and a recent experiment in Hong Kong also demonstrated the technical feasibility of ERP (The Chartered Institute of Transport, 1990).

Based on the above discussion and the higher technology of ERP, it is worth expanding the use of ERP to include a time-varying charging system. The charges for road users, under this system, are completely differentiated with the time of day and the particular location they use. Thus, this system of charging will be capable to bring prices more into lines with costs imposed by the users on each other. This system of charging is called Time-Dependent Road Pricing (*TDRP*) and discussed in the next section.

1.4 Time-Dependent Road Pricing (*TDRP*)

In practice, it is probably true to say traffic junctions are the bottlenecks in urban transport systems and one of the most important source of traffic congestion. And unless, these bottlenecks are tackled, there is little hope of alleviating the problem of traffic congestion in urban areas. Larsen and Ranjerdi (1990) have reported that the whole time profile of arriving vehicles within a period, may be critical to the delays at the bottleneck. They also added that the delay a specific vehicle causes for other vehicles may last for a long time after this specific vehicle has left the bottleneck.

Besides, Goodwin and Jones (1989) pointed out that the detailed consideration of the behaviour of queues at intersections (bottlenecks) indicates that the first vehicle in the queue is causing more delay than the end vehicle. Accordingly, each vehicle should be theoretically charged at a different rate.

Based on the above two considerations, the system of Time-Dependent Road Pricing (*TDRP*) attempts to reduce traffic congestion at traffic network in urban areas by levying charges on individual users related to the actual amount of delay (i.e. the queuing delay) they impose on others at every individual bottleneck (i.e. traffic junction). On this basis, since the traffic conditions (i.e. demand and/or capacity) vary continuously over junctions by time of day and accordingly the queue lengths, then the delay imposed by individual users on one another will be time-dependent. Typically, the charges will be responsive to the minute to minute local conditions over the bottleneck in question. Thus, the congestion could be measured and paid for as it occurs and there is no predetermination of where or when it occurs. As a result, charges can be highest when congestion is worst, and reduced at other times when there is spare road capacity. This, in turn, will indicate the strength of demand for road space in different places and at different times of the day.

In effect, motorists will be selectively charged on the times and locations where congestion needs to be reduced. Although, this system does not distinguish individual from (nor encourage) car pooling, it leaves individual motorists alternatives and determines whether it is worthwhile to make such a particular trip to a particular place at that particular time. Therefore, it is desirable that intending drivers should know the charges payable before making a journey. For this reason, charges should be directly related to the expected conditions (for example, from the previous day) and drivers should be informed of charges by automatic roadside signs at the approaches to the locations of charges.

Of course, the only practical way of charging individual in this system is by using the technology of ERP. This technology has been demonstrated by the Hong Kong ERP-scheme as technically viable, administratively feasible and of a significant potential benefit in helping to deal with heavily congested road network.

The theoretical basis and the approach used for setting the charges in this system (i.e. *TDRP*) are explained in Chapter (3).

1.5 Setting the Study Objectives

Having discussed the state-of-the art as well as the basic concept of road pricing, and the rationale for further development, this section identifies the study objectives. Keeping in view the aim to investigate the various implications of applying *TDRP* in reality, the study objectives are presented as follows: –

The main objective is to: –

Develop a system of charging in which charges are differentiated by time of day, according to the queuing delay imposed by individual vehicle on other vehicles at a specific location. The distinctive feature of this system is that congestion will be measured and paid for as (whenever and wherever) it occurs.

In order to assess the likely effects and the prospective advantages of this system, the following two secondary objectives are defined: –

(I) Develop a network traffic assignment model that embraces route choice, departure time choice and the TDRP model. This model could be capable of assessing any other charging policy.

(II) Evaluate the different effects of TDRP system on the user equilibrium, departure pattern, travel time, schedule delay (i.e. the difference between the desired arrival time and the actual arrival time to the destination), and the overall average speed.

Finally, to demonstrate the significance of this work, several simulation experiments are conducted with the aim to: –

- *Carry out a comparative analysis with other methods of charging to show the distinctive features of TDRP.*
- *Check the sensitivity of the results under TDRP system to different travel aspects, e.g. demand level, work start time flexibility, and the key parameters of the user cost function.*

1.6 Structure of The Study

This section sets out the development of the chapters of this study and their contents. This may throw some useful light on the structure of the study which consists of nine chapters (including the introduction chapter).

This structure is organised as follows. The next chapter (Chapter 2) is a background review of road pricing history and theory. It also discusses the current pricing policies as well as the different methods of collecting the road charges.

Chapter (3) identifies the approach used in this study to estimate the amount of queuing delay imposed by every individual on one another. This has been done by reviewing the theory of queuing and delay as well as the different solutions of the time-dependent queuing and delay problem. Then, the approximate solution of the time-dependent queuing problem is used to derive a mathematical model for the amount of charge imposed by every individual on other vehicles as a function of the departure time of this individual from the bottleneck.

The existing models for temporal distribution of the peak period demand for a bottleneck are reviewed in Chapter (4). This chapter also examines the User Equilibrium “*UE*” under *TDRP* and gave an analytical demonstration to show how *TDRP* could lead to system optimal “*SO*” (i.e. it is equivalent to the optimal toll). Chapter (5) then puts a framework for the simulation models used for a single bottleneck as well as a typical network to evaluate the different effects of *TDRP* system. The Chapter also covers the simulation solution and its results for a single bottleneck under deterministic and stochastic user equilibrium.

The network traffic simulation model as well as the importance of the phenomenon of interactions between nodes and its impacts on the value of *TDRP* are presented in Chapter (6). Chapter (7) describes the typical traffic network used for the evaluation and discusses the output results of the simulation model under different demand levels as well as before and after *TDRP*. The results of the comparative analysis between *TDRP* and the other methods of charging are covered in Chapter (8). The chapter also represents the results of the sensitivity analysis of the results under *TDRP* to different travel aspects (e.g. demand level, work start time flexibility, and the key parameters of the user cost function).

Finally, Chapter (9) contains a brief summary and the final conclusions of the study. It also outlines the study limitations and proposes a number of areas for further research.

CHAPTER TWO

2. Road Pricing Background

2.1 History of Road Pricing

- 2.1.1 Previous Studies.
- 2.1.2 Economics of Road Pricing.
- 2.1.3 Is Road Pricing fair?

2.2 Paying for Roads

- 2.2.1 Why pay for roads?
- 2.2.2 Who will pay the road user charges?
- 2.2.3 Who will benefit from road user charges?

2.3 The Costs of Roads

- 2.3.1 The costs of providing the roads.
- 2.3.2 The costs arising out of journeys.

2.4 Pricing Strategies (Policies)

- 2.4.1 Arbitrary pricing.
- 2.4.2 Average cost pricing.
- 2.4.3 Marginal cost pricing.
- 2.4.4 Monopoly pricing.

2.5 Theory of Road Pricing and its Drawback

- 2.5.1 Theory of road pricing.
- 2.5.2 Determining the size of charge in practice.
- 2.5.3 Drawback of the theory and the line of development.

2.6 Practical Issues Involved in the Implementation of RP

- 2.6.1 Methods of charging.
- 2.6.2 Exemptions for certain vehicles or persons.
- 2.6.3 Spatial charging structures.
- 2.6.4 Temporal pattern of charges.
- 2.6.5 Other aspects.

2.7 Singapore, Hong Kong and Norway Experiences with RP

2.8 Recent UK Studies Concerning RP

(RP systems proposed for Cambridge)

2.9 Summary

2.1 History of Road Pricing

2.1.1 Previous Studies

Road Pricing (RP) has an established history in the literature of transport economics and has attracted the intermittent attention of policy makers for many years. In fact, it was a French engineer, Dupuit (1844), who laid the foundation for RP. Although, Ellet (1840) and Dupuit (1849) addressed the issues of pricing and investment of transport facilities many years ago, the starting point of much of the discussion can be traced back to the writings of Pigou (1920) and Knight (1924). The former developed the general case for marginal cost pricing (discussed later in this chapter under section 2.5) and the latter advocated the imposition of congestion taxes as a means of alleviating congestion and encouraging the more efficient use of road space.

Over the years, however, as congestion problems become more widespread in cities throughout the whole World, the interest in RP as a means of traffic restraint and management increased. This interest was simulated more particularly by the work of such as Beckmann, McGurie, and Winsten (1956), Walters (1961 and 1968), Jonhson (1964), and Beesly (1969).

In Britain, in the early 1960s, the first and the best known study was carried out by a committee chaired by professor R. J. Smeed set up by the Ministry of Transport 1962. This study set out the basic theory of RP and how the benefit could be estimated: it laid down 17 desiderata (see Smeed, 1964) and compared direct charging for road use with other possible congestion taxes such as a differential fuel tax, poll tax on employers in congested areas, parking tax and daily licences.

In the USA, Alan Walters (1968) and Gabriel Roth (1967) argued that it is better to adjust travel demand to a limited road network through the application of congestion charges; variable charges which raise prices and discourage travel during peak hour traffic. One of the initial attempts to empirically estimate the magnitude of highway congestion charges was the study by Walters (1961). He recommended a 10–15 cents per mile charge for peak periods in the more congested areas in conjunction with a general urban fuel tax of 33 cents per gallon.

In the early 1970s, Wigan and Bamford (1973) showed that even at a low elasticity of demand under RP, there will be a redistribution of travel within the capacity of links of the network. By the mid 1970s congestion was out of the root cellars of economic journals and into the consciousness of transportation planners (see Elliott, 1986). The World Bank was urging congestion charges on Third World Countries, the Smeed Committee had recommended them for London in the 1960s, and the US Urban Mass Transit Administration, with research support from the Urban Institute, was also urging congestion charges on any American jurisdiction willing to risk their political side-effects. These agencies had already begun the task of translating the economists' theories into practical strategies. They had high-technology strategies, based on Automatic Vehicle Identification (AVI); low-technology strategies, based on dated stickers; mid-technology strategies, based on stickers that would change colour after a day of use. They had also estimates of what each strategy might cost (Bhatt, 1974 and Elliott, 1976).

In 1975, the first genuine road pricing system in the world was introduced in Singapore in the form of an area licensing system and has proved to be practically successful (see Watson and Holland, 1978). This system is briefly discussed later in this chapter under section 2.7.

In the late 1970s, Boardman and Lave's (1977) work focuses on the estimation of speed-flow curves, which are the foundation for deriving the optimal congestion charges. Inman's (1978) work is also concerned with estimating the speed-flow relationship. He developed a new approach to the specification and estimation of the speed-flow relationship, called a generalised congestion function (GCF). The GCF contains all of the most commonly used speed-flow relations as special cases (see Morrison, 1986).

De Vany and Saving (1980) incorporate uncertainty into their theoretical model of highway pricing and investment in the form of random demand: i.e. at any given price, the number of trips demanded is not fixed, but is a random variable. Kraus (1982) also incorporates uncertainty but in a different way. In his model, the demand function itself is non-stochastic, but depends on a parameter that is unknown to highway planners who must determine the optimal charge and capacity.

Martin Wachs (1981) argued that pricing will play a central role in transport policy and will be an important tactic to achieve the objectives of more efficient equitable use of existing transportation investment. Focusing on the reactions of policy makers, and the role of analysts, business interests, and the media in determining the study outcomes, Thomas Higgins (1981) suggests that risk management strategies will help to implement RP. On the other hand, Button and Pearman (1983) raise resources allocation issues on the socio-economic aspects on RP. They consider that the basic theory is technically correct, but it is doubtful whether this analysis will be accepted as a basis for policy making in the real world solution.

In responding to the problem associated with RP raised by Button and Pearman, Else (1984) argued that because the value of time varies among different groups using the same congested road, optimal RP requires price discrimination between groups such as those with the lowest value of time pay the highest charge and suffer the greatest losses. He considered that most governments' lack of interest for RP stems as such from political considerations as the more technical problems of applying it (Fong, 1985).

The experiment in Hong Kong has demonstrated the technical possibility of Electronic Road Pricing ERP, (see Hong Kong Government, 1985). Also the first, rather simple, road pricing schemes have appeared in European Cities of Milan (Buchanan and Partners, 1986), Bergen and Oslo (see Larsen, 1988). More extensive and sophisticated schemes are planned for Stockholm and Randstat area of Holland (see Ramerjdi, 1989).

The main conclusion emerges from these previous studies, is that RP has a very long history and could be the most promising solution to the problem of traffic congestion in urban areas.

2.1.2 Economics of Road Pricing

From an economic point of view, a proposal for road pricing may be made on several different grounds. Essentially, there are five arguments (Hewitt, 1989): –

Firstly, the congestion costs imposed by each driver on others should be borne by the driver 'internalising the externalised costs', thus ensuring that road space is used as efficiently as possible. In this view, RP is seen primarily, or solely, as an instrument to reduce congestion. Secondly, RP offers a way to finance new road building, or road improvements. The best example of this is Bergen 'toll ring' in Norway, built for sole purpose of revenue raising. Thirdly, RP should be used as an incentive to generate private investment in road-building, with RP revenues going to private investors. This argument is particularly associated with the Adam Smith Institute, and was adopted by the Department of Transport (UK) in the conclusion paper, "New Roads by New Means", which proposes private toll roads, running near the public highways with which they could compete. Fourth, on the basis of 'the polluter pays principle', RP should be used as an environmental charge, designed to reduce private car use in the interests of reducing pollution. As with congestion costs, the external costs imposed by car-users on the environment should be internalised in the price of using a car. Fifth, RP in city centres should be used to raise revenue for public transport improvements.

Although all the arguments may lead to apparently similar conclusions, they are quite different. The first and fourth arguments see RP as a means of reducing private car use in the interests of reducing congestion or pollution; the fifth argument is usually combined with them. In contrast, the second and third arguments use RP as a route to more roads and, thus, greater car use. From one perspective, RP is an incentive to road-building; from the other, it is a disincentive to excessive car use.

However, RP could be used as a charge designed to reduce both congestion and pollution. Where congestion is severe, RP is justified on the classic economic grounds of internalising externalised costs and promoting more efficient use of road-space. From the standpoint of environmental economics, environmental costs must be taken into account. Road pricing is also an environmental charge; a specific application of 'the polluter pays principle'. Therefore, RP is best described as a 'congestion and pollution charge'.

It is interesting that in Holland, RP is designed not only to reduce congestion and pollution, but also to raise revenues for the creation of three new privately financed road tunnels. The central objectives of the Western Region Access Plan are, however, to eliminate congestion costs and control the growth of motoring, particularly in urban areas and during the rush-hours. Electronic Road Pricing 'ERP' is seen by the Dutch Government as making a vital contribution to both objectives.

2.1.3 Is Road Pricing Fair?

The most common objection to RP is that it is unfair; as the rich car user, or the driver subsidised by his company, pays while the poor car user is priced off the road. The car users who can afford to pay get the benefit of less crowded roads, while the poor car users either have public transport or, if public transport is not available, abandon the journey or find the money by cutting back on something else.

RP in any form, of course, has the greatest impact on the poorest car-users. But, this is not the whole story since the effects of RP do not depend simply on the user's income. For example, people who place a higher value on their time (not all of whom fall into the higher income groups) may be willing to 'pay and stay'. Those who would have to pay the highest charges, by passing more charging points, might be deterred from using their cars in peak hours. People who could change the time of their journeys in order to reduce the charge, or avoid it altogether, or who could change reasonably easily to public transport would obviously be more likely to do so than with fewer alternatives.

In a study for road pricing in Central London, MVA/Buchanan report (1989) showed that the benefits are enjoyed not only by remaining road users— including those who pay for the benefit of less congestion through the charge— but also by public transport passengers, cyclists and pedestrians who, of course, would pay no charge at all. Transport policy quite often ignores these groups and include some of the poorest members of the community. Children, young people, women (particularly those caring for dependants) and elderly people are disproportionately likely to use the buses or to walk. They would be amongst the main beneficiaries of any effective measure of restraint, including RP. The MVA/Buchanan report has also concluded that fears that RP in Central London would be unfair are unjustified and it is the absence of RP, not its introduction, which is unfair.

Also Hewitt (1989) advocated the above opinion using a different approach. He argued that in the absence of RP, car drivers on congested roads are imposing costs on other people for which they are not charged themselves. This, in other words, means that drivers are being subsidised. Some of the subsidy comes from taxpayer (who pays, for instance the costs of National Health Service NHS); some comes from businesses and their customers, on whom many of the costs of congestion fall; some comes from private residents affected by noise, local pollution and the effect of major roads on the value of nearby houses. This subsidy is going to the better-off members of the community— those who use cars— not the poorest. Since RP on congested roads removes this unfair subsidy to car drivers, it would, in fact, be fairer than the present situation.

2.2 Paying for Roads

This section discusses, from an economic view point, why pay for roads?; who will pay the road use charges?; and who will benefit from them?.

2.2.1 Why paying for roads?

The question of 'why pay for roads?' has been answered by Roth (1966). He suggested that the principles on which the allocation of the vast majority of goods and services, can be usually applied to the allocation of road space. Thus, the objectives of charging for the use of roads are those that taken for granted in the case of most other commodities: –

- To ensure that the best use is made of the existing facilities.
- To indicate where– and to what extent– the existing facilities need improvement; and
- To provide sufficient funds– no more no less– to cover the costs of the facilities for which the charges are made.

According to the first two objectives, a pricing system should encourage the best use of existing roads and it should provide information on the need for improvement. Roth also indicated that the third objective might be disagreed by many readers. Why should the revenues collected for road use be equal to the costs of the roads?. Is there not a case for treating roads as a social service, and supplying them at prices that do not cover their costs?. Conversely, might road use not be regarded as a source of general revenues, and charge more for it than its costs?.

Perhaps the question of subsidies is the easier of the two to deal with. Road users come from all classes; those who use road most (i.e. the car-owners) belong to the better-off rather than to the poorer sections. Therefore, there is no evidence that general subsidies to road users would improve the distribution of income or bring any benefits to people in most need of them. Until such evidence is forthcoming there is no general case for providing roads at prices that do not cover their costs.

Should the price of road use include a contribution to general revenues?. The stock answer to this question is that 'the government can do what it likes' and that the question is one for politicians, not for economists. But the economic aspect is, of course, still important (Roth, 1966).

The following two sections answer the questions of who will pay the road use charges? and who will benefit from the road user charges?.

2.2.2 Who will pay the road user charges?

Kulash (1974) indicated that a motorists' willingness to pay road user charges is related to many things: the availability of other modal alternatives; the feasibility of travelling at other times of day; the amount of time saved which depends, in turn, on where the traveller is going to or coming from; and the value of time saved, which is closely tied to income level. The relation between factors of this sort and the traveller behaviour, which would be evident if road user charges were instituted, is a very complicated matter. However, poorer motorists will tend to alter their travel behaviour to avoid road user charges more than wealthier motorists will. Moreover, motorists who place a lower value on their time may not be willing to pay.

As a consequence, road use charges have been attacked as a way of pricing the poor off the roads so that the rich can travel unimpeded, but similar arguments could be made about economic goods. A key consideration in the case of road use is the essential character of the service; if road use is held to be a basic right, then pricing it for reasons of efficiency may not be allowable.

2.2.3 Who will benefit from the road user charges?

Vickrey (1973) and Wohl (1970) indicated that the application of road user charges can produce a variety of impacts on road users. The chief benefit resulting from road user charges are travel time savings. These are enjoyed by motorists passing through part or all the Traffic Restraint Area (TRA), by their passengers, and by bus passengers in the area as well. The improvements in environmental

damages (impacts) are also enjoyed by everybody in the area particularly pedestrians and local residents.

However, the ratio of persons benefiting to those facing the charges falls sharply as earnings increase. This is because as earnings (i.e. levels of income) increase the willingness to pay also increases and the scheme will be less effective in achieving its main objective (i.e. reducing congestion), and consequently the percentage of beneficiaries from the scheme will fall.

2.3 The Costs of Roads

In considering the costs of roads, it is necessary to distinguish between the costs of providing roads and the costs arising out of the journeys. The costs of providing roads are the fixed costs that have to be incurred to make road journeys possible. These costs do not vary in the short run with the amount of road use. They are often described as capital costs, indirect costs, and by other names. Costs that arise out of journeys are variable— e.g. fuel, time, wear-and-tear of the road surface and so forth. They are often described as direct or running costs. The essential distinction is that the fixed costs of providing roads cannot be avoided by avoiding journeys. As once a road is built, no money can be retrieved by keeping it idle. On the other hand, the costs arising out of journeys can be avoided by avoiding the journeys. However, Roth (1966) argued that the difference between the unavoidable costs of providing roads and the avoidable costs that arise out of journeys is not one of principle but of degree. This is simply because even the costs of making the road can be avoided when the time comes to renew it. Thus, the costs arising out of journeys can be more accurately described as short-term, and the costs of providing roads as long-term. Accordingly, most road costs can be classified without difficulty into the short-term or the long-term category.

2.3.1 The costs of providing roads

The costs that have to be incurred in order to provide roads include: —

(a) Rent for use of land

The appropriate rent to pay for a piece of land used as a road is the rent it could earn in an alternative use, on the assumption that the road system continued in its existing function. It is sometimes suggested that it is not reasonable to charge rent for road space, because if it were not for the roads, the value of all the property in the neighbourhood would be insignificant. This argument is not convincing since the rental values in any district are not only due to road systems but also to the presence of shops, buildings and land uses of different kinds. Thus, this no reason for holding that some land use should not be charged. For example, the value of property in any district is most likely to be enhanced by the availability of electricity and water suppliers. But this does not mean that the electricity sub-station and water pumping station should be excused the payment of rent for the space used by their installations. Fitch (1963) has pointed out that:

“In ordinary accounting, land required for motor vehicle use which has already been paid for, including land used for streets, is treated as a free good. This is true also of park land and other land already owned by the governmental jurisdictions which is converted to motor vehicle use. However, for purposes of policy analysis, the imputed value of such land, that is the amount that it would command in alternative uses, should be considered in computing the costs of motor vehicle transportation.”

(b) Construction costs

These costs include the costs of labour and materials used in the road construction. Dawson (1961) examined a sample of 205 road improvement schemes and found that the costs of roadwork (which he defined as the total costs of road construction less the costs of land, bridges and subways, and ancillary works) vary less than the costs of land. Thus in 88% of all the schemes he examined, the costs of road works came less than £1.5 per square foot and in 71% to less than £1.0 per square foot. In six schemes road works cost more than £3.0 per square foot. Hence, the construction costs for roads per square foot could be more or less than the costs of land used for roads. This depends on the value of land in the area provided for roads.

(c) Interest on capital

The capital assets locked up in roads, represent resources that could have been used for other purposes. In order to prevent waste in the use of these resources, it is desirable that interest on capital investment should be charged to the users as a cost. For example, if an enterprise relies on private investment, it has no option but to pay interest charges. Also, if it raises the money out of its own resources it loses the interest that this money could earn in an alternative investment. However, it cannot raise money without paying for it.

(d) Maintenance and administration costs

In this category, it is difficult to separate the fixed costs of providing roads from the costs that arise out of journeys. For example, if the road surface is worn as a result of use, the resulting maintenance is clearly a cost arising out of road use. On the other hand, some maintenance is also necessary to prevent deterioration due to weather conditions. However, it could be assumed that only the costs of maintenance due to weather and vegetation come under the category of fixed costs of providing roads, while the costs of making good wear-and-tear due to vehicles, the costs of lighting, of traffic control, of police and of accidents come under the costs which arise out of journeys and can be avoided by avoiding journeys.

2.2.3 The costs arising out of journeys

There are four classes of costs that arise out of journeys. Each of these classes is briefly discussed below.

(a) Private costs

Private costs are the costs borne and paid directly by those who make journeys without giving rise to any special problems. These costs vary substantially with traffic conditions. This variation can be described on the basis of two relationships:

- (i) The effects of changes in traffic volumes on traffic speeds.
- (ii) The effects of changes in traffic speeds on the private costs of road use.

The relation between the volume of traffic and its speed under conditions of congestion has been investigated in a number of cities (see Roth, 1966). Generally speaking, as the traffic volume increases the average speed falls. The second relation can be described by the following formula:—

$$Cost = a + \frac{b}{Speed} \dots\dots\dots (2.1)$$

The constants 'a' and 'b' differ for different speed ranges and for different types of vehicles (see Roth, 1966).

(b) Road use costs

These costs include the costs of wear-and-tear of the road, lighting, traffic control, and all the costs of road system arising directly out of road use. These costs are usually obtained by dividing the total expenditure incurred over the whole country, or over broad categories of roads, by the estimated vehicle mileage on these roads. The main weakness in the estimates of road use costs lies in the absence of detailed statistics. In other words, there is a need for more information about the cost incurred on individual stretches of road and on the classes of vehicles that gives rise to those costs. This is particularly important in the case of certain classes of heavy lorries.

(c) Congestion costs

'Congestion' can be said to occur when road users impede the movements— and raise the costs— of one another. Thus, the costs imposed by road users upon one another under conditions of congestion are called congestion costs. It is convenient to distinguish between two classes of congestion costs:

- (i) The costs imposed on, and by, cyclists and pedestrians.
- (ii) The costs imposed by motor vehicles on one another.

In order to calculate the congestion costs resulting from a particular journey, it is necessary to know the effect of one additional unit of traffic (i.e. an additional vehicle) on other road users (see Chapter (3)). While no information is available on the first category of congestion costs, there is considerable information about the congestion costs imposed by vehicles on one another.

The effects of one vehicle in delaying the traffic can be calculated from the knowledge of the speed-flow relationship characteristics of the road network, i.e. the effect on the traffic speeds as a result of changes in traffic flows. The costs of congestion are the costs of delay. They also include higher labour costs, loss of peoples' time, higher fuel and running costs, and lower utilisation of vehicles and their loads.

In fact, congestion costs arise out of scarcity— the scarcity of road space. This scarcity can enable the owners of a congested road to levy a 'rent' from the users. That rent is equivalent to the rents chargeable by land owners, theatre and hotel operators and all those who own scarce resources and make them available to others. It is evident that the benefits obtainable from a congested road are largest when the rent required from each user just equals the congestion costs resulting from his presence. For if the rent demanded is put less than congestion costs imposed, some users would be attracted to the road even if the benefits to them fall short of the costs they inflict on others. Whereas if the rent demanded is in excess of the costs imposed on others, some potential users will be unnecessarily deterred from using the road (RPL, 1965).

(d) Community Costs (or Social Costs)

These costs describe the costs inflicted by road users not only on one another but on the community (society) at large. They include time delay to other classes of road users (bus passengers, pedestrian, and separate considerations of commercial vehicles, emergency service, etc.), accidents, visual intrusion, noise and atmospheric pollution. The existence of all these costs raises a large variety of problems. In the first place, there is the problem of evaluating the costs of each individual component of the community costs (i.e. time delay, noise, pollution, etc.). Secondly, the problem concerns the relationship between the amount of road use and the community costs resulting from it.

In conclusion, costs caused by vehicles vary greatly from one time and place to another; and to some extent these costs can be measured and charged for. Goodwin and Jones (1989) emphasised that the more classes of external costs that can be included, the greater the potential advantage of road user charges— though it becomes more difficult to have confidence in the precise estimates of optimum user charge. However, the next step is to see how these costs can be met; and whether there are any

charging methods that can be used to bring prices more into lines with costs. Therefore, the next section (i.e. section 2.4) considers briefly the different pricing policies (or strategies), then sections 2.5 and 2.6 review the theory of RP and the practical issues involved in its implementation respectively.

2.4 Pricing Policies (Pricing Strategies)

From an economic view point, there are alternative pricing policies that can be applied to a public service such as roads. These alternatives include: arbitrary pricing; average cost pricing; marginal cost pricing; and monopoly pricing. Each of these alternatives is discussed separately and very briefly.

2.4.1 Arbitrary Pricing

The service under this policy, is financed from proceeds of general taxation, which no attempt to relate prices to costs or the amounts spent on the service to the revenues collected from users. Therefore, it is likely that some users will pay more than the costs incurred on their behalf and some will pay less. It is not possible to say *a priori* whether under arbitrary pricing, road users in total will pay more or less than the cost to the community of providing roads.

2.4.2 Average Cost Pricing

In this pricing policy, the total costs of the service are divided among the users but without any attempt to charge individual users the costs incurred by the service on their benefit. This will result in users of non-congested roads being overcharged and in users of congested roads being undercharged. Accordingly, under a system of this kind, the demand for road space in congested areas appears virtually limitless, and the provision of adequate road capacity a physical and economic impossibility (Roth, 1966).

This point was put as follows by Deen (1963):

"When all users of both high cost and low cost facilities pay the same tax, the result is equivalent to the situation of an electric company which decides to bill customers not on the basis of individual consumption, but by measuring total power usage for the community and charging each consumer an equal part of the total bill. Not only is this inequitable; more importantly, it would eliminate the incentive for conserving electricity. Many new houses would, without any doubt, be heated with electricity, since an individual's cost would not be increased by a decision to install electric heat. Demand for power would soar, and new investment would be needed for new generating facilities. There would be no real basis for determining the proportion of total resources which should be devoted to power generation."

The analogy between electric power and road service raises the same difficulty when the costs of roads are averaged among users. Deen also indicated that the ideal transportation facilities would be designed so that users would be called upon to pay only the costs of the transportation services which he himself consumes. Thus, this concept implicitly considers transportation as a commodity or service not fundamentally different from electric power, water, or food, in that each consumer pays for these items in direct proportion to the quality and quantity of the service (or commodity) which he himself uses.

2.4.3 Marginal Cost Pricing

Under this policy, the user is charged only the extra costs resulting from a particular use. It is necessary, here, to distinguish between the case in which the user is charged only the extra costs borne by the supplier and that in which he is charged the costs imposed on the community as a whole (i.e. community or social costs). Since the former case arises only when goods are produced under

conditions of competition, it can be called 'competition pricing'; the latter can sometimes be termed 'marginal social pricing'. However, charging for all costs inflicted on the whole community raises a great many problems that cannot be dealt with easily. It is, therefore, proposed by Roth (1966) to consider only one type of marginal social cost pricing, this is the case in which users are charged the costs imposed on the supplier and on one another. This will be called 'user cost pricing'. Both competitive pricing and user cost pricing are discussed very briefly below.

(a) Competitive Pricing

In competitive pricing, because goods are produced under conditions of perfect competition, suppliers will maximise their profits if they increase their production to the point at which the cost of producing an additional unit just equals the additional revenue obtained by the sale of that unit. If production is restricted below this point, some units will be produced at a loss.

Under conditions of competition, road suppliers would attempt to maximise their profits, and road users to choose the cheapest route. If any road supplier were to make abnormally high profits, additional suppliers would be attracted to expand the road system in this area (as there is no restriction to the entry of new suppliers). Such expansion would continue until normal profits were earned (i.e. without attracting any new suppliers). In this way competition among profit-seeking suppliers would be likely to lead to the provision of a road network most suited to the needs of road users.

Although, it is possible to find competition in the provision of roads connecting points a long distance apart, the position is difficult with access roads. For example, in cities and in villages, roads are used not only for passage of vehicles but also to provide access to homes, shops and factories. In most cases access is provided by one road only, and the provision of further roads is impossible because of the technical layout of built-up areas. In these circumstances competition in any area is effectively impossible because any firm or individual owning an access road would be in a monopoly position. Thus, competitive pricing cannot be applied in built-up areas, the areas in which changes in road charging methods are most urgently needed. However, Roth (1966) argued that the user cost pricing, discussed below, is likely to produce the optimum use of existing roads, and the optimum investment in new ones.

(b) User cost pricing

Users, under this pricing policy, are charged the costs imposed on the supplier (such as wear-and-tear), and also the costs imposed on one another (i.e. congestion costs discussed in section 2.3.2). These congestion charges can be regarded as a rent paid to the supplier to ensure that the highest possible productivity is obtained from scarce road space. It is necessary to recognise that the total costs of providing the road system are not relevant to user cost pricing. Roth reported that the policy of any road supplier running his roads on the basis of user cost pricing would be guided by the following three rules: –

1. He should charge all road users amounts approximating to the main costs, including congestion costs, arising out of their use of roads, no more no less.
2. He should pay the costs incurred in providing his roads, based on the current value of the resources consumed.
3. He should expand those parts of the system on which revenues exceed outgoing. On the other hand, he should contract (or at least not expand) those parts on which outgoing exceed revenues, unless users are prepared to pay more than the costs arising out of road use or unless subsidies are provided by public authorities.

Thus, road user pricing has two main advantages over other methods of pricing. First, it is more likely to lead to the efficient utilisation of existing facilities, as road users would be encouraged by low charges to use non-congested roads and discouraged by high charges from using congested ones. Second, user cost pricing provides its own built-in criterion for investment– the criterion of

profitability. Charging systems that start by taking the existing costs of the road system as given, allocating the total costs among users, give no guidance as to which sections of the system should be expanded and which should be contracted. Under user cost pricing, this guidance is given by the existence of profits or losses in each particular section. Vickrey (1969) also supported this opinion and argued that appropriate patterns of congestion tolls are essential, not only for the efficient utilisation of existing facilities, but to the planning of future facilities. He added that in the absence of the information that would be provided by the charging of such appropriate tolls, planning of investment in expanded transportation facility is half blind, and resort is sometimes had to arbitrary rules of thumb, such as that of providing capacity adequate to handle the traffic during the thirtieth heaviest hour of traffic out of the year. The capriciousness of such a rule should be fairly obvious.

2.4.4 Monopoly Pricing

Under monopoly pricing, competition is restricted and this enables the suppliers to fix prices at levels which will maximise their profits. Therefore, this pricing policy is contrary to public interest. Monopoly pricing differs from user cost pricing in that it aims to maximise the profits of the suppliers, whereas under user cost pricing the object is to maximise the productivity of the facility being charged for. The practical difference lies in the extent of the restriction. For example, if user cost pricing were applied to a hotel, the owner would aim at restricting bookings to the extent that all rooms were occupied, and that there was no waiting list. Under monopoly pricing, the hotel owner aiming at maximising his profits might find it to his interest to charge a price so high that some of his rooms were hardly ever let. Similarly with roads, a monopolist might find it to his interest to restrict traffic unduly by charge more than user cost or else to charge less than user cost to profit from expensive congestion.

Having discussed the different pricing policies, it appears difficult to find how the optimal pricing could be decided in reality. Meyer and Straszheim (1971) discussed this issue in the following paragraph: –

"In general, the choice of a pricing strategy depends, at least to some extent, on subjective performances and objectives of public policy. This is true even within a relatively limited and static view of technologies and demand structures. Among the range of issues to be considered are development objectives, administrative questions, and welfare issues. In short, any viable generalisation about what constitutes an optimal pricing strategy is likely to be difficult, if not impossible, to obtain. Accordingly, an effort to improve the process of identifying goals and their relationship to alternative pricing strategies seems worthwhile."

Finally, it is worth mentioning that economists, for many years, have argued that governments should adopt marginal cost pricing for congested roads. To do so would increase economic welfare, measured in terms of consumers' surplus, and would encourage commuters to shift to public transport, thereby reducing the amount of money to be invested in road infrastructure (Borins, 1984 and 1988; Vickrey, 1963; Smeed, et al, 1964). OECD (1973) also indicated that the preferred philosophy of RP is to use marginal social cost charges so as to render possible the optimum utilisation of road and street networks. The next section, therefore, reviews the theory of road pricing and discusses its drawback.

2.5 Theory of Road Pricing and its Drawback

2.5.1 Theory of road pricing

The theory of RP goes back, at least, to the classic paper by Walters (1961). It is on the principle that prices of goods and services should, ideally, equal social marginal cost. Therefore, it is also called the "theory of marginal cost pricing".

In the absence of RP, when a decision is made concerning trip timing, route choice and whether to use the car at all, a driver using a congested road takes account only of his private costs (i.e. average costs)— what it costs him in fuel, time and possibly parking fee. He does not take into account the fact that his vehicle also slows down other vehicles. Each vehicle is impeded perhaps to a very small degree, but summed over all vehicles in the traffic flow, the impedance imposed is much larger and will be proportionally greater the higher the density of traffic (Starki, 1986). Marginal social cost (as discussed in the previous section) includes both the average costs and other additional costs (i.e. delay, running costs and environmental costs) that are inflicted on other users and non-users of the road network when an additional trip is undertaken. Larsen and Rumjerdi (1990) argued that the marginal social costs under congested driving conditions, could be two to five times the average costs. Of course, if the driver took into account the costs the rest of the community had too, then it might be that he would decide that the journey was not worth making— at least at that time or by that route.

Therefore, RP is theoretically designed to overcome this underestimation of costs by levying a charge on road users which, as nearly as possible, confronts them with the full costs of congestion that their use of the road causes. This charge, ideally, equals the difference between the marginal social cost and the driver's average cost. Thus, drivers will use the roads only if the value to them of doing so exceeds the amount the society must pay. Accordingly, low-priority users will be priced off the road; they will change mode, time, go to another destination or not make the journey, while high-priority users are accommodated with fewer deterioration effects from crowding, lower travel times and higher speeds. As a result, the charge acts to bring about a more efficient allocation of a scarce resources— something which is of benefit to road users themselves as well as to the economy as a whole.

The theory of RP is summarised in Figure (2.1), which shows the relationship between flow (q), average cost and marginal social cost. The demand (Q) may be specified by equation (2.2), while the average cost (C_a) is given by equation (2.3) below.

$$Q = g(q) \dots\dots\dots (2.2)$$

$$C_a = f(q) \dots\dots\dots (2.3)$$

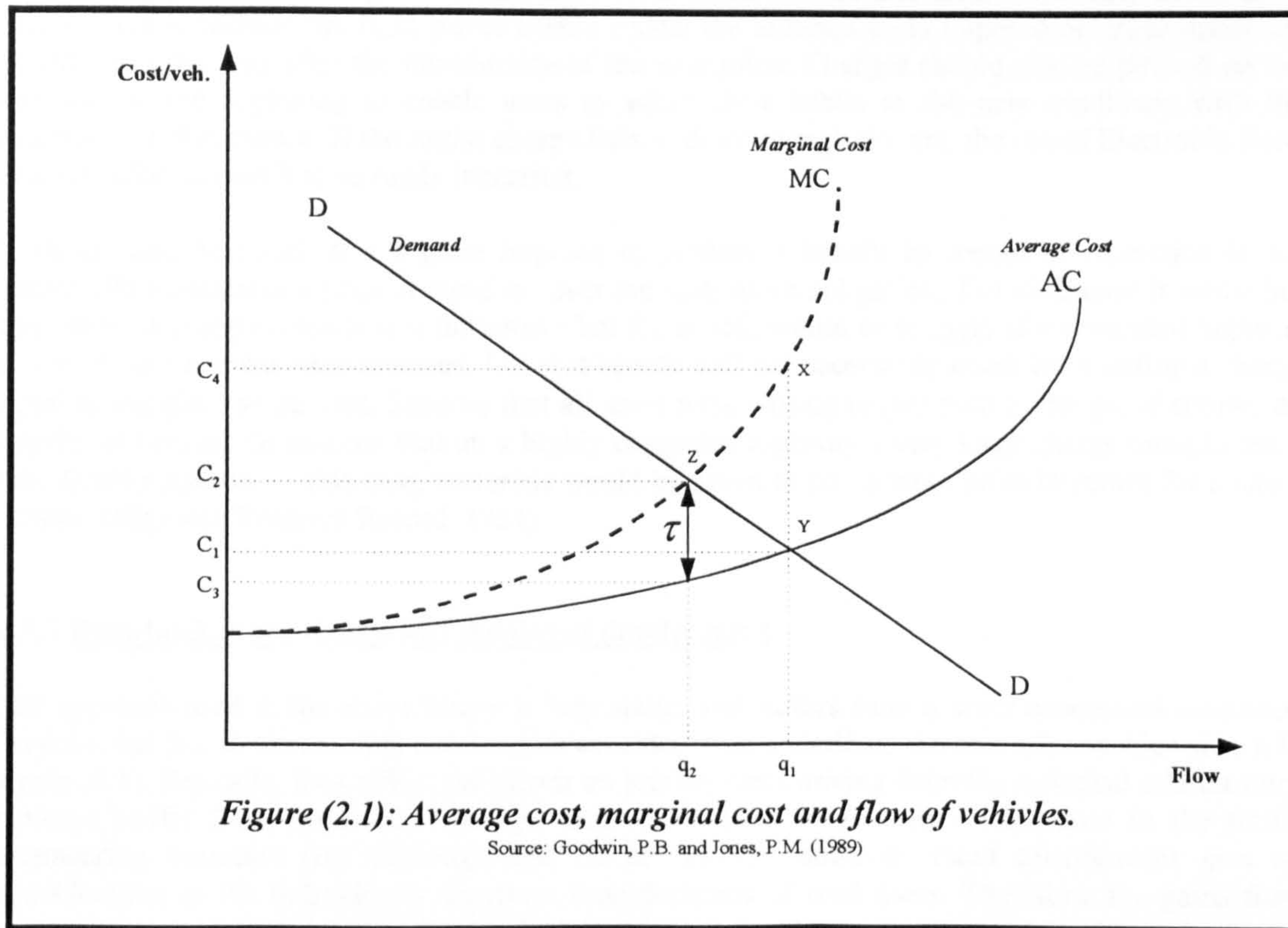
The marginal social cost (C_m), which shows the additional costs of additional vehicle is given by differentiating total cost (T_c) with respect to flow (q) as in equations (2.4) and (2.5) below.

$$T_c = q \cdot f(q) \dots\dots\dots (2.4)$$

$$C_m = q \cdot f'(q) + f(q) \dots\dots\dots (2.5)$$

In the absence of RP, as motorists will not be aware of the additional costs they impose on others (i.e. $q \cdot f'(q)$), then equilibrium occurs at q_1 where the demand curve (DD) intersects the average cost curve (AC) at Y. Thus any user who values the trip at more than the cost he bears, travels. However, at this point the extra cost to society, i.e. including other users, exceeds the benefit derived by the last traveller. The same is true for all trips beyond q_2 . The amount by which the additional cost of these (q_1 - q_2) trips exceeds the additional benefits is equal to the area XYZ in the diagram. This area represents the welfare loss from non-optimal pricing. The total cost incurred by all vehicles in this case is $q_1 C_1$ and the consumer surplus is

$$\int_0^{q_1} g(q) \cdot dq - q_1 C_1$$



To ensure that the road is optimally used, each user must be charged a toll τ equal to the difference between the marginal and average costs at the optimal flow (q_2), i.e. ($\tau = C_2 - C_3$) as shown in Figure (2.1). The marginal cost curve (MC) and the demand curve (DD) now intersect at Z; flow will be reduced to q_2 , and cost for each driver increased to C_2 . Total cost incurred by all vehicles is now $q_2 C_2$, this value may be more or less than it was previously (i.e. $q_1 C_1$), depending on the shape of the curves. The new apparent consumer surplus is

$$\int_0^{q_2} g(q) \cdot dq - q_2 C_2$$

This value must be less than the old consumer surplus since q_2 is less than q_1 .

2.5.2 Determining the size of charge in practice

On the one hand, the size of the charge, from the above theory, depends primarily on a knowledge of the shape of the demand curve. In practice, this is very difficult to establish and perhaps impossible to establish because of the continual shift of the demand curve through time as a result of changing public taste. On the other hand, although it is possible to make reasonably accurate estimates of some costs, for instance of congestion, it is much more difficult to calculate a 'price' for other external costs including the damage to the environment. Therefore, Hewitt (1989) argued that it may not be necessary to charge the full amount that would be justified if all the social costs of using the roads were reflected in the charges in order to persuade them to drive less. Hence, precise calculation of the external costs involved is not, however, necessary. Instead, the price can be fixed at whatever level is needed to persuade enough car-users to switch to other forms of transport, or to change the time or route of their journeys. The question of which level of road charges is 'enough' depends on judgements about the overall objectives of the scheme (i.e. efficiency and accessibility of using road network, the desired levels of reduction in congestion and vehicles emissions, as well as the amount of revenue to be raised).

However, in practice, these prices would have to be fixed by trial and error, especially in the early stages. This is because the right prices should reflect the external costs imposed on other under the conditions prevailing after the introduction of the new price. Charges should also be pitched on the low side at the beginning to enable users to adjust their habits to the new conditions with the minimum of disturbance. If the initial charge fails to deter enough drivers, the use of Electronic Road Pricing (ERP) allows it to be easily increased.

It should also be noted, that a price imposed to produce a benefit by means of restriction is not necessarily identical to a price charged to cover the costs as an obligation. The difference is subtle but important. A marginal social cost indicates what the benefit would be to users of a congested highway if one of their number were removed. But that benefit will not necessarily result from setting a charge equal to marginal social cost. Suppose that all users were willing to pay such a charge, of course, no benefit will result. Or suppose that on a highly congested highway a very large charge brought but a very small response. In this case, motorists would be asked to pay a large price in return for a small benefit. (Highway Research Record, 1964).

2.5.3 Drawback of the theory and the line of development

The approach used in the above theory is very static, and suffers from a severe conceptual weakness. To point out this weakness, it is necessary to consider what underlines the cost relationships shown in figure (2.1). Basically, they reflect the effects on journey costs arising from the technical relationships between traffic flow, speed and density, which have been extensively documented in the traffic engineering literature (see Gerlough and Huber (1975)). However, these relationships give no consideration to the individual's departure time decisions of road users. Therefore, the gains from such charges is due to changing the level of demand but not the time pattern of the demand over the peak period. The recognition of this weakness goes back to Vickrey (1963), who derived the departure rate along a single route subject to queuing congestion as the outcome of individual cost minimisation. In addition to characterising the no-toll equilibrium, Vickrey (1969) determined the social optimum and solved for the toll which decentralises it. He also provided an illuminating discussion of the benefits of capacity expansion on a single route and for routes in parallel, with and without the toll.

Vickrey's model was independently formulated by Hendrickson and Kocur (1981) and Fargier (1983). Fargier considered both the morning and afternoon rush hours (characterised by a desired departure time). Subsequent work has extended these papers. For example, Smith (1983) proved the existence of equilibrium under general assumptions on the function relating trip costs to travel time and schedule delay, Daganzo (1985) provided a proof of a uniqueness of equilibrium, and Hendrickson and Kocur (1981) provided an empirical application.

The work in this study follows up this line of development by examining the user equilibrium for a single bottleneck as well as a typical traffic network under a cost function of three components. These three components are travel time and delay; schedule delay (i.e. the difference between the actual arrival time to destination and the preferred one) and; a time varying toll (i.e. *TDRP*). This will be discussed in more detail in Chapter 4.

2.6 Practical issues involved in the implementation of RP

At the same time as the improvements in theory were appearing to make road pricing more complex and demanding, there was also a gathering of experience of the practical issues that have to be considered in the implementation of a practical system. These issues are discussed below.

2.6.1 Methods of Charging

Roth and Thomson (1963) suggested that the ideal method of charging would be flexible and selective. A flexible system of charging would try to take account of the wide variation of traffic in

congested areas by discouraging traffic from using congested roads and encouraging it to use non-congested ones. Under selective charging systems, the vehicle deterred from using the congested streets should be, by and large, those whose need to use them is least. Roth and Thomson also added that an efficient charging system gives the road user the chance to assess the importance of each journey and decide for himself whether to pay the price charged or not.

There are also some other conditions which must be satisfied if a charging system is to provide a practical means of collecting revenue. It must be cheap to operate, easy to understand, and yet difficult to evade by fraud or error. It should allow for the possibility of payment in advance in a small amounts and at fairly frequent intervals; any devices used should be highly reliable; it should be fair in its incidence and it should not raise awkward enforcement problems. Furthermore, and perhaps most important from one view point, Roth (1966) emphasised that the charging method should indicate the strength of demand for road space in different places and at different times of day, and it should enable the payments made over alternative routes to be known in some detail.

In practice, the implementation of pricing policies (strategies) depends on the charging methods available. It is useful to distinguish between two alternatives; direct and indirect methods of charging for roads. In short, direct methods involve charging for the use of the road as such, while indirect methods involve charging for something (such as fuel) used in associated with roads. Different methods of direct and indirect road charging are illustrated in Figure (2.2).

The UK Ministry of Transport study into road pricing (HMSO, 1964) examined two basic approaches to vehicle direct-charging:—

(i) Off-vehicles recording systems, whereby the charges payable by vehicles are recorded elsewhere than on the vehicles. These systems can be compared with those used for charging for telephone calls.

(ii) Vehicles metering systems, whereby the charges are registered on the vehicles themselves. These systems are analogous to taxi-meter methods of charging. They may be driver-operated or automatic.

A further subdivision is between “continuous” charging and “point” charging. Continuous charging involves setting up pricing points on the borders of congested areas, and charging the vehicles according to the time spent (or distance travelled) in the zone, as deduced from recordings at the points of entry and exit. Points charging involves setting up pricing points throughout the congested areas, vehicles being debited with the appropriate charge when passing any pricing point.

Given the technology of the early sixties, the study team came out strongly in favour of a driver-operated meter as being cheaper, more flexible, more reliable and better able to reflect the philosophy behind road pricing than an off-vehicle system. Twenty years later the balance had swung in the other direction in the Hong Kong Electronic Road Pricing Demonstration, which used a passive, off-vehicle system, on the grounds that it was cheaper and less prone to tempering.

Goodwin and Jones (1989) indicated that although considered technically superior by the Hong Kong consultants, the off-vehicle system has a number of consequences for the way in which road pricing can be implemented:—

- All vehicles have to be fitted with an Electronic Number Plate “ENP”, including those which rarely use an ERP-controlled area of the country, thereby adding to the scale of the set-up problem, and to the difficulty of containing the system.
- The use of loops in the road as the means of detection makes it more difficult to introduce extensive changes to the system and favours some type of point rather than continuous pricing system.

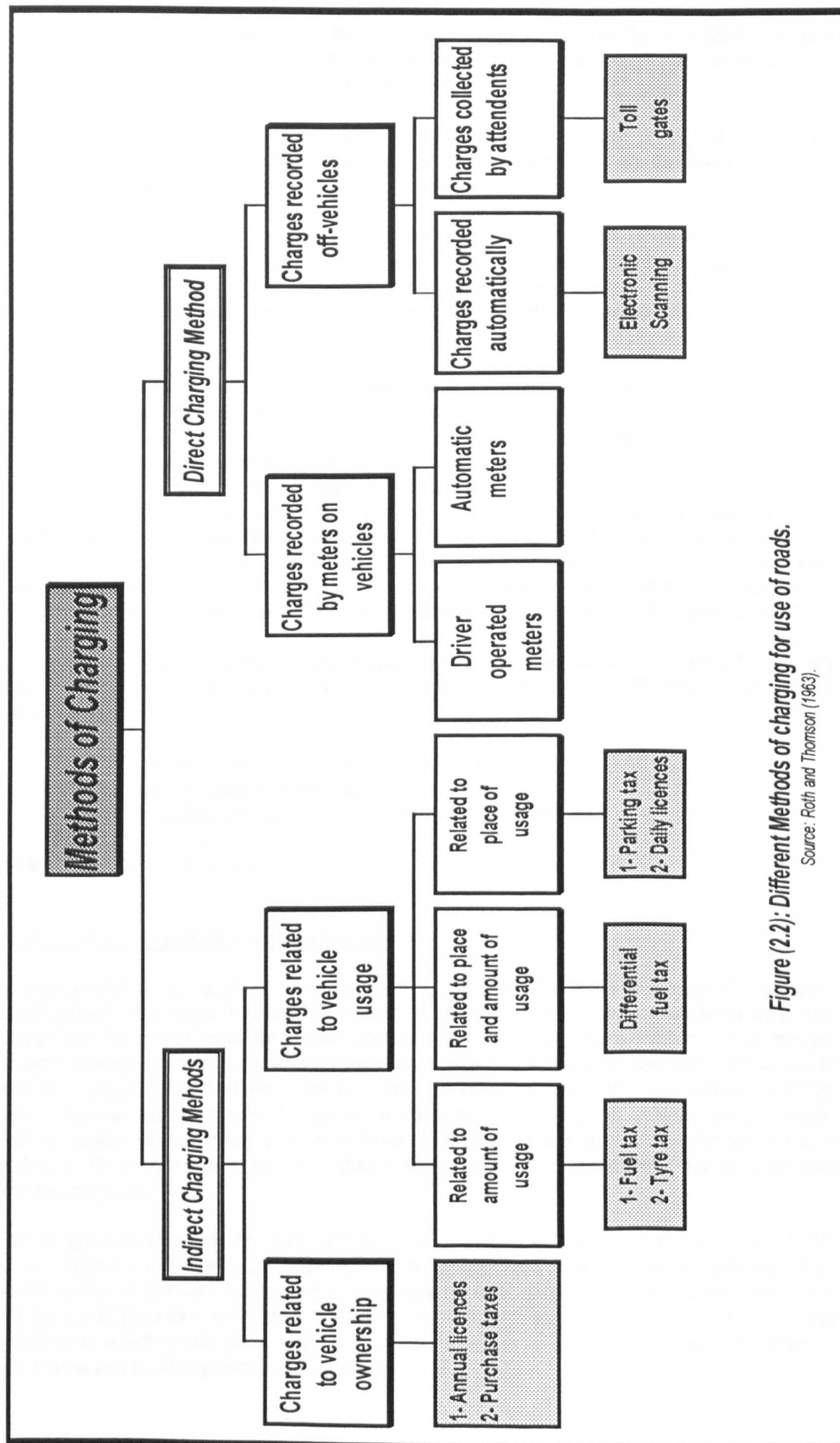


Figure (2.2): Different Methods of charging for use of roads.

Source: Roth and Thomson (1963).

- There is little information, at the time a car journey is being made, as to the ERP cost of the journey and because charges may be collected some weeks after the journey is made the link between journey and cost may be weakened.
- Because of the lack of immediate association between travel and price, and the need for taxis to travel in the charged areas without passengers, there are problems of deciding how to pass on the congestion charge to taxi passengers.

A third charging option— and the one which has been most widely used on tolled roads and in area licensing schemes in Singapore, Bergen, etc.— is a manual system, where charges are either collected on site or permits are purchased in advance and displayed on the vehicles. The former has serious implications for congestion and the latter severely restricts the spatial and temporal complexity of the road pricing scheme.

A more advanced method is to use a “*smart card*” which will contain a number of units, say, 50 and would be purchased for 10 pounds if the price per unit is 20 pence. For safety and convenience reasons however, the system will go into negative mode when a card is empty until the next time the engine is switched off. It will then be totally immobile until a new card is inserted which will instantly deduct the negative units. A visual display unit could give the number of units left on the card. Electronic Toll Collection using ‘smart card’ as a payment and data recording tool is considered the most desirable method of collecting fares in many countries e.g., the United Kingdom, Germany, Sweden, Netherlands, Singapore and France. The result of expertise and experience in this field concluded that smart cards create a secure, efficient and reliable revenue collection mechanism for electronic toll collection, parking and other transportation applications (see Farid Amara, 1994).

Accepting the technical feasibility of road pricing— which was established in the Hong Kong study beyond reasonable doubt (for that particular type of system)— there are three further implementation issues to be decided:—

- Which vehicles to be exempt from the road use charge..
- How to vary charges by time of day.
- How to distribute charges across the city (areas, direction, etc.).

These are considered each in turn.

2.6.2 Exemptions for certain vehicles or persons

From a theoretical viewpoint, there is no reason why all vehicles should not be charged the economic or social cost of the journeys they make in congested conditions, but in practical applications to date the main costs have been borne by private motorists, with buses and goods vehicles being exempt. Costs per bus passenger would probably be quite low at peak times, but there has been a reluctance to impose the charge because of the wish to encourage travel by this more space-efficient mode; similarly, there are concerns about increasing manufacturing costs if goods vehicles were charged— though the impact of such charges on some firms might well be less than the inefficiencies which arise through the present restrictions on loading and unloading in congested areas, and so might lead to net cost reductions.

The main problem is caused by taxis which, both in Singapore and Hong Kong, were initially excluded from the scheme, but experience has shown that they need to be included because of the high cross-elasticities of demand between cars and taxis in such densely populated areas; given the substantial use of taxis by non-car owners, however, it is probably expedient to charge taxis at a lower rate than cars, which would encourage some transfer from car to taxi, but this would be offset by transfers from taxi to other public transport modes.

Pressure for exemptions also comes in another form, from particular groups of people who feel they are unfairly disadvantaged by road pricing or have a special right to road space in the affected area. Examples include local residents— who are exempt from access restrictions in many schemes such as the no-car area in Florence— and disabled drivers. Local business people, too, may argue that they have as much right to special treatment as the local delivery driver.

2.6.3 Temporal Pattern of Charges

Traffic conditions change continually over the road network and vary from day to day. It would be possible electronically to measure variations in traffic flow minute by minute and vary charges accordingly. Goodwin and Jones (1989) argued that this would not be practical or desirable from a behavioural viewpoint: motorists would be accelerating past charging points as prices were about to change, or do U-turns, and would have no clear idea of the cost of travel when the travel decision is taken (aside from minor variations of route). The constraint is thus one of comprehensibility, stability and simplicity: patterns of charges, once established, should remain in force for months at a time and the range of charges should be simple enough for motorists to understand and remember.

Balancing these considerations in Hong Kong led to a six-period charging system, based around morning and evening peak charges, an inter-peak charge, an off-peak (zero) charge and shoulder charges to stagger the rise from the off-peak to peak charges. There was also provision for special charges, perhaps levied during road works or on days when heavy traffic loads are anticipated.

However, the system of RP proposed in this study (i.e. *TDRP*) is considered to be a great challenge for the temporal pattern of charges as the charge is expected to vary by time of day over every individual traffic junction throughout the road network. The model for estimating the value of this time-varying charge is developed in Chapter (3) and the stability of the user equilibrium under this system of charging is examined in Chapter (4) for a single bottleneck and in Chapter (7) for a typical road network.

2.6.4 Spatial charging structure

Goodwin and Jones (1989) identified five different charging structures, each of them is considered briefly below.

(a) Link charges. Motorists would be charged for using certain links of the road network, perhaps where capacity is especially restricted, or where there is a concern to limit traffic for environmental reasons.

(b) Junction charges. In this case, charges would be levied according to junction basis. Thus, electronic loops would be laid at major junctions to control traffic levels over a certain area of the network. This is probably more efficient than (a) especially if some turning movements could be exempted.

(c) Cordon charges. Motorists would be charged for entering or leaving a designated area. This is similar in concept to Singapore Area Licensing Scheme (see section 2.7). In this case, charges could be varied by time of day and direction of travel.

(d) Boundary charges. Boundary charges are similar to cordon charges, but charging lines extend right across travel corridors, so that there is little scope for switching route to avoid the charge— unlike (a) & (b) where, to varying degrees, route switching is an option.

(e) Zonal charges. Conceptually this is quite simple, but is difficult to achieve in practice. A combination of a cordon/boundary and a junction or link configuration would be used to charge motorists for journeys made within and between zones— rather like public transport fare structure now in London.

2.6.5 Other aspects

This section has discussed just a few of the practical considerations involved in implementing a road pricing scheme. There are also some other issues which are not mentioned in the above discussion for example, the sophisticated accounting systems or the design of the electronic toll sites. These issues are out of the scope of this study.

Having discussed the most important issues involved in the implementation of a road pricing system, the following two sections (sections 2.7 and 2.8) are giving genuine examples of implementing road pricing in Singapore, Hong Kong, Norway and UK.

2.7 Singapore, Hong Kong and Norway Experiences with RP

The usual objectives for RP are a combination of the following— to increase efficiency, to improve the environment and to extend access (i.e. raise money for transport improvements). Singapore, Hong Kong and Norway are the principle exhibits in the real world for selective road pricing that is relevant to all these objectives. This section throws some light on these three schemes.

In Singapore

During the late 1960s and early 1970s, the government carried out two major transport studies, both of which concluded that limitations on the ownership and use of private motor vehicles were necessary. The government's goal was to reduce peak hour traffic by 25 to 30 percent. Several alternative measures, including increased parking charges in the central business district and higher taxes on the import, purchase, and registration of automobiles, have been considered and adopted, but they were considered insufficient by themselves, due to their lack of temporal or geographical specificity. Accordingly, during 1975 the Area Licensing Scheme (ALS) was implemented. This scheme, the first and currently the only one of its type in the world, is part of a package of transport measures designed to reduce the growth of car ownership and use, particularly during the peak periods, and to improve public transport and the environment.

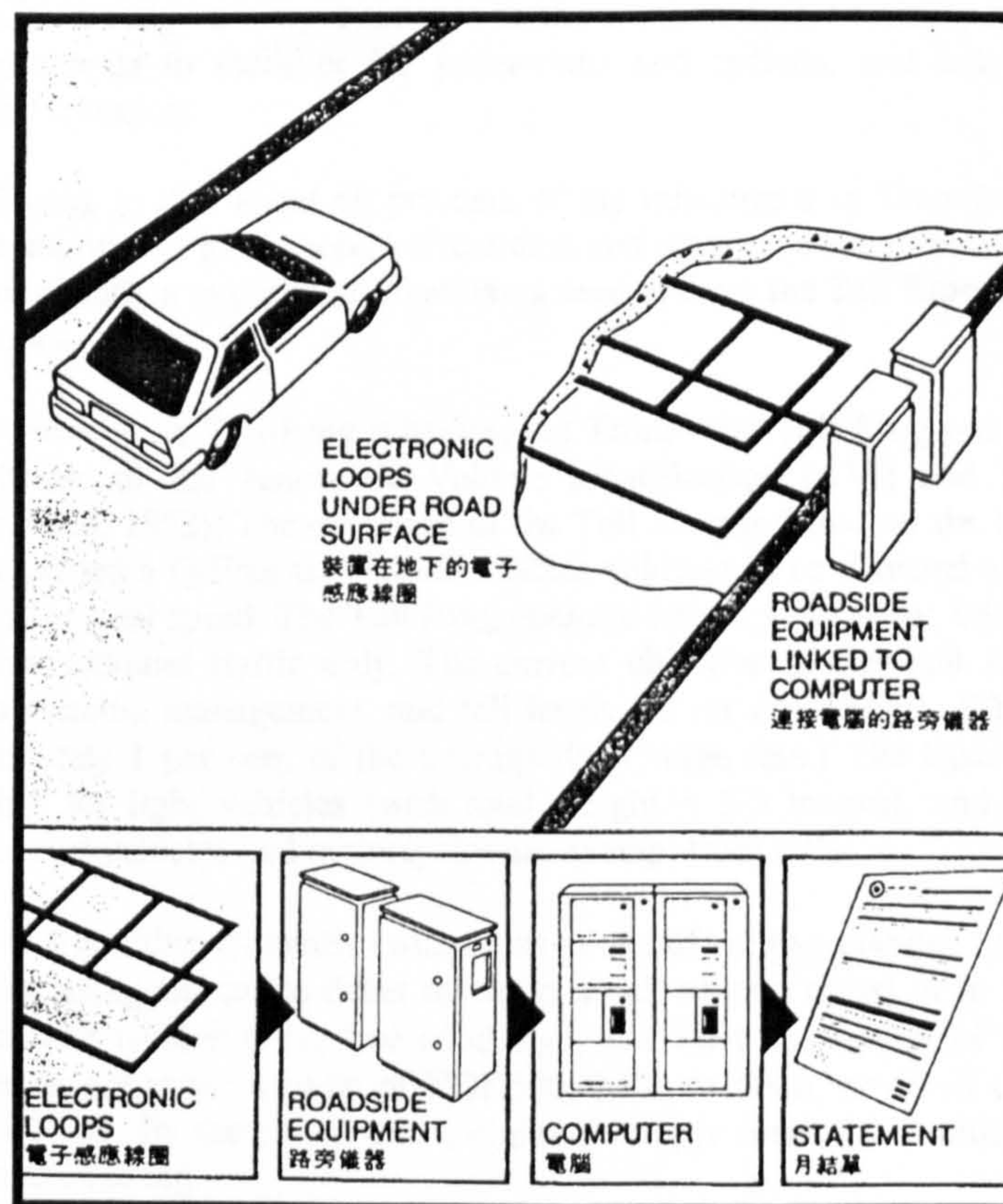
A restricted zone of about 2.4 square miles was established in the central business district. During restricted hours (initially 7:30 a.m. – 9:30 a.m., subsequently extended to 10:15 a.m.), entry into the zone requires purchase and display of a special licence. Monthly licences are sold at the Registry of Motor Vehicles and at some post offices. Daily licences are sold at post offices and at road-side sales booths. Company cars pay as twice the licence for private cars; taxis pay only forty percent; buses, goods vehicles, motorcycles, and car pools with at least four persons are all exempt.

The Area Licensing Scheme 'ALS' has significantly reduced congestion and helped to control the growth in vehicle use in Central Singapore. The reduction in traffic was caused by a substantial shift from auto to bus and car pool and settled down to about 44%, with an increase of 22% in traffic speeds. Also administration and enforcement have not proven to be problems.

Holland and Watson (1976) believe that it should be possible to design and implement a workable scheme for a wide variety of cities. However, they do cite particular factors that made implementation of the ALS easier in Singapore than might be expected elsewhere. First, the relative isolation of the region from outside traffic makes administration and enforcement easier. Secondly, because cars were used by only a relatively small fraction of downtown commuters prior to the ALS, there were fewer potential opponents, as well as a greater relative capacity of public transport to accommodate those who switched from auto to transit. Political problems common to such schemes elsewhere are much less important in Singapore because Singaporeans tend to believe that the government acts in the general social interest and are thus more accepting of rules and their costs. Finally, planning and implementation are much easier in a city-state with only one level of government.

In Hong Kong

In the early 1980s, the most ambitious road pricing scheme was the Electronic Road Pricing (ERP) system planned for Hong Kong. During 1982, although there were less than five vehicles for every one hundred persons in Hong Kong, there were 300 vehicles for every kilometre of the road network. As was true in Singapore, general increases in the cost of automobile ownership and operation were not considered sufficient in themselves to tackle the congestion problem, due to their lack of selectivity. ERP, on the other hand, is the most selective of all auto restraint schemes. The technology of the experimental Electronic Road Pricing Scheme in Hong Kong is described in Catling and Harbord (1985), and illustrated in Figure (2.3). The plan for Hong Kong was to fit every vehicle with an Electronic Number Plate (ENP). The ENP is a sealed unit, about the size of a video cassette tape, fitted underneath the vehicle. The unit is "interrogated" by transmitters buried beneath the road surface at each of the 300 planned electronic toll points. The ENP then transmits its unique number to a computerised receiver that stores the information. The vehicle's account is charged for a predictable number of units, depending on the location and time of day, with a bill sent to the vehicle's owner at the end of the month. Vehicles failing to respond with a valid or operative ENP are photographed for subsequent assessments of penalties. The technology for a full ERP system was fully demonstrated by the pilot scheme began in April 1985. Results were extremely encouraging, and exhaustive tests of the system during the last few months of the project confirmed that the Hong Kong system is accurate, reliable and robust enough to be extended to a full system. The transport studies (by the Government) had been successfully completed and indicated that ERP is fair and efficient restraint policy option open to the Hong Kong Government for dealing with the continued problems of traffic congestion.



Source: Catling and Harbord (1985).

Figure (2.3): Technical operation of the Electronic Road Pricing Scheme Developed for Hong Kong.

Also Dawson and Catling (1986) concluded that the pilot-stage project has been successful, and that ERP has been demonstrated as technically viable, administratively feasible and of significant potential benefit to deal with Hong Kong's heavily congested urban road network.

The Hong Kong project has brought together a number of significant recent technological advances and has combined these with established theory to demonstrate the practicality of road pricing as a very important method of dealing with traffic problems. However, the politics of any traffic restraint scheme are significant, whether for reasons of environmental management or control of congestion, and ERP generated public debate in Hong Kong throughout the pilot stage project. Because of the success of the 1982 car ownership restraint measures in keeping congestion at a manageable level, in many quarters, the need for ERP in Hong Kong was not felt immediately. The scheme has therefore been shelved while traffic conditions are monitored.

In Norway

For a number of years, Norway has financed new road infrastructure by charging tolls. Initially this was done on bridges and tunnels outside built-up areas, but in recent years Toll Rings have been introduced around major urban centres including Bergen, Oslo and, most recently, Trondheim. The original legislation under which these tolling schemes were introduced restricted the use of money collected through tolls to the provision of new road infrastructure. Now, however, it is possible to use part of this money for public transport infrastructure and urban environmental improvements.

The Trondheim Toll Ring started operating on 14 October 1991. In accordance with national policy, the main objective for the Toll Ring is to raise revenue to be used to finance the 'Trondheim package'. This package includes a range of improvements to the local transport system, including new road construction, improvements to facilities for pedestrians and cyclists, and enhanced priority and segregation for public transport.

The Toll Ring is aligned so that about 60 per cent of the inhabitants in Trondheim live outside its boundary, while the majority of jobs, shops, recreational and other public services lie inside, at or near the city centre. Consequently a majority of inhabitants need to cross the Toll Ring at some point in the course of their activities.

In contrast to other Norwegian Toll-Ring schemes, the Trondheim Toll Ring was designed from the outset to make full use of the Automatic Vehicle Identification (AVI) and automatic debiting technology (see Waersted, 1992). The operation of the Toll Ring is based on the use of a passive in-vehicle transponder (called a Q-Free tag) which enables vehicles to be detected and charged as they cross the Toll Ring at normal speed. The Toll Ring operates Monday to Friday, 06:00h to 17:00h, and collects revenue from inbound traffic only. The current objective of the Toll Ring is to generate revenue rather than demand management, and toll levels are set accordingly. (The toll per crossing amounts to approximately 1 per cent of the average daily wage rate.) The basic toll level is set at NOK10* per crossing for light vehicles (with total weight < 3.5 tonnes), and NOK20 for heavy vehicles. Public transport vehicles and motorcycles are exempt from tolls.

Drivers can choose to pay either manually (with coins at one of the two manned toll stations, or at an unmanned stations by using automatic ticket machine which accepts coins) or to subscribe to the Q-Free system. Subscribers receive a tag free of charge, and have the option of using it in several payment arrangements: prepayment units of NOK500, 2500 or 5000, or direct debit on a monthly basis from a bank account. By the end of 1994, eighty-three per cent of the vehicles paying to cross the Toll Ring used a Q-Free tag.

* NOK11 $\cong$ £ 1.00.

Before and after travel surveys were carried by SINTEF (at the Norwegian Institute of Technology) in collaboration with the Transport Studies Unit, University of Oxford. Based on the results of these surveys and an automatic counting, Meland (1994) has concluded the following effects on:-

- **Total travel by car**, there is no evidence that the Toll Ring reduced the overall car use during the whole week, neither in the number of trips nor in trip duration.
- **Temporal distribution**, the toll Ring has shifted the timing of car driver trips, away from the toll charging periods, and over to evenings and weekends when tolls are not charged (the car trips showed a decrease in trip frequency during the tolled periods and an increase during the toll-free periods).
- **Spatial distribution**, during the tolled periods there was a decrease in the number of car trips crossing the cordon in each direction, while there was an increase in the number of trips between places outside the Toll Ring. After the toll period there was a decrease in the number of trips between places outside the Ring and an increase in trips crossing the Ring. This indicates two different types of changes in travel behaviour: changes in choice of destination- for shopping trips, for example- during periods with tolls, and shifts in time, where trips that do not cross the Toll Ring are more frequently made during the tolled periods, and trips that cross the Toll Ring are postponed till after the end of the tolled period. There was a general decrease in trips between places inside the Toll Ring in both the tolled and the untolled period.
- **Modal split**, in Trondheim as a whole there was a decrease in number of trips for all modes, except public transport, which had an 8 per cent increase. Car passenger and bicycle trips were reduced by 14 to 15 per cent.

2.8 Recent UK Studies Concerning RP

(RP systems proposed for Cambridge)

Since Smeed produced his report (Smeed, 1964), there have been many studies of different aspects of road pricing. More recent studies include Goodwin (1989), who considered problems associated with the division of the income derived from the prices paid (i.e. 'rule of three'), and Evans (1992), who emphasised motivational problems. Most recently, Milne (1993), May et al (1994), Smith et al (1994a) and Milne et al (1994) have given results obtained by using the network models SATURN and CONTRAM to evaluate four possible road pricing, or road charging, systems on a detailed model of the Cambridge road network. These four road charging systems are : -

- Toll Cordons;
- Distance-Based Charging;
- Delay-Based Charging; and
- Time-Based Charging.

This section briefly describes these four charging systems as three of them are used later in this study for the comparative analysis with *TDRP* (see Chapter 8).

• Toll Cordons

Toll cordons represent the most conventional form of road user charging. A fixed charge is levied at specific points in order to permit travel within a specific area, defined by a single or series of boundaries. Cordon charges are normally one-way. The charge incurred for a trip is directly related to the number of boundary crossings and charges may varied by time of day, direction and vehicle type/user class.

In implementing cordon tolls for Cambridge, a fixed penalty toll is added to the generalised cost of traversing each link in the CONTRAM assignment process when either a cordon or a screen line is crossed.

- **Distance-Based Charging**

Here each vehicle is charged according to the distance travelled, by that vehicle, in the charge area. This encourages drivers to make greater use of regions that have no charge, and it will also have the effect of reducing congestion within the charge area. On the other hand, it also encourages drivers to use shorter distance routes within the charge area, and this will have the effect of increasing congestion within this area.

In implementing distance-based charging in Cambridge, a penalty in pence per kilometre is added to the generalised cost of travelling a link in the charge area, in the CONTRAM assignment process.

- **Delay-Based Charging**

Here each vehicle is charged according to a long run estimate of that component of travel time spent in delay (mostly in traffic queue) in the charge area, thereby encouraging the vehicle to avoid congestion points and so to choose routes that reduce queuing delays. This charge would be expected to vary at only a slow rate both within each day and from day to day. In this way the charge is predictable and so may be expected to influence drivers' choices in a beneficial way.

Perhaps this is the closest to the congestion metering system explored by Cambridge, where a charge is only levied when some fixed delay threshold is exceeded (the most commonly suggested threshold is three minutes to travel half a kilometre). However, long-run delay-based charging is easier to model and properly influence the decisions of drivers as the prices are fairly stable and predictable.

In implementing delay-based charging in Cambridge, a penalty in pence per hour of delay, is added to the generalised cost of traversing each link in Cambridge, in the CONTRAM assignment process.

- **Time-Based Charging**

Here each vehicle is charged in proportion to the time spent travelling in the charging area. Delay and free-flow travel time are charged equally. Drivers are expected to choose routes which reduce their own journey time in the charge area. This may perhaps be easier to implement than congestion pricing; but it has the drawback that non-congested and congested travel are equally priced and so departs very significantly from the delay approximation to marginal cost pricing described in Smith and Ghali (1992).

In implementing time-based charging in Cambridge, a penalty in pence per hour of travel time is added to the generalised cost of traversing a link in Cambridge in the CONTRAM assignment process.

For more details about road user charging systems in Cambridge see Smith et al (1994b) and (1994c).

2.9 Summary

This chapter has reviewed the historical background of road pricing and answering the questions of why paying for roads, who will pay the road user charges and who will benefit from road user charges. The different costs of roads have been discussed and the different pricing policies (strategies) for allocating these costs on road users are considered. The theory of road pricing as well as its drawback are discussed. Then, the Chapter addresses the line of development in the theory and shows how this study follows up this line of development.

As the costs caused by vehicles vary greatly by time of day and place and to some extent these costs can be measured, the implementation issues of a road pricing system that could bring prices more into lines with these costs are presented in section (2.6).

The experiences of Singapore, Hong Kong and Norway in implementing road pricing schemes are outlined. The most recent UK studies concerning road pricing are highlighted with a brief description for the different RP systems proposed for Cambridge.

Finally, the conclusions that emerge from this Chapter include the following:—

- From a theoretical viewpoint, every individual road user should be charged the economic cost (or social cost) of the journey he/she makes in congested conditions to achieve the efficient utilisation of the existing facilities. Thus, the RP scheme that could fulfil this theoretical viewpoint is worth exploring.
- Road pricing is better able to restrain traffic levels in congested areas than parking controls or car ownership restraint.
- Road pricing enables higher car ownership levels to be absorbed than would otherwise be possible in a city.
- Road pricing provides an indication of the economic demand for road space and a source of revenue for future road building or improvements to other parts of the transport system, or environmental improvements.
- Road pricing schemes will be most effective when introduced in conjunction with other forms of restraint, where a good public transport system already exists or improvements are credibly planned as part of the road pricing package, and there is some scope for activity re-scheduling.
- The practical experiences of RP in different countries have demonstrated the technical feasibility of Electronic Road Pricing (ERP). Electronic toll collection, using smart-card technology is emerging to be the preferred method of charging as it creates a secure, efficient and reliable revenue collection mechanism for toll collection, parking and other transportation applications. It also allows anonymous prepayment charging.

CHAPTER THREE

3. The Approach for Modelling TDRP

3.1 Introduction

3.2 Theory of queuing and delay

3.2.1 Characteristics of queuing process.

3.2.2 Mathematical solution of queuing problems.

3.2.2.1 Steady-State Solutions.

3.2.2.2 Deterministic Solutions.

3.3 Time-Dependent queues and delays at road junctions

3.3.1 Time-dependent demand-capacity interaction at traffic junctions.

3.3.2 The basic queuing problem.

3.3.3 The average queue length.

3.3.4 The average vehicular delay.

3.4 Solutions of time-dependent queuing problem

3.4.1 Queuing probabilities.

3.4.1.1 Known initial queue length.

3.4.1.2 General probability distribution for initial queue length.

3.4.2 Average queue length.

3.4.3 Average Delays.

3.5 Co-ordinate Transformation solution for the time-dependent queuing problem

3.5.1 queue length (L) as a function of demand, capacity and time.

3.5.2 The co-ordinate transformation technique applied to vehicle delays.

3.5.3 Modification for accuracy.

3.6 Modelling the time-dependent delay imposed by each vehicle on other vehicles

3.7 Summary

3.1 Introduction

Delay at any type of road junctions arises from several sources. Firstly, vehicles have to slow down to negotiate the junction. In the immediate approach they have to respond to the system of control—traffic signals, major/minor or roundabout control— and to be ready to stop or to give way to other traffic streams. They may have to queue before they can enter the junction. As they depart, acceleration to normal running speed involves further delay.

In general, there are no very clearly defined boundaries between the various components of delay. However, two main components have been separated conceptually (see Maycock, 1974): “geometric” delay— the intrinsic delay arising from the need to slow down, negotiate junction and accelerate back to running speed— and “congestion” delay. The first is defined for single isolated vehicles, while the second arises from vehicle-vehicle interactions and mainly because of the delay due to waiting in the queue before clearing the junction. It is sometimes called queuing delay.

The objective of this chapter is to describe the derivation of the mathematical model developed for estimating the queuing delay imposed by each individual vehicle arriving at a certain time to a traffic junction on other vehicles behind. Since the prediction of queuing delays at road junctions subject to time-varying demand and capacity, the approach for estimating queuing delay, in this study, is based on the solutions of the time-dependent queuing problem. Therefore, sections 3.2 and 3.3 review the theory of queuing and delay and time-dependent queues and delays at road junctions respectively. The exact and the co-ordinate transformation solutions of the time-dependent queuing problem are presented in sections 3.4 and 3.5 respectively. Then, the co-ordinate transformation solution is used in section 3.6 to derive the mathematical model for estimating the time-dependent delay imposed by each vehicle on the other vehicles.

3.2 Theory of queuing and delay

Queuing situations arise in all aspects of work and life and are typified by the ‘queuing for service’. The classical prototype problem is the following: customers (for example people) arrive at certain time instants at a service point (at bank counter, an air ticket counter, a highway intersection, etc.). The service facility requires a certain time to serve each customer but is capable of serving only a finite number of customers at a time (for example, one). If customers arrive faster than the facility can serve, then customers must wait in a queue. Typically, both the customer arrivals and the service times are specified to have some given probability distributions. One wishes to relate the delays in queue, queue lengths, etc., to the given properties of the arrivals and service. In practical applications one frequently wishes further to compare the operation of several possible modes of operation with regard to its type of service, cost, etc. (see Newell, 1971). The theory of queuing gives a basis for understanding the various aspects of the problems and enables a quantitative assessment to be made. Therefore the theory enables these ‘service situations’ to be more effectively designed and operated.

3.2.1 Characteristics of queuing process

Gross and Harris (1985) have defined six basic characteristics of queuing processes. These basic characteristics are: –

- (i) Arrival pattern of customers.
- (ii) Service pattern of servers.
- (iii) Queue discipline.
- (iv) System capacity.
- (v) Number of service channels.
- (vi) Number of service stages.

In most cases, these six basic characteristics provide an adequate description of a queuing system. Each of these characteristics is briefly discussed.

(i) Arrival pattern of customers

The arrival pattern or input to a queuing system is often measured in terms of the average number of arrival per some unit of time (mean arrival rate) or by the average time between successive arrival (mean interval time). On the other hand, if there is uncertainty in the arrival pattern (often referred to as random, possibilities, or stochastic), then these mean values provide only measures of central tendency for the input process, and further characterisation is required in the form of the probability distribution associated with this random process.

Another factor of interest concerning the input process is the possibility that arrivals come in batches instead of one at a time, the input is said to occur in bulk or batches. In the bulk-arrival situation, not only may the time between successive arrivals of the batches be probabilistic, but also the number of customers in a batch (i.e. the batch size).

(ii) Service pattern of servers

Much of the discussion above concerning the arrival pattern is appropriate in discussing service patterns. For example, service patterns can also be described by rate (number of customers served per some unit time) or as a time (time required to service a customer). One important difference exists, however, between service and arrivals. When one speaks about service rate or service time, these terms are considered on the fact that the system is not empty; that is, there is someone in the system requiring service. If the system is empty, the service facility is idle. Service may also be deterministic or probabilistic; hence in the latter case the probability distributions associated with service are conditional, based on a non-empty system. Service may also be single or batch.

(iii) Queue discipline

Queuing discipline refers to the manner by which customers are selected for service when a queue has formed. The most common discipline that can be observed in everyday life is first come, first served (FCFS), or first in, first out (FIFO), as it is sometimes called. However, this certainly not the only possible queue discipline. Some others in common usage are last come, first served (LCFS), which is applicable to many inventory systems when there is no obsolescence of stored unit as it is easier to reach the nearest items which are last in; selection for service in random order independent of the time of the arrival to the queue (RSS); and a variety of priority schemes, where customers are given priorities, regardless of their time of arrival to the system.

(iv) System capacity

In some queuing processes there is a physical limitation to the amount of waiting room, so that when the line reaches a certain length, no further customers are allowed to enter until space becomes available by a service completion. These are referred to as finite queuing situation; that is, there is a finite limit to the maximum queue size.

(v) Number of service channels

The number of service channels refers to the number of parallel service stations which can service customers simultaneously (for example, parallel traffic lanes). The multi-channel systems differ from single-channel systems in that the latter has a single queue, while the former allows a queue for each channel. It is generally assumed that the service mechanisms of parallel channels operate independently of each other.

(vi) Number of service stages

A queuing system may have a single stage of service such as traffic at road junctions, or it may have several stages. An example of a multi-stages queuing system would be a physical examination procedure, where each patient must proceed through several stages, such as medical history; ear; nose, throat examination; and so forth.

The six characteristics of queuing system discussed above are generally sufficient to describe completely a process under study. One can see from this discussion thus far that there exists a wide variety of queuing systems that can be encountered. However, before performing any mathematical analysis, it is very necessary to describe adequately the process being modelled.

3.2.2 Mathematical solutions of queuing problems**3.2.2.1 Steady-State solutions**

Briefly, provided that the service channel is capable of serving at a faster average rate than that at which customers arrive, then the steady-state is reached when the queue behaves independently of the initial state of the system, and the probability of having a given number, n , say, in the queue remains constant with time. The state of the queue soon becomes independent of the starting conditions when the present pattern of production was begun (Murdoch, 1978). This situation exists, more or less, in some where like a bank, the system starts every day with no people at all either being served or queuing, and the chance of finding, say, 6 people queuing depends on how soon after opening time the observation is made. Assuming that there are always the same number of clerks on duty and customers arrive at a constant average rate throughout the day, the chance of finding 6 people queuing immediately after the bank opens is likely to be very small indeed. As the day proceeds, the queue gradually achieve a steady-state and eventually the queue fluctuates about a fixed average size, the probability of finding 6 people queuing now being higher than it was at the very start of the day's business.

In traffic engineering applications for the queuing theory, the customer in service corresponds to vehicle waiting at the give-way line or stop line of a junction or bottleneck, and in this case, the term 'queue length' is sometimes taken to include this vehicle. Therefore, two queue lengths can be defined as follows: –

- (i) L_q = the number of waiting vehicles, excluding that at the stop line or give-way line.
- (ii) L_s = the number of waiting vehicles, including that at the stop line or give-way line.

The steady-state solution for a single service facility with random arrivals and service time is given by: –

$$L_q = \frac{\rho^2}{1 - \rho} \dots\dots\dots (3.1)$$

$$L_s = \rho + L_q = \frac{\rho}{1 - \rho} \dots\dots\dots (3.2)$$

Where:

- ρ is the traffic intensity and equal to λ/μ .
- λ is the traffic demand in vehicles per second.
- μ is the capacity (or service rate) in vehicles per second.

(The mathematical derivation of Equations 3.1 and 3.2 is given by Satty (1961))

If the demand or service are non-random, these two relationships become as follows: –

$$L_q = \frac{C\rho^2}{1-\rho} \quad \dots\dots\dots (3.3)$$

$$L_s = \rho + \frac{C\rho^2}{1-\rho} \quad \dots\dots\dots (3.4)$$

Where C is a constant depending on the arrival pattern and service distribution, and is given by equation (3.5) below (see Kimber et al (1986)).

$$C = \frac{1}{2} \left[\left(\frac{\sigma_a}{\bar{T}_a} \right)^2 + \left(\frac{\sigma_s}{\bar{T}_s} \right)^2 \right] \quad \dots\dots\dots (3.5)$$

Where: $\sigma_a, \bar{T}_a$ are the standard deviation and mean of the arrival distribution.

$\sigma_s, \bar{T}_s$ are the standard deviation and mean of the service headway distribution.

For random arrivals and service, $(\sigma_a / \bar{T}_a) = (\sigma_s / \bar{T}_s) = 1$, so $C=1$ and the standard M/M/1 result is obtained. For random arrivals and regular service $(\sigma_a / \bar{T}_a) = 1, (\sigma_s / \bar{T}_s) = 0$. So $C=1/2$, giving the M/D/1 result.

The congestion delay per unit time d_{cu} is numerically equal to the queue length L_q and excludes the time spent by vehicles in service. The total delay per unit time d_u is equal to L_s includes the time spent by the vehicles in service. The congestion delay per vehicle d_{cv} is obtained by dividing the congestion delay per unit time by the number of vehicles arriving per unit time (i.e. $\lambda = \rho\mu$). Thus,

$$d_{cv} = \frac{C\rho^2}{(1-\rho)} / \rho\mu = \frac{C\rho}{\mu(1-\rho)} \quad \dots\dots\dots (3.6)$$

Similarly, the total delay per vehicle d_{tv} is

$$d_{tv} = \left(\rho + \frac{C\rho^2}{(1-\rho)} \right) / \rho\mu = \frac{1}{\mu} + \frac{C\rho}{\mu(1-\rho)} \quad \dots\dots\dots (3.7)$$

In most practical solutions only the steady-state need to be considered, but occasionally only the time-dependent solution is applicable since the queue never reaches a steady-state. This latter is generally true if the arrival rate is greater than the service rate ($\lambda \geq \mu$) and applies in some cases when the system is not in operation long enough before reverting to the starting state, usually with no customers at all in the system (Murdoch, 1978).

However, the time-dependent queuing problems and their solutions are discussed in some detail under sections 3.3, 3.4 and 3.5.

3.2.2.2 Deterministic Solutions

The conceptually simplest class of queuing problems are those for which probability distributions are not necessary to describe arrival and service patterns. Instead, the units of input arrive at known points in time and service times are fixed constants. A queuing model that falls into this class is said to be deterministic, since there are no probability distributions associated in any way with the problem (Gross and Harris, 1985).

Consider then the elementary case of a constant rate of arrivals to a single channel which possesses a constant service rate. These regularly spaced arrivals are to be serviced first come, first served (FCFS). Let it also be assumed that at time $t=0$ there are no customers waiting and that the channel is empty. Let λ be defined as the number of arrivals per unit time, and $1/\lambda$ then will be the constant time between successive arrivals. Similarly, if μ is to be the rate of service in terms of completion per unit time when the service is busy, then $1/\mu$ is the constant service time.

Considering first the case where $\lambda > \mu$ (i.e. $1/\lambda < 1/\mu$); that is the arrival rate is greater than the service rate. In this situation, the queue length would keep increasing and grow beyond any bound. Each successive customer would wait longer than his predecessor for service until eventually customers would be waiting for ever. To prevent this, forced balking is imposed on customers whenever the number in the system gets to a certain size (i.e. imposing a finite system-capacity constraint).

Under the assumption that as soon as a service is completed another is begun, the number in the system L_s (including the customer in service) at time t is determined by equation 3.8 below: –

$$L_s(t) = \{\text{number of arrivals in the interval } (0,t]\} - \{\text{number of service completed in the interval } (0,t]\}$$

$$\begin{aligned} &= \left[\frac{t}{1/\lambda} \right] - \left[\frac{t - 1/\lambda}{1/\mu} \right] \\ &= t(\lambda - \mu) + \mu / \lambda, \end{aligned} \quad \dots\dots\dots (3.8)$$

and the number of waiting vehicles L_q excluding that at the stop line or give-way line is given by:

$$L_q(t) = t.(\lambda - \mu) \quad \dots\dots\dots (3.9)$$

Thus far only the case where $1/\lambda < 1/\mu$ has been considered; when $1/\lambda \geq 1/\mu$ a very simple situation is considered since there is never more than one in the system. Hence L_s is either one or zero.

3.3 Time-Dependent queues and delays at road junctions

Queuing at road junctions is a time-dependent phenomenon. During peak periods traffic flows are frequently very close to the capacities of junctions. In such circumstances neither steady-state solutions nor deterministic solutions could provide a realistic basis for estimating queue lengths and delays. Steady-state queuing solution (as discussed in the previous section) is widely used but predicts infinite queues and delays when demand flow reaches the capacity available to it. This behaviour highlights the fundamental difficulty of applying this solution to road traffic; in reality, when the capacity is exceeded for short periods, the queue growth lags behind the expectations of steady-state solutions, and the rate of variation of demand and capacity cannot be ignored. Deterministic solutions, on the other hand, in which the delay is obtained as a simple integral of demand minus capacity, can sometimes be used when demand and capacity vary in time. However, this treatment ignores the statistical nature of traffic arrivals and departures, and leads to serious underestimates in the delay unless the capacity is exceeded by a considerable margin (Kimber and Hollis, 1979). When demand only just reaches capacity zero delays are predicted.

In practical terms, the most important region of operation is that where demand and capacity are approximately equal, and it is this region which inadequately represented both by the steady-state and deterministic approaches. Methods are therefore needed which adequately treat the whole range of demand and capacity, and which take proper account of the statistical nature of traffic and of the variations in time of both demand and capacity.

A time-dependent queuing theory has therefore been developed and tested (Kimber and Hollis (1979), Kimber et al (1986), and Burrow (1987)), to reflect the stochastic time dependent nature of the queuing process.

3.3.1 Time-dependent demand-capacity interaction at traffic junctions

Traffic at road junctions can usually be divided into a number of separately identifiable streams, each with a defined capacity and demand flow. The capacity depends in general on the space or time available to the stream and on the type of control. Whether or not queuing occurs in a given stream depends on the ratio of demand flow to capacity—i.e. the traffic intensity; if the ratio is very much less than unity queuing is rare, but otherwise it occurs to a greater or lesser extent. Whenever there is queuing, vehicles are delayed. As an example, Figure (3.1) below shows the traffic flows to be considered for three-arm major/minor priority junctions. Traffic approaching the minor road gives way to major road traffic, and, apart from major road vehicles turning right (movement 2-1), delays are experienced only by traffic in the left and right-turning minor road streams. These delays depend on the demand flow in these streams and the capacity available to them, both of which will vary in time. The capacity is determined at any time by the major road flows and can be calculated by linear regression techniques (see Kimber, 1976).

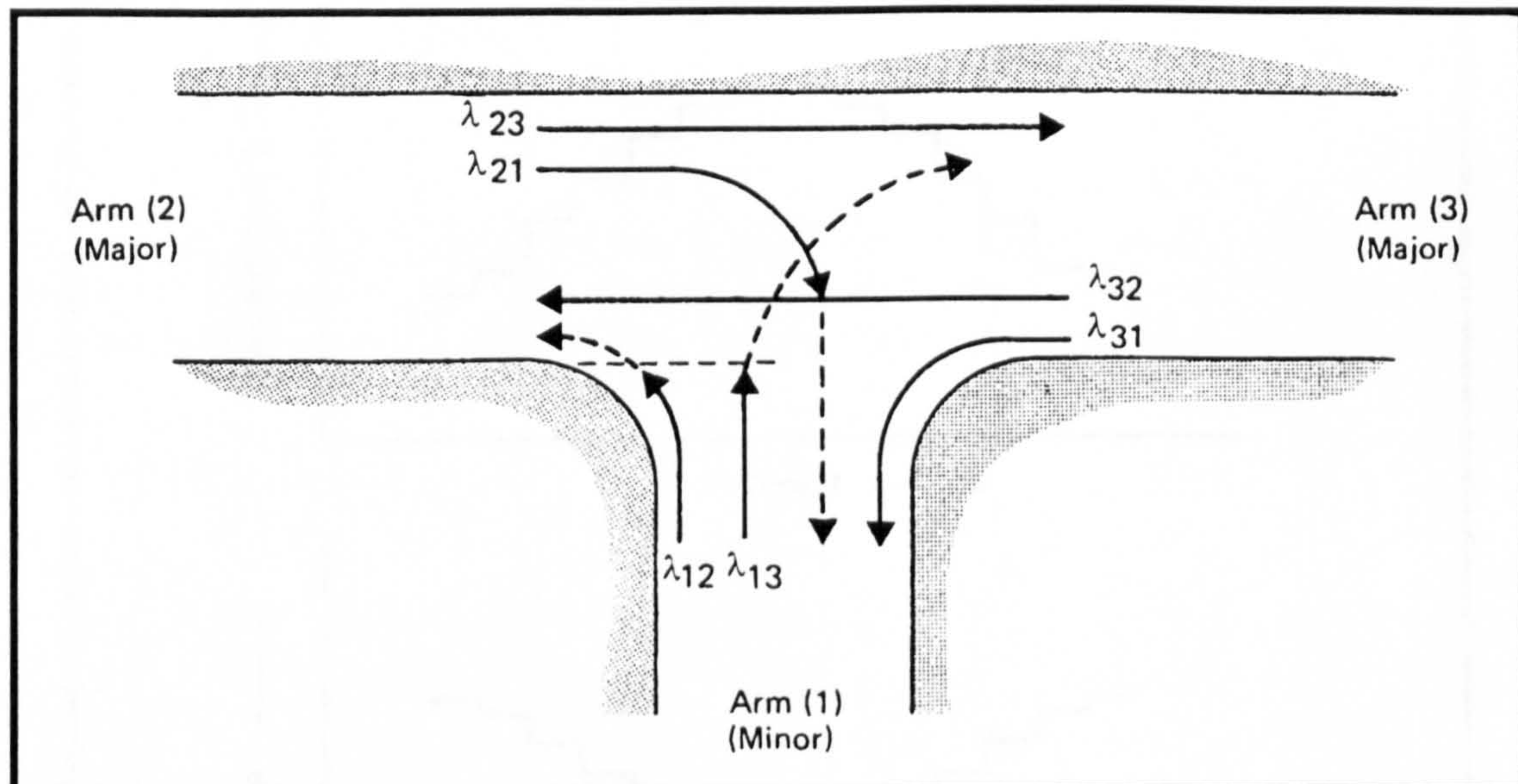


Figure (3.1): Traffic movements at a three-arm major/minor junction.

In the following sections, the queuing of vehicles in a single stream is considered; no assumption is made about the type of junction involved, nor about the nature of other, possible interacting, streams.

3.3.2 The basic queuing problem

Suppose the traffic demand in a stream is λ and the capacity available to it is μ (both expressed in vehicles per unit time). If the system is in equilibrium, i.e. the time-averaged values of λ and μ are stable over a long period, and the vehicle arrivals and the waiting times at the give-way line or stop line are randomly distributed, the probability of a queue of n vehicles is P_n :

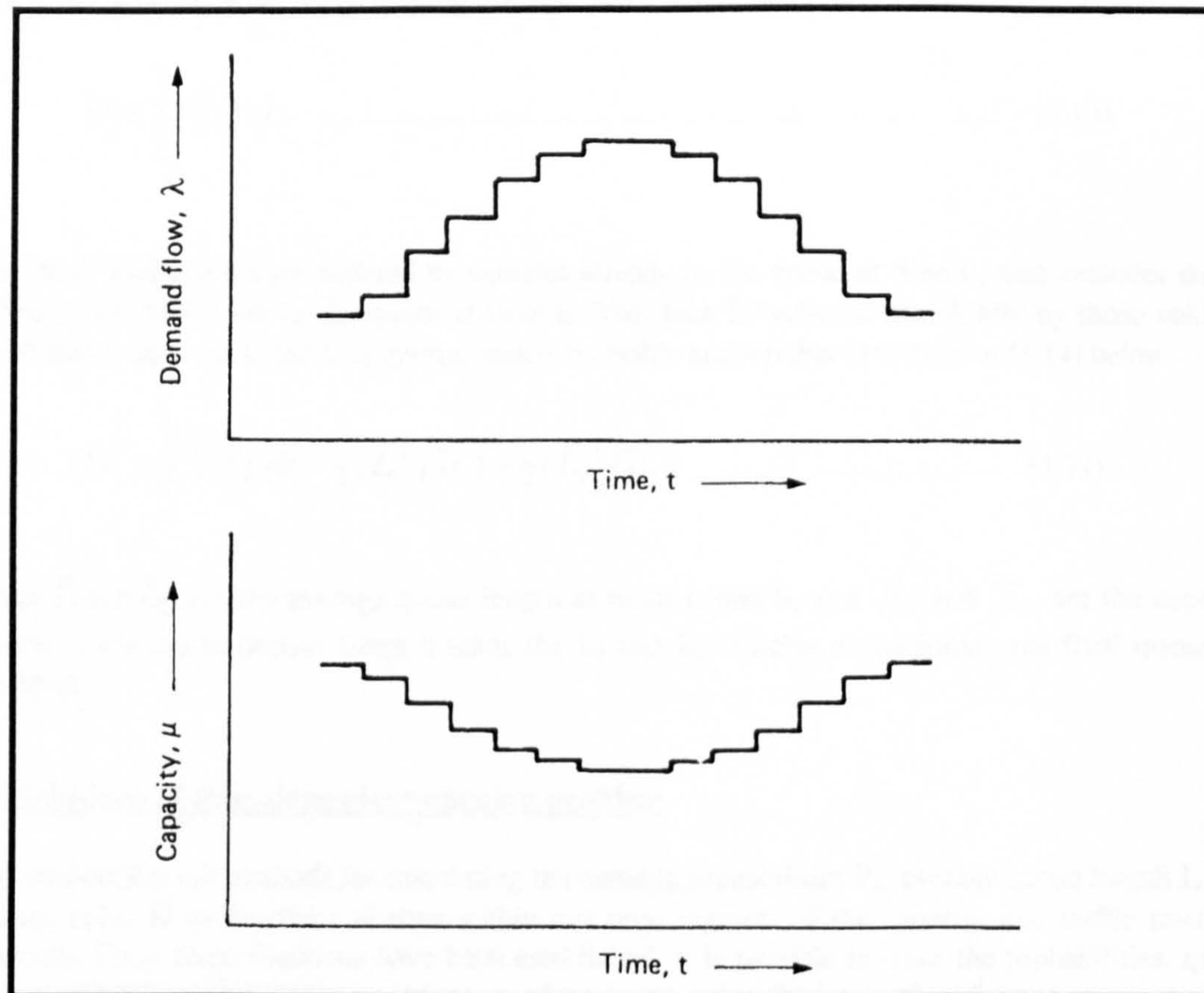
$$P_n = \rho^n (1 - \rho) \quad \dots\dots\dots (3.10)$$

Where ρ is the traffic intensity, i.e. $\rho = \lambda/\mu$.

The queue length fluctuates in time according to the probability distribution of equation (3.10), and P_n represents the proportion of the time for which there are n waiting vehicles.

Now suppose that the traffic demand and capacity vary in time, and that this variation is described by stepped functions of the form illustrated in Figure (3.2) below. The value of λ and μ can no longer be treated as long term averages in times; instead they represent average values at a given stage in repeated 'trials' with the defined sequence of demand and capacity. Each trial represents a possible set of arrivals to the queue and departures from it consistent with the prescribed sequence of λ and μ values. Consequently the queue evolution is different from trial to trial. The queuing probabilities are now time-dependent and can be interpreted as relative frequencies amongst the trials of occurrence of a queue of given length at a given time; thus $P_n(t)$ is proportion of trials for which there is a queue of n vehicles at time t .

The basic queuing problem is to determine the probabilities as functions of time, given the sequence of λ and μ values. Once this has been achieved, the most important quantities— the average queue length and the average vehicular delay— can be derived as functions of time from the probabilities $P_n(t)$.



(λ and μ are represented as averages in a sequence of short time segments)

Figure (3.2): Time-Varying Traffic Demand, λ and Capacity, μ

3.3.3 The Average queue length

The average queue length, L , evaluated over a large number of trials, varies with time, and can be obtained from:

$$L(t) = \sum_{n=1}^N n \cdot P_n(t) \quad \dots\dots\dots (3.11)$$

Where N is the maximum number of waiting vehicles that can be accommodated. $L(t)$ includes the vehicle "in service" i.e. at the give-way or stop line. Alternatively, the queue length $L'(t)$ excluding the vehicle in service is given by:

$$L'(t) = \sum_{n=2}^N (n-1) \cdot P_n(t) \quad \dots\dots\dots (3.12)$$

Since the $P_n(t)$ depends on the demand flow λ and capacity μ , and these are defined as stepped functions, not as continuous variables, queue length L has to be evaluated in successive time segments. The segment-width can be made as fine as the demand and capacity information allows.

3.3.4 The average vehicular delay

Once the variation of the average queue length in time has been established, it is a straightforward matter to calculate vehicular delays. The total delay, D , suffered between a time t_1 and a time t_2 is simply the area under the curve relating average queue length and time:

$$D = \int_{t_1}^{t_2} L(t) \cdot dt \quad \dots\dots\dots (3.13)$$

This delay includes delays suffered by vehicles already in the queue at time t_1 , and excludes delays suffered by vehicles left in the queue at time t_2 . The delay D' suffered specifically by those vehicles which arrive between t_1 and t_2 is approximated by Hollis and Kimber (1977) as in (3.14) below:

$$D' = \int_{t_1}^{t_2} \bar{L}(t) \cdot dt - \frac{1}{2} (\bar{L}_1^2 / \bar{\mu}_1) + \frac{1}{2} (\bar{L}_2^2 / \bar{\mu}_2) \quad \dots\dots\dots (3.14)$$

Where $\bar{L}_1$ and $\bar{L}_2$ are the average queue length at times t_1 and t_2 , and $\bar{\mu}_1$ and $\bar{\mu}_2$ are the capacity averages over the respective times it takes the L_1 and L_2 vehicles of the initial and final queues to discharge.

3.4 Solutions of time-dependent queuing problem

This section sets out methods for calculating the queuing probabilities P_n , average queue length L and average delay D as functions of time within any time segment of the capacity and traffic intensity functions. Once these functions have been established, it is possible to trace the probabilities, queue lengths and delays throughout any sequence of segments, using the final values for one segment as the initial values for the next and so forth.

P_n and L depend on:

- The value of capacity and traffic intensity within the segment, and
- The probability distribution for the queue length at the beginning of the segment.

The random arrival/random service queuing model, M/M/1, is employed in this section, more general cases in which the traffic arrival and the service mechanism for the stream are not completely random are discussed in the next section (i.e. section 3.5). For computing reasons, it is convenient in the following sections to use solutions for queues in which the maximum number of vehicles permitted,

N , is finite, but to set N to a large value in order to simulate the case of unrestricted waiting space. Usually $N=500$ suffices for this purpose (Kimber and Hollis, 1979).

Suppose the time segment considered represent the interval beginning at t_I and ending at t_F .

3.4.1 Queuing probabilities

3.4.1.1 Known initial queue length

The simplest case is when the queue length at time t_I is known with certainty; say it is m vehicles. Then:

$$P_{n=m}(t_I) = 1$$

$$P_{n \neq m}(t_I) = 0$$

The probabilities evolve with time according to:

$$P_n^m(t) = P_n + \frac{2\rho^{\frac{1}{2}(n-m)}}{N+1} \sum_{i=1}^N \frac{1}{x_i} \left\{ \sin \frac{im\pi}{N+1} - \rho^{\frac{1}{2}} \sin \frac{i(m+1)\pi}{N+1} \right\} \left\{ \sin \frac{in\pi}{N+1} - \rho^{\frac{1}{2}} \sin \frac{i(n+1)\pi}{N+1} \right\} e^{-\mu x_i(t-t_I)} \quad (\text{for } t > t_I) \quad (3.15)$$

Where:

- $P_n^m(t)$ is the probability of a queue of n vehicles at time t , given that there were m vehicles at time t_I ;
- $P_n = [(1-\rho)/(1-\rho^{N+1})]\rho^n$, is the equilibrium probability (i.e. the probability at $t=\infty$, if ρ remains constant) of finding n waiting vehicles for a finite queue (for $\rho < 1$, $P_n \rightarrow \rho^n(1-\rho)$ as $N \rightarrow \infty$); and
- $x_i = 1 + \rho - 2\rho^{\frac{1}{2}} \cos \frac{i\pi}{N+1}$.

Some useful examples of the time evolution of the queuing probabilities in the simple case where the initial queue length is zero are given by Kimber and Hollis (1979) and shown in Figure (3.3).

3.4.1.2 General probability distribution for the initial queue length

Usually the initial queue length is not known with certainty, and the initial probabilities extend over a range of possible lengths. The solution is then:

$$P_n(t) = \sum_{m=0}^N P_n^m(t) P_m(t_I) \quad (3.16)$$

Where $P_m(t_I)$ is the initial probability of a queue of length m .

3.4.2 Average queue length

Given the probabilities $P_n(t)$ at any time t within a time segment, the average queue length can immediately be calculated according to equation (3.11). By the time the end of the segment is reached there is a range of queuing probabilities, so that for the next segment the initial queue length is not

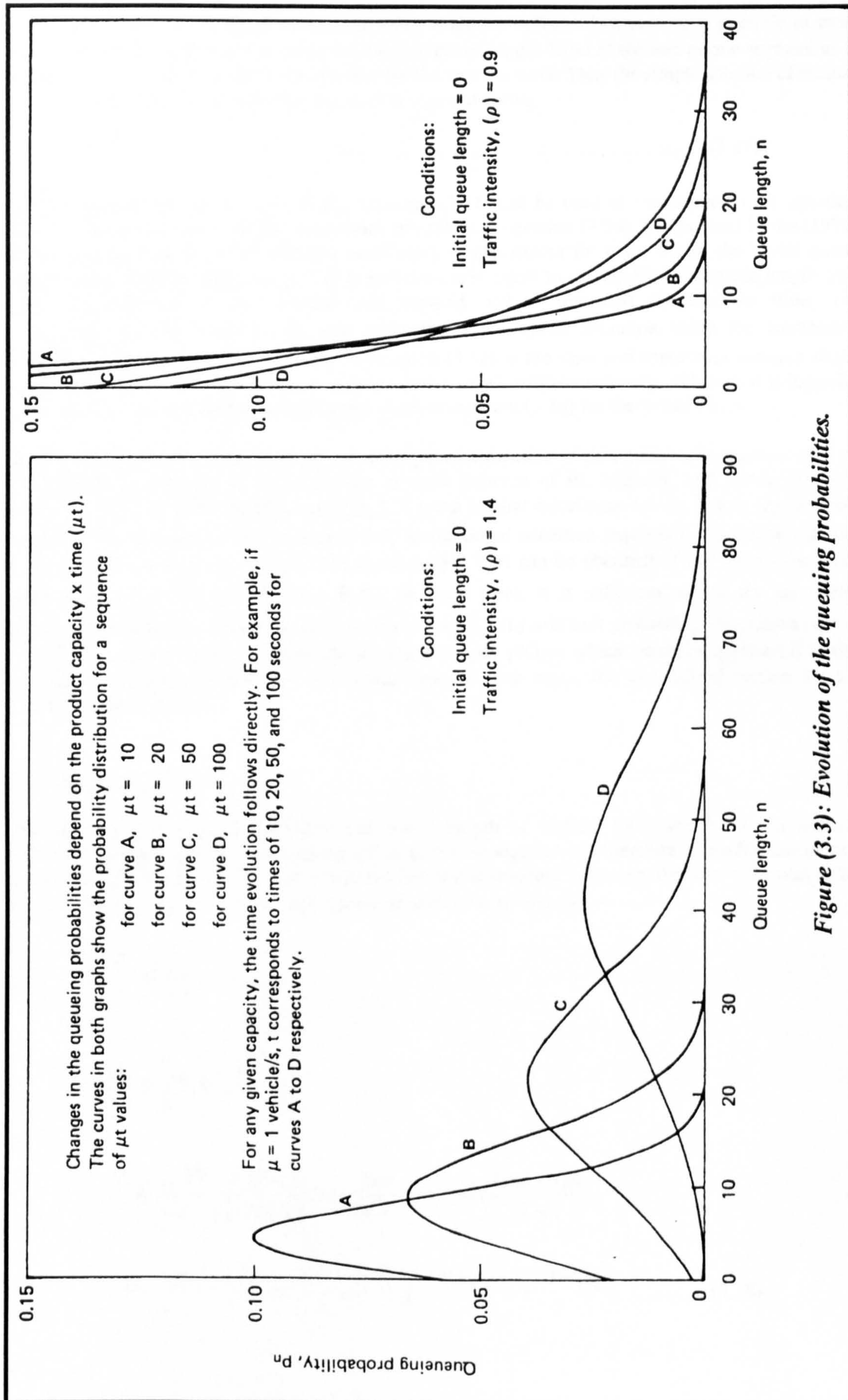


Figure (3.3): Evolution of the queuing probabilities.

known with certainty, although its average value is clearly defined. However, it is possible in most cases to simplify the problem by using the average queue length $L(t_F)$ at the end of one segment as if it were a known ('single value') initial queue for the next segment. Thus the simple solution of section 3.4.1.1 can be used repetitively from segment to segment taking

$$L(t_F) \rightarrow m \quad \dots\dots\dots (3.17)$$

at each segment boundary. Accordingly, equation (3.11) can be used in conjunction with equation (3.15), without recourse to the full summation procedure of equation (3.16). Kimber and Hollis (1979) concluded that this procedure provides satisfactory results except for cases where the initial queue length (determined by equation (3.17)) is approximately equal to the equilibrium queue length that would be obtained if the demand and capacity values remained constant in time; i.e. $m \cong \rho / (1 - \rho)$ for large N . In such cases the initial queue evolution using the elementary probabilities of equation (3.15) directly in equation (3.11) is too slow and sometimes shows a slight oscillatory behaviour. This does not impair the queue length estimate seriously, although it is logically better in these cases to employ the full summation of equation (3.16) for the probabilities.

Simply, the evaluation of the average queue length as a function of time within the segment may be summarised as follows. At the beginning of each segment of the capacity and traffic intensity functions, the probability distribution $P_n(t_i)$ is given by that calculation for the end of the previous segment. This distribution can be used directly as the initial condition required to obtain the solution at the end of the segment $P_n(t_F)$. The queue length $L(t)$ can be obtained at any time t from the probability distribution at that time, $P_n(t)$. In most cases, it is sufficient to use for the initial probability distribution a single defined queue of length $L(t_i)$ and unit probability. Exceptions occur when $L(t_i)$ is nearly equal to the equilibrium queue length $\rho/(1-\rho)$ which would be obtained if traffic intensity ρ remained constant for an infinite time. In such cases, the methods of section 3.4.1.2 provide a better solution.

3.4.3 Average delays

The expressions for the probabilities and queue length of section 3.4.1 and 3.4.2 are directly integrable, and average delays occurring within each time segment can therefore be readily calculated. For example, using the simple case where the initial queue length is known, the total vehicular delay occurring in the segment which begins at t_i and ends at t_F is $D(t_i, t_F)$: -

$$\begin{aligned} D(t_i, t_F) &= \int_{t_i}^{t_F} L(t).dt \\ &= \sum_{n=0}^N n. \int_{t_i}^{t_F} P_n(t).dt \\ &= \sum_{n=0}^N n \frac{2\rho^{\frac{1}{2}(n-m)}}{(N+1)\mu} \sum_{i=0}^N \frac{1}{x_i^2} \left\{ \sin \frac{im\pi}{N+1} - \rho^{\frac{1}{2}} \sin \frac{i(m+1)\pi}{N+1} \right\} \cdot \\ &\quad \left\{ \sin \frac{in\pi}{N+1} - \rho^{\frac{1}{2}} \sin \frac{i(n+1)\pi}{N+1} \right\} \cdot \{e^{-\mu x_i t_i} - e^{-\mu x_i t_F}\} \quad \dots\dots\dots (3.18) \end{aligned}$$

3.5 Co-ordinate transformation solution for the time-dependent queuing problem

The methods described in section 3.4 can be used to obtain queue length and delay predictions for systems in which capacity and demand vary in time; they also allow the calculation of more detailed queuing information, such as the time-dependence of the variance of queue length and any other quantity which can be derived from the probability distribution functions. However, the computer calculations required are in some cases lengthy, and for quick and convenient application it is necessary to use approximation methods. Therefore, this section describes a method developed for this purpose; its use is justified on the basis that it gives predictions close to those of section 3.4. This method employs a co-ordinate transformation technique to smooth the steady-state stochastic relationship for queue length or vehicle delay into the over-capacity deterministic results obtained by integrating the excess demand over capacity. The resulting formulae are used to predict directly the queue lengths and vehicular delays as functions of time within which successive segments defined by the intensity-capacity functions. A more detailed description of the derivation of these formulae is given by Kimber (1978) and Hollis and Kimber (1977).

3.5.1 Queue length (L) as a function of demand, capacity and time.

Suppose, at time $t=0$, there are L_0 waiting vehicles and the traffic intensity changes rapidly to ρ . In deterministic terms the queue simply grows or decays according to the difference between demand and capacity. Thus, after time t the number of waiting vehicles is:

$$\begin{aligned} L &= (\lambda - \mu).t + L_0 \\ &= (\rho - 1).\mu t + L_0 \end{aligned} \quad (3.19)$$

The steady-state result for the queue length L including the vehicle in service when the intensity is ρ , is given by:

$$L = \rho + C\rho^2/(1 - \rho) \quad (3.20)$$

Where C (as defined earlier) is a constant depending on the arrival and service patterns; for regular arrivals and service $C=0$, for random arrivals and service $C=1$. For simplicity consider this latter case; then

$$L = \rho/(1 - \rho) \quad (3.21)$$

The co-ordinate transformation technique is to transform the steady-state result (equation 3.21) so that instead of L becoming infinite at $\rho=1$, it approaches the deterministic equation (3.19) (for each value of time t) at high ρ values. Figure (3.4) illustrates the process graphically for the simple case $L_0=0$. For a given number of waiting vehicles, L , the steady-state intensity ρ_e is transformed into the new value ρ_n such that, $X=Y$ as in shown in Figure (3.4); this leaves the deterministic line as asymptote to the transformed curve. Thus

$$1 - \rho_e = \rho' - \rho_n$$

and hence

$$\rho_e = \rho_n - (\rho' - 1)$$

Where ρ' is the intensity corresponding to L in the deterministic case; from equation (3.19):

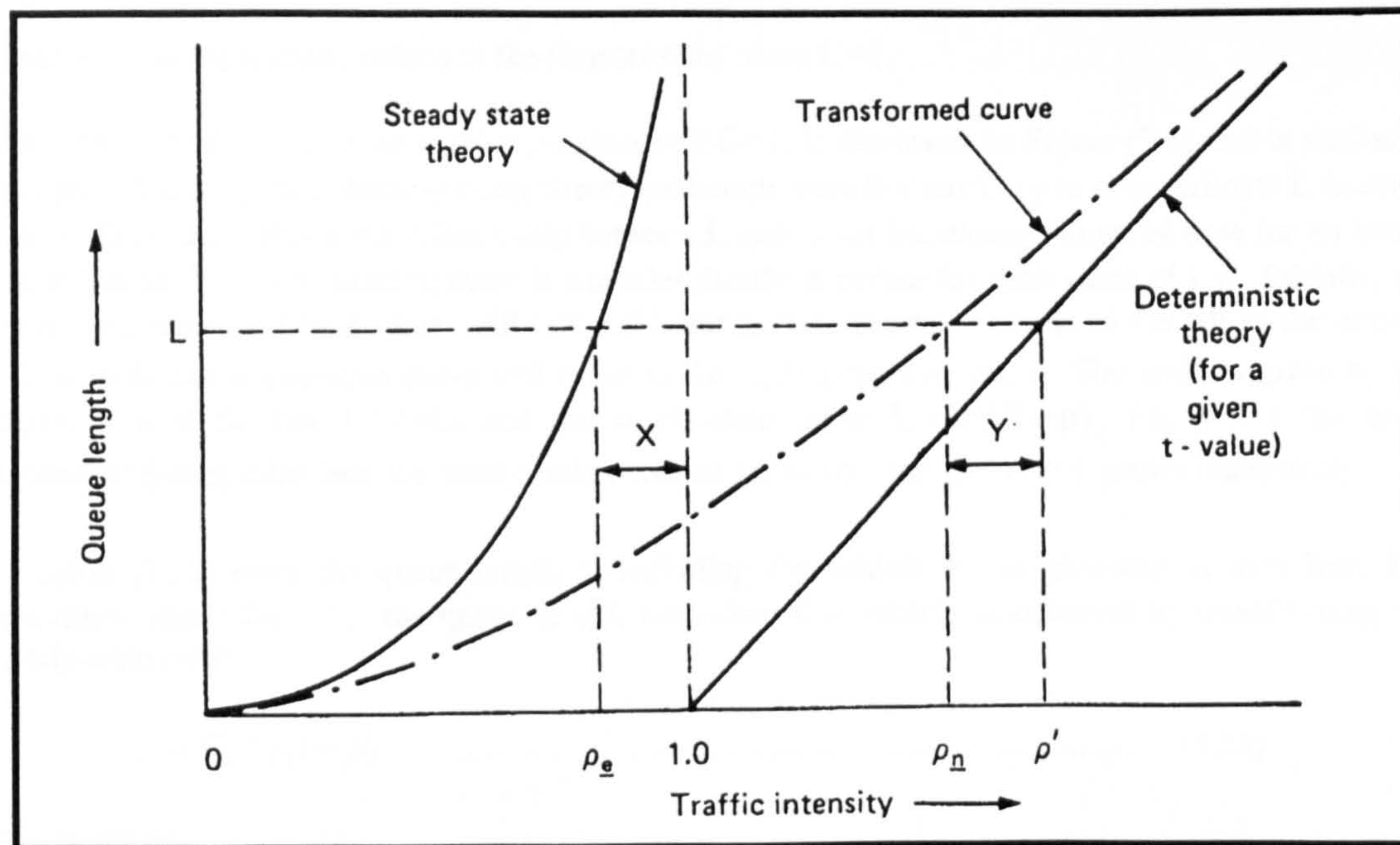


Figure (3.4): The principle of the Co-ordinate Transformation Method.

$$\rho' = \frac{L - L_0}{\mu} + 1$$

and the transformation is equivalent to setting

$$\rho_e = \rho_n - (L - L_0)/\mu t$$

in the steady-state equation (3.21), the transformed curve is therefore given by:

$$\frac{\rho_e}{1 - \rho_e} \rightarrow \frac{\rho_n - (L - L_0)/\mu t}{1 - \rho_n + (L - L_0)/\mu t} = L$$

Solving for L, the required result is:

$$L = \frac{1}{2}((A^2 + B)^{1/2} - A) \quad (3.22)$$

Where

$$A = (1 - \rho)\mu t + 1 - L_0$$

$$B = 4(L_0 + \rho\mu t)$$

(Only the positive root of the quadratic equation in L is required, see Kimber and Hollis (1979)).

If the more general steady-state equation (3.20), is used, the result is equation (3.22) with the following values of A and B.

$$A = \frac{(1 - \rho)(\mu t)^2 + (1 - L_0)\mu t - 2(1 - C)(L_0 + \rho\mu t)}{\mu t + (1 - C)}$$

$$B = \frac{4(L_0 + \rho\mu t)[\mu t - (1 - C)(L_0 + \rho\mu t)]}{\mu t + (1 - C)}$$

Clearly, these expressions reduce to the simpler ones when $C=1$.

The evolution of L in time for various ρ -values and $C=1$, is illustrated in Figure (3.5) and is similar to that predicted by probabilistic queuing theory (although there is a tendency to over estimate L in some cases). This figure shows the relationship between L and ρ for increasing values of time for an initial queue length, L_0 , of 5 vehicles; there is a similar family of curves for each value of L_0 . Initially, the curves are equivalent to the line AXB ($\mu t = 0$), but as time increases points to the left of the node X fall towards the steady-state curve and those to the right grow towards it. The node is given by the intersection of the line $L(0)=L_0$ and the steady-state curve $L = \rho/(1-\rho)$. For $\rho < 1$ the time-dependent curves relax into the steady-state curve as $\mu t \rightarrow \infty$, but for $\rho \geq 1$ grows indefinitely.

Equation (3.22) gives the queue length L including the vehicle at the give-way or stop line. The equivalent result for L' , the queue length excluding this vehicle is obtained by transforming the steady-state result:

$$L' = C\rho^2/(1-\rho) \quad \dots\dots\dots (3.23)$$

The result is:

$$L' = \frac{1}{2}((D^2 + E)^{1/2} - D) \quad \dots\dots\dots (3.24)$$

Where

$$D = \frac{(1-\rho)(\mu t)^2 - \mu t L'_0 + 2C(L'_0 + \rho \mu t)}{(\mu t - C)}$$

$$E = \frac{4C(L'_0 + \rho \mu t)^2}{(\mu t - C)}$$

and

$$L'_0 = L'(t = 0)$$

3.5.2 The co-ordinate transformation technique applied to vehicle delay

(a) The delay per unit time

The delay per unit time can easily be obtained (as in equation 3.13) by evaluating the relevant area under the queue length curve given either by equation (3.22) or equation (3.24). However, it is possible to apply a co-ordinate transformation to the delay expression directly. Suppose that, $L=L_0$ at time $t=0$ and the traffic intensity changes rapidly to ρ at that time. Then in deterministic terms the delay per unit time, $D_1(d)$, is given by:

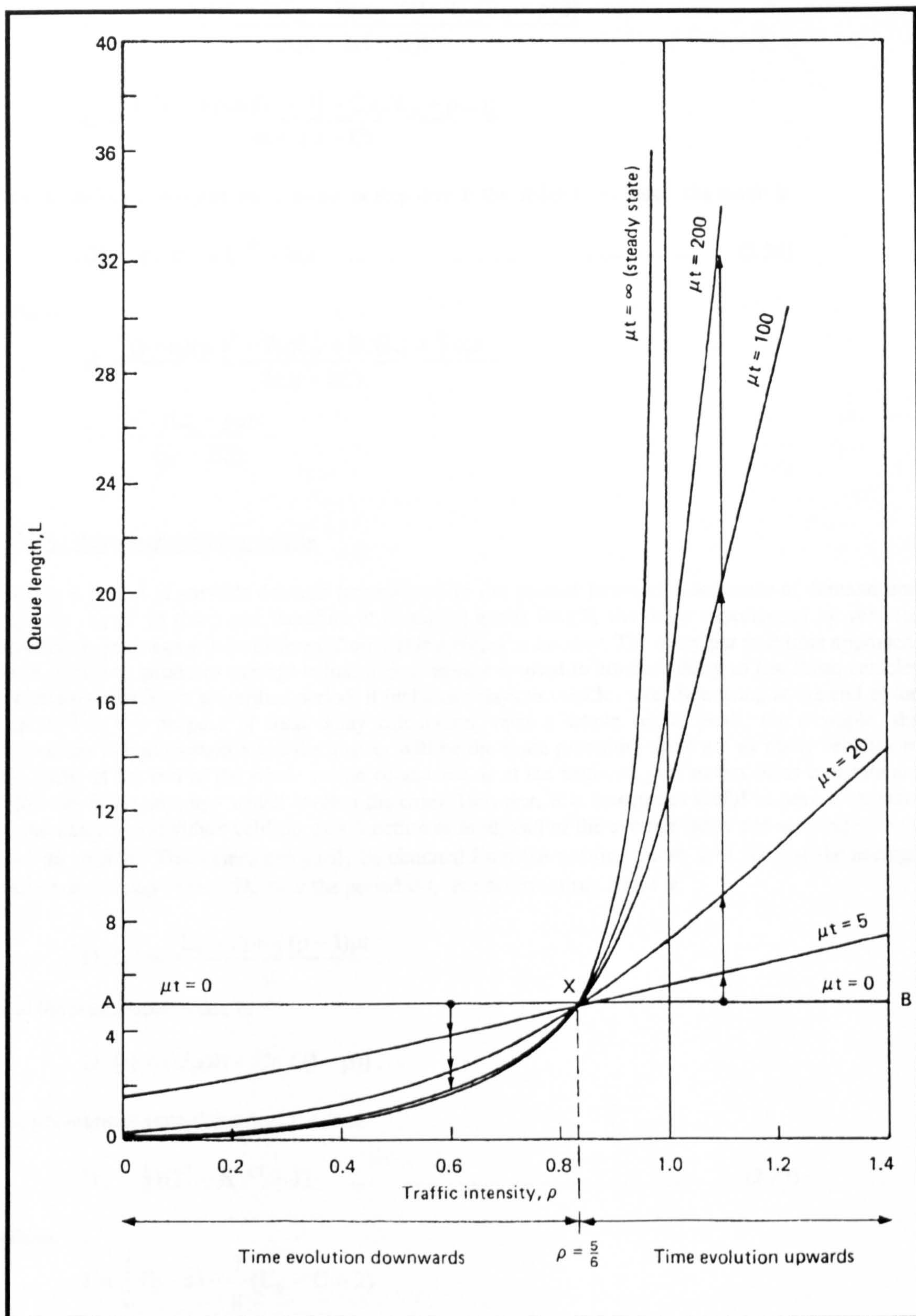
$$D_1(d) = \frac{1}{2}(\rho - 1)\mu t + L_0$$

The steady-state delay per unit time $D_1(s)$, is equal to L in equation (3.20).

$$D_1(s) = \rho + C\rho^2 / (1 - \rho)$$

Transforming this equation in exactly the same way as described for queue lengths in the previous section (i.e. section 3.5.1) gives the result:

$$D_1 = \frac{1}{2}((F^2 + G)^{1/2} - F) \quad \dots\dots\dots (3.25)$$



Source: Kimber and Hollis (1979).

Figure (3.5): The average queue length L as a function of the traffic intensity ρ according to the co-ordinate transformation method.

Where

$$F = \frac{(1-\rho)(\mu t)^2 - 2(L_0 - 1)\mu t - 4(1-C)(L_0 + \rho\mu t)}{2(\mu t + 2(1-C))}$$

$$G = \frac{2(2L_0 + \rho\mu t)[\mu t - (1-C)(2L_0 + \rho\mu t)]}{\mu t + 2(1-C)}$$

This includes the delay at the give-way or stop line. If this delay is excluded, the result is:

$$D'_1 = \frac{1}{2}((H^2 + I)^{1/2} - H) \quad \dots\dots\dots (3.26)$$

Where

$$H = \frac{(1-\rho)(\mu t)^2 - 2\mu t L'_0 + 8C(L'_0 + \frac{1}{2}\rho\mu t)}{2(\mu t - 2C)}$$

$$I = \frac{2C(2L'_0 + \rho\mu t)^2}{(\mu t - 2C)}$$

(b) The delay per arriving vehicle

During a period of variable demand (represented in the present terms as a sequence of demand and capacity 'steps' in time) and therefore of changing queue length, the delay experienced by vehicles arriving at one stage will be different from those arriving at another. The delay per unit time approach, as in (a) above, produces average values and so cannot be used to attribute delay to just those vehicles which arrive within a prescribed period: it includes delays to vehicles already waiting at the end of the period. For the purpose of total delay calculation, over a whole traffic peak, for example, the distinction is unimportant since the answer will be the same providing there are as many vehicles in the queue at the end of the whole period considered as at the beginning (or unless these numbers are relatively small anyway— which is often the case). However, it is sometimes useful to have a measure of the delay to individual vehicles as a function of time, and of the average delay per arriving vehicle over an interval. The former can easily be obtained from the queuing curve, $L=L(t)$. For the average delay per arriving vehicle D_v over the period 0-t, the deterministic result is:

$$D_v(d) = \frac{(L_0 + 1) + \frac{1}{2}(\rho - 1)\mu t}{\mu},$$

and the steady-state result is:

$$D_v(s) = (1/\mu)(1 + C\rho / (1 - \rho));$$

the transformed time-dependent result is:

$$D_v = \frac{1}{2}((J^2 + K)^{1/2} - J) \quad \dots\dots\dots (3.27)$$

Where

$$J = \frac{t}{2}(1 - \rho) - \frac{1}{\mu}(L_0 - C + 2)$$

$$K = \frac{4}{\mu} \left[\frac{t}{2}(1 - \rho) + \frac{1}{2}\rho t C - \left(\frac{L_0 + 1}{\mu} \right)(1 - C) \right]$$

This includes the delay at the give-way or stop line. If this delay is excluded, the result is: —

$$D'_v = \frac{1}{2}((P^2 + Q)^{1/2} - P) \quad \dots\dots\dots (3.28)$$

Where

$$P = \frac{1}{2}(1 - \rho)t - \frac{1}{\mu}(L_0 - C)$$

$$Q = \frac{2Ct}{\mu}(\rho + \frac{2L_0}{\mu t})$$

3.5.3 Modification for accuracy

The formulae of section 3.5.1 and 3.5.2 are not rigorous in the way that time-dependent probabilistic queuing theory expressions are, but provide extremely useful and tractable forms for the queuing relationships. In limiting cases (high and low ρ and t) their results are correct, and in the intermediate regions their functional behaviour is sensible. However, in some cases they over-estimate the queue length at intermediate values of time by about one to two vehicles (and the delay by about one to two service time per vehicle, i.e. by about $(1/\mu)$ to $(2/\mu)$ per vehicle). For some purposes these errors are not serious, but when they could be, it is advisable to adopt another procedure developed by Hollis and Kimber (1977) for cases where the accuracy of the formulae of sections 3.5.1 and 3.5.2 are not adequate. The expressions for queue growth and decay, in this procedure, are derived from the special case of equation (3.22) with $L_0=0$, suitably modified to provide for all starting conditions. A full description of this procedure is given by Hollis and Kimber (1977).

3.6 Modelling the time-dependent delay imposed by each vehicle on other vehicles

Suppose the time variations of traffic demand and capacity, for a certain junction, are given in the form of 'histogram' of equal strip-width (5, 10, 15, 20 or 30 mins.), and the initial queue length at t_i (the beginning of the study period) is zero (i.e. $L(t_i)=0$). For each time strip the initial queue length would be taken as the average queue length calculated at the end of the previous strip, then the demand and capacity values of this strip are used to calculate the queue length as a function of time until the beginning of the next strip and so forth. By repeating this process, for example, over the peak period (or the whole day), the queue evolution can be built up according to the given time variation of demand and capacity over the entire period. Once the queue length is determined in this way, it could be immediately possible to obtain the total delay between the initial time t_i and the time at which the queue is entirely dispersed t_F from equation (3.29) below.

$$D = \sum_{n=0}^N n \cdot \int_{t_i}^{t_F} P_n^m(t) \cdot dt \quad \dots\dots\dots (3.29)$$

To estimate the delay imposed by any vehicle arriving at time t_a on the rest of the traffic, it is necessary to know the effect of an additional vehicle at that time in increasing the total delay over the whole period between t_i and t_F . Then the difference between this total delay and the total delay before adding this vehicle will represent the delay imposed by the vehicle arriving at time t_a on the other vehicles behind (including the delay experienced by the arriving vehicle itself).

The process of estimating the delay imposed by a vehicle arriving at time t_a on other vehicles can be summarised in the following steps:-

1. Suppose that the initial queue length at the beginning of the study period is known, then the queue evolution can be built up according to the given time variation of demand and capacity over the entire period (as explained above).

2. From (1) above, the queue length at any time t_a is calculated, say m , and the total delay D between t_a and the time at which the queue is entirely dispersed T , can be calculated from the area under the queue length curve as in equation (3.30) below.

$$D = \sum_{n=0}^N n \cdot \int_{t_a}^T P_n^m(t) \cdot dt \quad \dots\dots\dots (3.30)$$

3. Increase the queue length at time t_a by one vehicle, (i.e. it becomes $m+1$), and build up the queue evolution over the entire period (after t_a) according to this change in the initial queue length and the variation of demand and capacity.
4. As in (2) above, the total delay D' between the time t_a and the time at which the new queue is entirely dispersed T' is given by:-

$$D' = \sum_{n=0}^N n \cdot \int_{t_a}^{T'} P_n^{(m+1)}(t) \cdot dt \quad \dots\dots\dots (3.31)$$

5. The delay d' imposed by the vehicle arriving at time t_a on other vehicles behind is then given by:

$$\begin{aligned} d' &= D' - D \\ &= \sum_{n=0}^N n \cdot \int_{t_a}^{T'} P_n^{(m+1)}(t) \cdot dt - \sum_{n=0}^N n \cdot \int_{t_a}^T P_n^m(t) \cdot dt \\ &= \sum_{n=0}^N n \cdot \left[\int_{t_a}^{T'} P_n^{(m+1)}(t) \cdot dt - \int_{t_a}^T P_n^m(t) \cdot dt \right] \quad \dots\dots\dots (3.32) \end{aligned}$$

Equation (3.32) is simplified mathematically (see Appendix (A)) using the approximate solution of the time-dependent problem (discussed in section 3.5) and yields:

$$d' = \begin{cases} T - t_a & \dots\dots\dots t_a \in [T_0, T] \\ 0 & \dots\dots\dots \text{otherwise} \end{cases} \quad \dots\dots\dots (3.33)$$

Where T_0 and T (or t_q and t_q') are the queue starting and vanishing times measured from the beginning of the study period respectively. It should be noted that the difference between T and T' is equal to $1/\mu$, however, for simplicity it is assumed that $T \cong T'$.

This result is also demonstrated using a numerical example. The time variations of queue length and delay at a simple bottleneck is examined during a three-hour study period (from 7:00 a.m. to 10:00 a.m.) under the demand and capacity rates depicted in Figure (3.6). Two solutions are adopted in this example: the deterministic queuing approach and the approximate solution of the time-dependent queuing problem. The random arrival/random service queuing model, $M/M/1$, is employed for the second solution (i.e. both the arrival and service distributions are given by a negative exponential function, in other words the randomness constant C is equal to 1). Figures (3.7) and (3.8) illustrate the variation of queue length and the total delay over the study period. It is shown in Figure (3.7) that until the demand exceeds the capacity at 7:45 a.m., no queue develops by the deterministic solution, while the expected queue given by the time-dependent solution started to build up earlier (at 7:30 a.m.) and

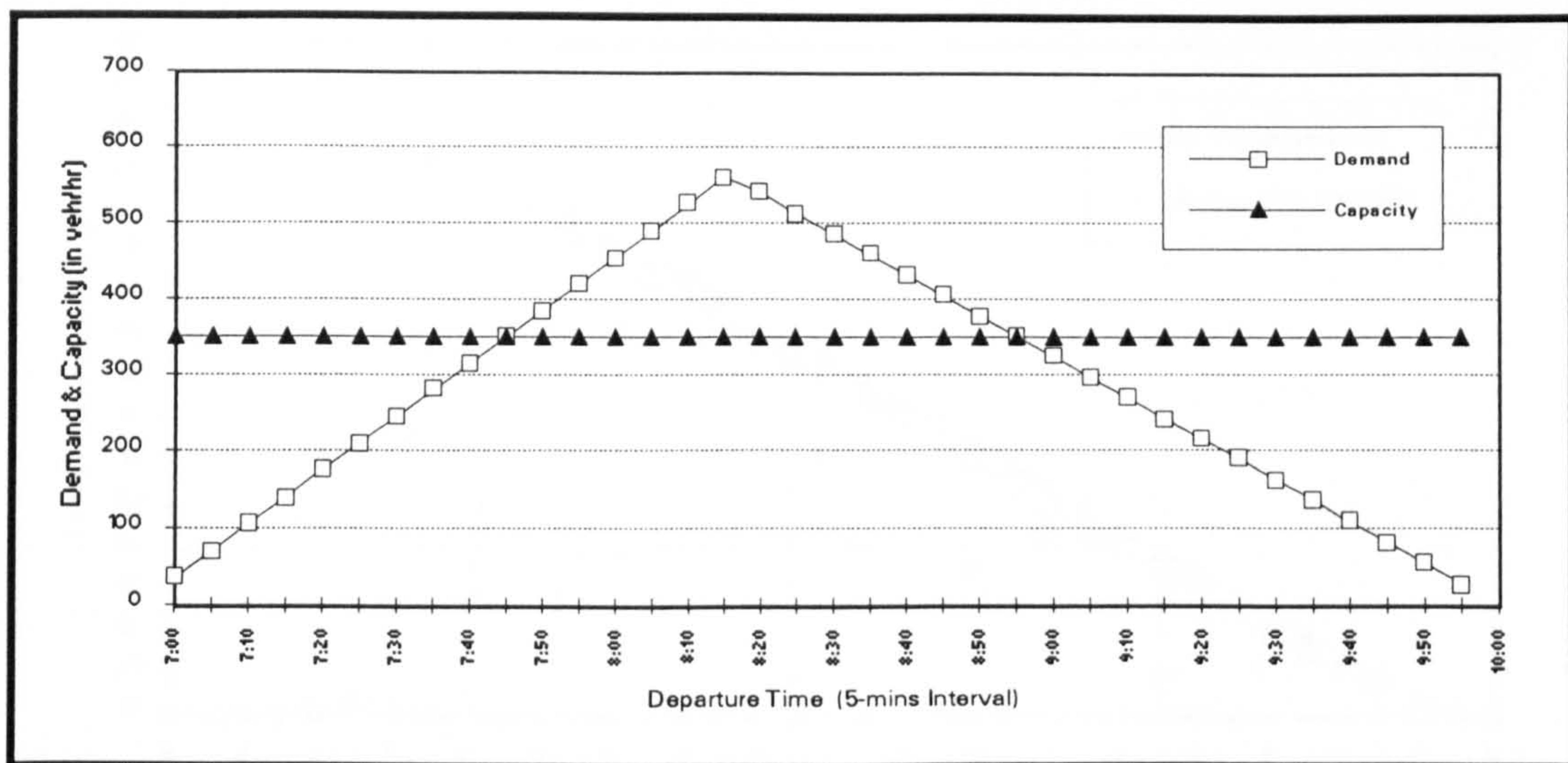


Figure (3.6): The Traffic Demand and Capacity During the Peak Period.

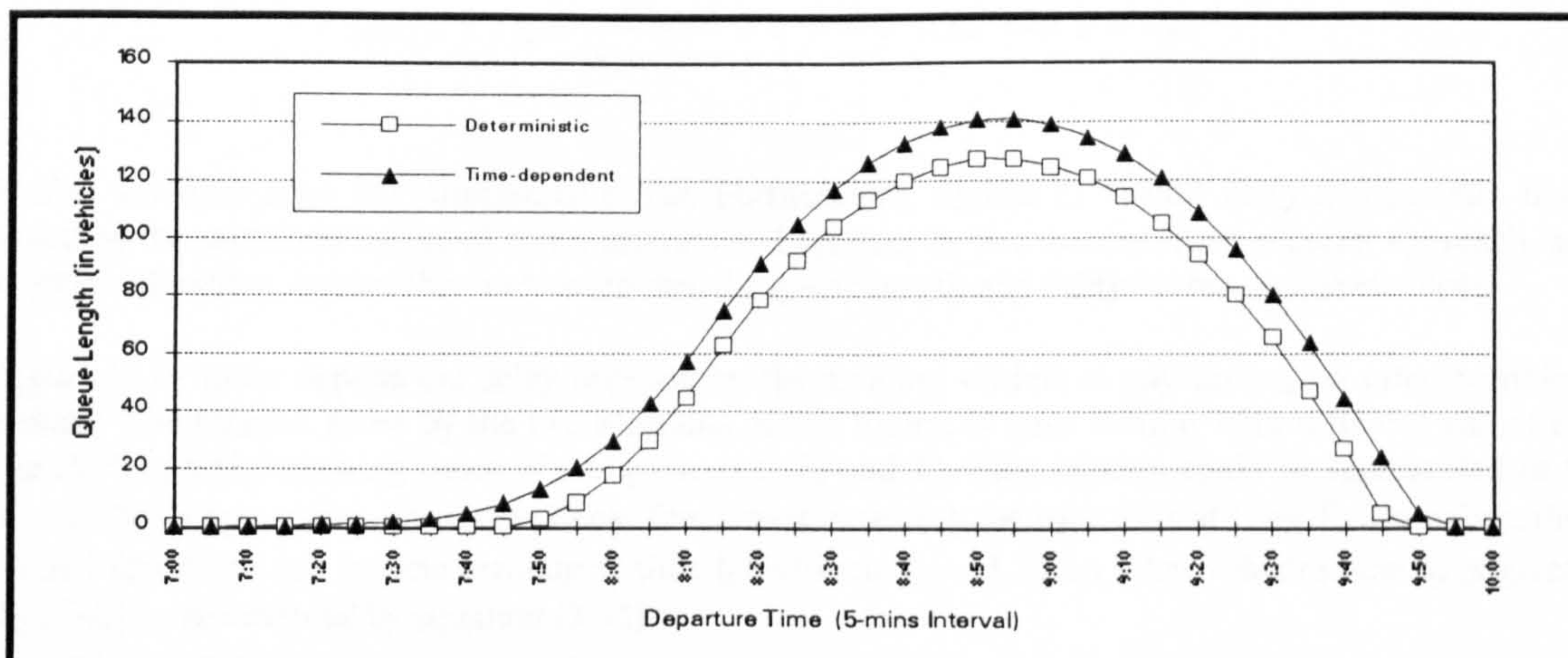


Figure (3.7): The Queue Length Distribution Over the Peak Period.

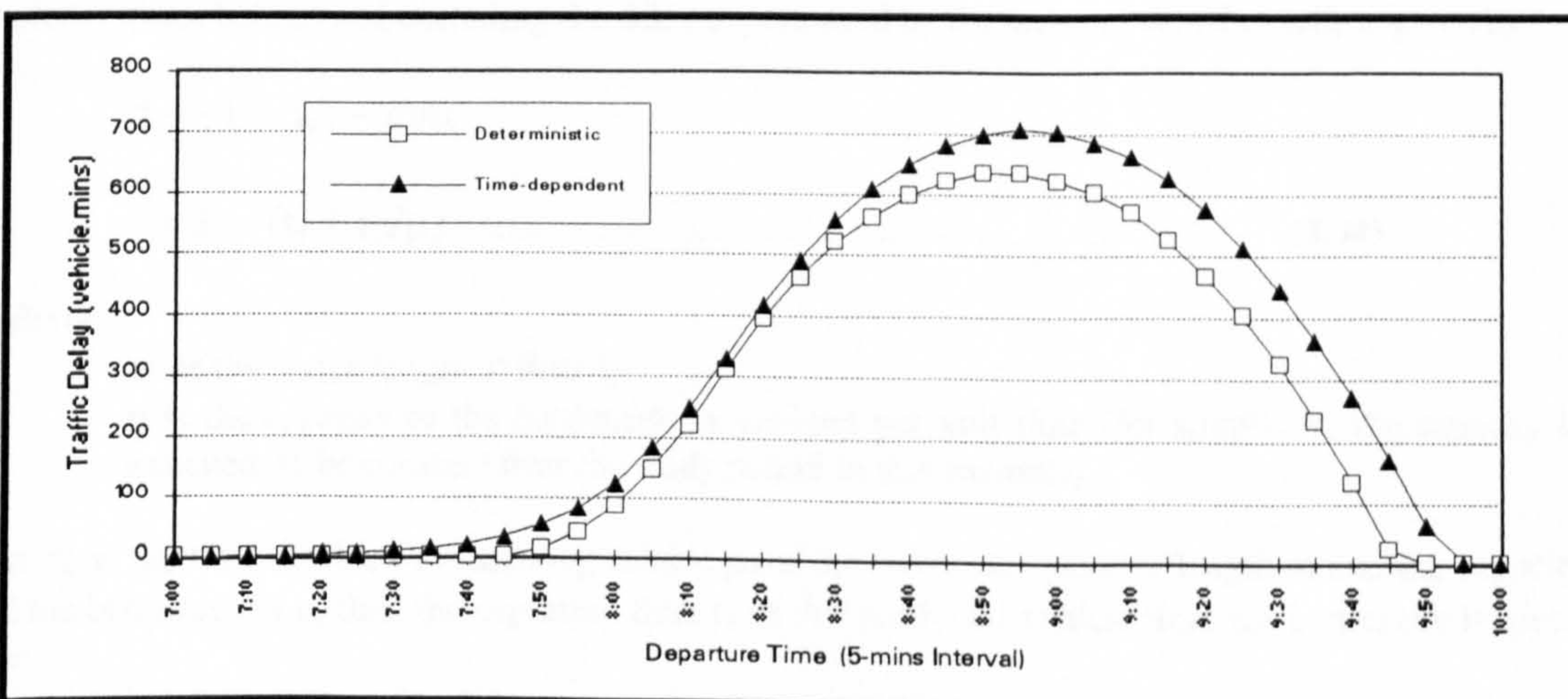


Figure (3.8): The Distribution of the Traffic Delay During the Peak Period.

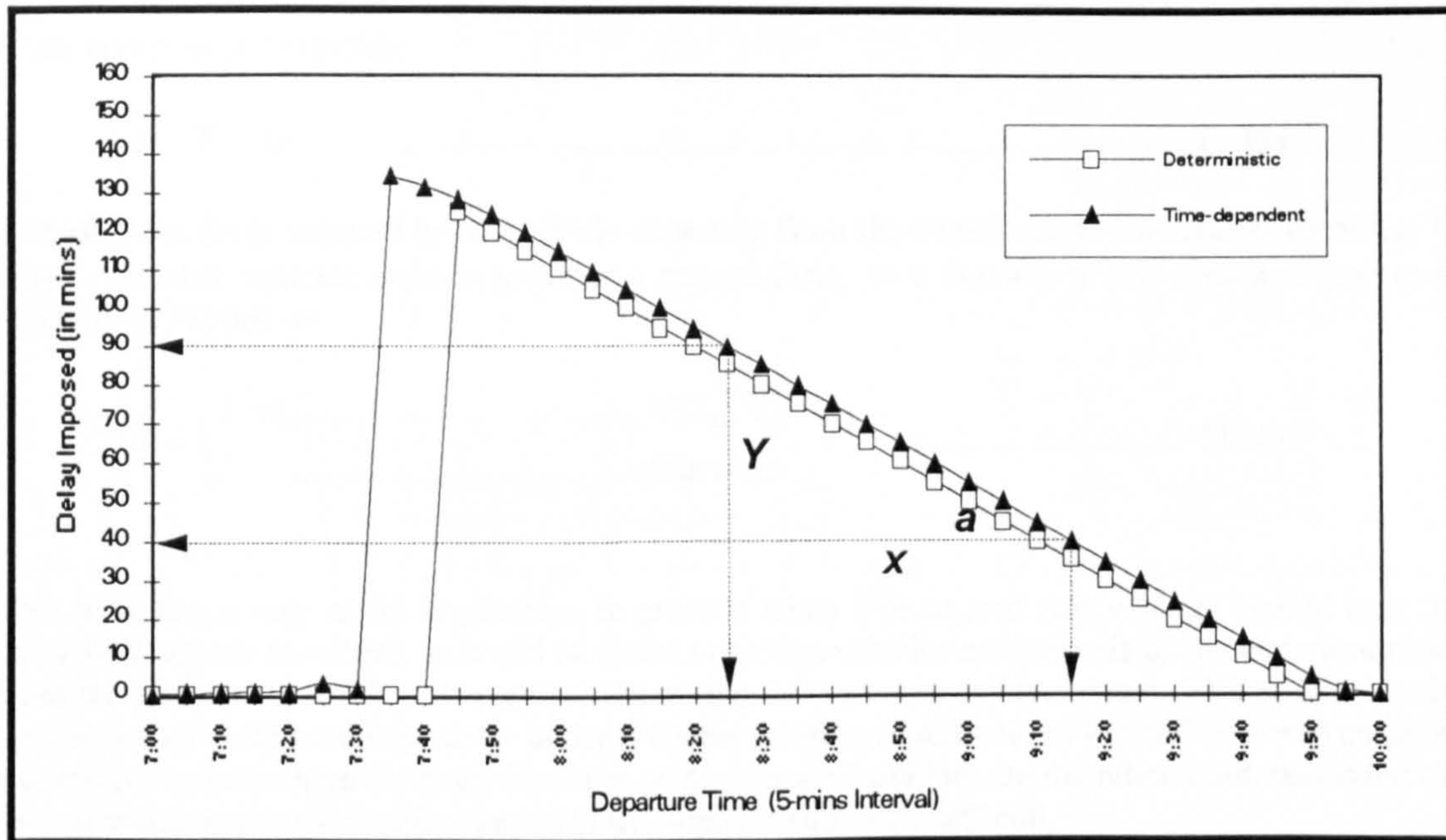


Figure (3.9): The Delay Imposed by a Vehicle Arriving at any Time on Other Vehicles.

$$(\tan(a) = Y / X = \frac{90 - 40}{08:25 - 09:15} = -\frac{50}{50} = -1.00, \text{ thus } a = -45^\circ)$$

it also vanishes after the deterministic one. Furthermore, figures (3.7) and (3.8) demonstrate that ignoring the statistical nature of travel arrivals and services by the deterministic queuing approach, as mentioned earlier, appreciably underestimates the queue length and delay in this simple example.

Figure (3.9) above depicts the delay imposed by the arriving vehicle at any time t_a on other vehicles behind. The relation given by the two solutions in this figure, is quite similar with different values of the starting and vanishing times of the queue (i.e. T_0 and T). This relation could be represented in a general form by a minus 45-degree linear function intersecting the time axis at time T . Therefore, the delay imposed by any vehicle arriving at time t_a (between T_0 and T) on other vehicles behind is given by $(T - t_a)$ as represented by equation (3.33).

The delay d' imposed by the vehicle arriving at t_a on others and given by equation (3.33), includes the delay experienced by the arriving vehicle itself. This is because the vehicle arriving at time t_a is expected to join the tail of the queue at that time. Thus, the delay d imposed by the arriving vehicle on other vehicles behind and excluding the delay experienced by the arriving vehicle itself, is given by:

$$\begin{aligned} d &= (T - t_a) - m/\mu \\ &= T - (t_a + m/\mu) \end{aligned} \quad (3.34)$$

Where:

m is the queue length at time t_a .

μ is the capacity of the bottleneck in vehicles per unit time (for simplicity, the capacity is assumed to be constant over the study period in this example).

As t_a is the time at which the arriving vehicle joins the tail of the queue of length m and the capacity of the bottleneck is μ , then the departure time t_d of this particular vehicle from the bottleneck is given by:

$$t_d = (t_a + m/\mu).$$

Thus, equation (3.34) yields:

$$d = T - t_d \quad \dots\dots\dots (3.35)$$

Therefor, the delay imposed by any vehicle departing from the bottleneck at any time t_d (between T_0 and T) on other vehicles could be given, in a general form, as a function of this departure time as in equation (3.36) below:

$$d = \begin{cases} T - t_d & \dots\dots\dots t_d \in [T_0, T] \\ 0 & \dots\dots\dots \text{otherwise} \end{cases} \quad \dots\dots\dots (3.36)$$

This form has a very useful implication in practice when it is argued that vehicles should bear the delay they impose on others, as it will be easier to charge vehicles on the basis of the departure times from the junction rather than the arrival times (the difference is quite obvious). This is because the former is easy to observe in practice under different queuing conditions, as it could be observed at the stop/give-way line where the road side loops or devices are installed. On the other hand, observing the time at which each vehicle joins the tail of the queue is extremely difficult.

Following Fargier (1983), Small (1992) and Ghali and Smith (1993), this result could be reached using a very simple deterministic approach by assuming that the vehicle arriving at the bottleneck at time t_a will join the tail of a queue of size m vehicles at that particular time, and this queue disappeared at time T . Then the number of vehicles M clearing the bottleneck between t_a and T is:

$$M = \mu(T - t_a) \quad \dots\dots\dots (3.37)$$

Where μ is the capacity of the bottleneck as defined earlier.

This number of vehicles, M , includes the number of vehicles that has been in the queue at time t_a , (i.e. 'm' vehicles). Hence, the total number of vehicles delayed by the arriving vehicle at t_a is $(M-m)$, and every vehicle is delayed by $(1/\mu)$ as a result of the arriving vehicle. Thus, the delay imposed by the arriving vehicles is given by:

$$d = \frac{1}{\mu}(M - m) \quad \dots\dots\dots (3.38)$$

Substituting in equation (3.38) by equation (3.37) yields to:

$$\begin{aligned} &= \frac{1}{\mu}[\mu(T - t_a) - m] \\ &= T - (t_a + \frac{m}{\mu}) \\ &= T - t_d \quad \dots\dots\dots (3.39) \end{aligned}$$

Where t_d is the vehicle departure time from the bottleneck for the vehicle arriving (i.e. joining the tail of the queue) at time t_a .

The result given by equation (3.39) is similar to the one given by equations (3.35) and (3.36) above.

3.7 Summary.

Queuing at road junctions is a time-dependent phenomenon. During peak periods traffic flows are frequently very close to the capacities of junctions. In such circumstances neither steady-state solutions nor deterministic solutions could provide a realistic basis for estimating queue lengths and delays. Therefore, two methods are described for the prediction of queue lengths and vehicular delays at road junctions subject to time-varying traffic demand and capacity. The properties of individual traffic streams are given firstly in terms of the probability distribution of queue lengths, and secondly by approximation formulae which allow queues and delays to be predicted directly. The second approach is used to derive a mathematical formula for the *TDRP* function. The formula derived from this approach is demonstrated and compared with the one obtained from the deterministic queuing approach using a numerical example. It is also demonstrated that the same formula can be reached using a very simple deterministic approach employed by Fargier (1983), Small (1992) and Ghali and Smith (1993).

The conclusion emerges from this chapter is that the *TDRP* function has a general fixed form given by a linear relation (i.e. the wedge-shape) and ignoring the statistical nature of the travel arrivals and services does not affect that shape, but it leads to underestimate the toll values obtained from this function.

Since *TDRP* is modelled as a function of the vehicle departure time from the bottleneck, it is expected to have a great effect on the departure time decisions for travellers. Therefore, the next chapter reviews the existing models for predicting the temporal distribution of traffic demand during the peak period and examines the stability of the results under *TDRP*.

CHAPTER FOUR

4. A Review of Models for Temporal Distribution of Peak Period Demand for a Bottleneck

4.1 Introduction

4.2 General description of the dynamic models of peak period

4.2.1 Deterministic User Equilibrium (*DUE*) Approach.

4.2.2 Stochastic User Equilibrium (*SUE*) Approach.

4.2.3 System Optimal (*SO*) Approach.

4.3 The day-to-day evolution of the departure-time decisions

4.4 The stability of DUE under Time-Dependent Road Pricing (*TDRP*)

4.4.1 Modifying *TDRP* to include the schedule delay changes to other vehicles.

4.4.2 Modifying *TDRP* using the day-to-day adjustment process.

4.5 Summary

4.1 Introduction

Most travellers during the (morning) peak period are commuters who usually have predetermined times they wish to arrive at their destinations. Let Destination Target Time (*DTT*) be defined as the time a traveller wishes to arrive at this destination. This *DTT* may have an associated tolerable band which varies for different travellers; for example, a traveller's *DTT* may be 8:00 a.m., but he may be content with, or indifferent to, arriving at this destination at any time between 7:45 a.m. and 8:15 a.m.

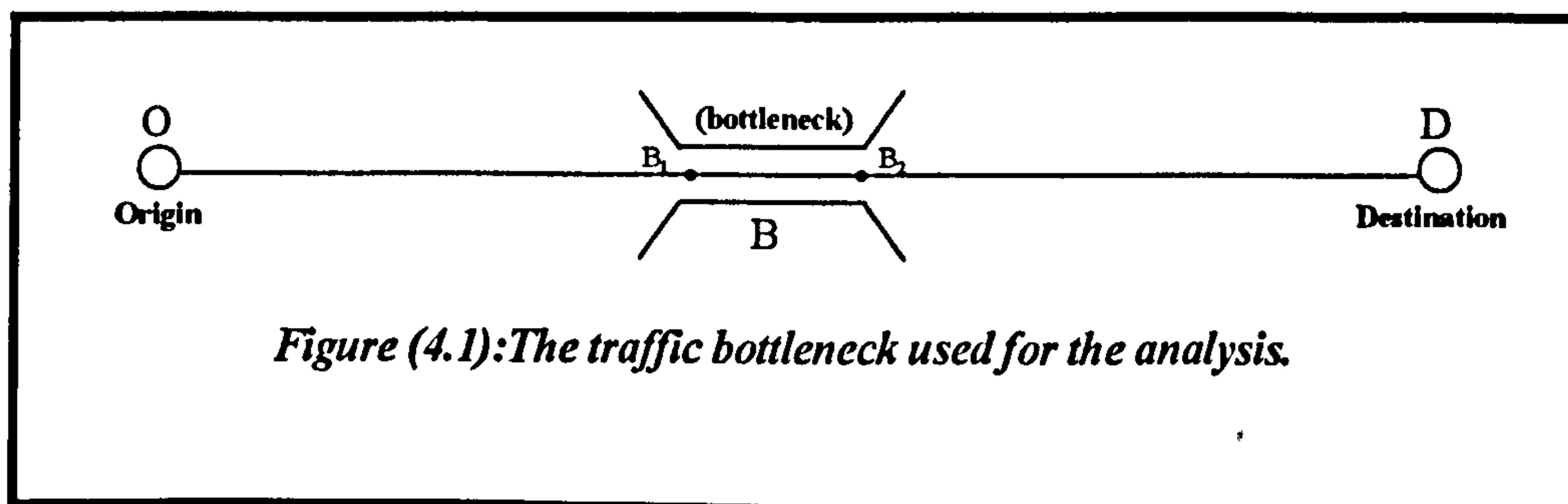
Demand for travel in a system is a function of the related cost (utility) of travel; the impedance related to travel in the system (i.e. congestion cost or queuing delay cost) constitutes a major portion of this cost of travel. The need for a traveller to arrive at his destination at a particular time (*DTT*) introduces an additional cost which varies with the time the traveller starts and completes his journey. This cost is called *schedule delay cost* (and it is equal to the difference between actual and desired arrival time). This added cost due to the *DTT* affects a traveller's choice of departure time from home, route, and mode of travel; but most significantly it affects his departure time. The temporal distribution of traffic demand, observed on the road network, during the peak period is mainly the result of the interactions of all commuters' selected departure times during this period. The spread of the distribution is determined by how rigid the *DTT* is and how the commuters perceive the costs for arriving at their destinations earlier or later than the *DTT* and the costs for excess travel time in the system.

In early studies of peak-period traffic congestion in which Pigou (1932) played a pioneer role, no consideration was generally given to the departure time decisions of road users. This oversight was remedied by Vickrey (1969) who derived the departure rate along a single route subject to queuing congestion as the outcome of individual cost maximisation. The departure time problem was rediscovered by Henderson (1974, 1977 and 1981), Hurdle (1981), and others.

This chapter, however, reviews the existing models for predicting the temporal distribution of traffic demand during the peak period, i.e. which portion of a total *OD* demand (for any *OD* pair) departs from the origin at each time slice. For more details about the review of these models see Alfa (1986). The purpose of this revision is to define an approach for examining the stability (and the quality) of travelling under *TDRP*, then using this approach in the next chapters to evaluate the different effects and the efficiency gains from applying *TDRP* (using the simulation technique) for a simple bottleneck as well as a typical traffic network.

4.2 General description of the dynamic models of peak period

This section outlines the general basic formulation of the dynamic models.



Most of the existing models considered mainly a simple network type which consists of one origin-destination pair connected by one route with a bottleneck in between (see Figure (4.1) above).

Consider the system depicted in Figure (4.1) where Q (travellers) go everyday from point O to point D . This can represent a home to work trip with the following three segments:–

- OB_1 is never congested. The travel time from O to B_1 is a constant t_1 .
- B_2D has the same property as OB_1 . The travel time from B_2 to D is t_2 .
- B_1B_2 is a bottleneck of fixed capacity μ . The travel time from B_1 to B_2 is a constant t_0 ; but at point B_1 a queue begins to develop whenever the arrival rate at B_1 becomes greater than μ . Denote the waiting time for a driver at B_1 at time t as $w(t)$.

Thus, the total travel time, $tt(t)$, from O to D for a driver arriving at time t at B_1 is:

$$tt(t) = t_1 + w(t) + t_0 + t_2 = t_c + w(t) \quad (4.1)$$

Where

$$t_c = t_1 + t_0 + t_2 \quad (4.2)$$

i.e. t_c is the sum of the line haul times on the links along the route between a particular OD pair and it is assumed to be constant. Alfa (1986) argued that in reality t_c may vary depending on traffic volumes along the links but this variation during the peak period is negligible, and the main source of delay to travellers is usually at the bottleneck, e.g. intersection, where there is limited capacity.

It should be noted that an arrival at B_1 at time t and a departure from O at time $t-t_1$ are equivalent. Also it is assumed that the travel time from O to B_1 is constant whether there is a queue or not. In other words, the queue length is assumed to be very short relative to the distance OB_1 , as if cars in the queue were stacked up at point B_1 .

Now consider a traveller who departed from O at time t . If $w(t)$ is the delay to this traveller, and $t_a(t)$ is his arrival time at D , then

$$t_a(t) = t + t_c + w(t) \quad (4.3)$$

The delay $w(t)$ is a function of μ , the bottleneck capacity, and the arrival rates at the bottleneck from the time of arrival of the first vehicle up to t_1+t .

Suppose a traveller who wishes to arrive at this destination, D , at time t_D (i.e. $t_D = DTT$) departed from O at time t and arrived at D at $t_a(t)$. If $t_a(t) < t_D$, this traveller will be early by $t_D - t_a(t)$ time units; and if $t_a(t) > t_D$, he will be late by $t_a(t) - t_D$ time units. In either case traveller attaches some costs, and the relative weighting of these costs depends on how he perceives lateness and earliness penalties. Let $C_e(t_D - t_a(t))$ be the cost for earliness and $C_l(t_a(t) - t_D)$ be the cost for lateness. In addition to these costs traveller attaches some cost $C_w(w(t))$, to the delay $w(t)$. There is also an "inevitable" cost to all travellers due to t_c , the minimum travel time (or free-flow travel time). This "inevitable" cost will not be considered in this case; as it does not affect the general structure of the problem of a simple system consisting of one OD pair and only one single route (but for traffic network applications it should be considered). Let the total perceived cost to this traveller who started his journey at time t be $C(t)$. Then

$$C(t) = C_w(w(t)) + C_e(t_D - t_a(t))^+ + C_l(t_a(t) - t_D)^+, \quad (4.4)$$

Where $(e)^+ = \max(0, e)$.

Equation (4.4) gives the general form of the cost structure of how a traveller perceives travel cost if he has a DTT. There have been minor modifications to this cost structure by different researchers but the basic principles are all the same. For example, de Palma et al (1983) converted the cost to a utility measure; Vickrey (1969) considered value of time instead of costs of time and maximised a traveller's value of time. In other cases, cost as in equation (4.4) was considered, but one of the three components was ignored.

It is believed that each traveller wishes to select his departure time from O in such a way that his total cost $C(t)$ is minimised. Most models were set up on this basis. The major differences in the models are in the assumptions of how the traveller perceives this minimum cost and how it affects his decisions. These differences are in the principles adopted. Deterministic User Equilibrium (*DUE*), Stochastic User Equilibrium (*SUE*) and System Optimum (*SO*) are the three main principles that have been used to study the problem of the temporal distribution of peak traffic demand. These principles are discussed in some detail each in turn.

4.2.1 Deterministic User Equilibrium approach (*DUE*)

The principle of the deterministic user equilibrium (*DUE*) as used for the choice of departure time is similar to that of route choice as proposed by Wardrop (1952). The *DUE* for choice of departure time is stated as follows: "at equilibrium, no traveller can reduce his total cost by changing his departure time."

Suppose the first group of traveller depart from O at time T_1 and the last group of travellers at T_2 ($T_1 < T_2 < \infty$). If the departure rate is $D(t)$, $T_1 \leq t \leq T_2$, and Q is the total number of travellers going from O to D, then the *DUE* principle for the selection of departure time from O states that at *DUE*:

$$C(t) = K \quad \text{if } D(t) > 0, t \in [T_1, T_2] \quad \dots\dots\dots (4.5a)$$

$$C(t) > K \quad \text{if } D(t) = 0, t \notin [T_1, T_2], 0 < K < \infty \quad \dots\dots\dots (4.5b)$$

Where $D(t)$ is a continuous function between T_1 and T_2 . The *DUE* problem for choice of departure time is thus to find $D(t) > 0$, which satisfies equations (4.5a) and (4.5b) for which Q is given by:

$$Q = \int_{T_1}^{T_2} D(t) dt \quad \dots\dots\dots (4.6)$$

All the travellers depart from O between T_1 and T_2 at rates $D(t)$, and they experience total average costs $C(t)$; no traveller can reduce his cost by changing his departure time, except the last traveller as pointed out by Hendrickson and Kocur (1981).

The *DUE* approach was used by Hendrickson and Kocur (1981), Fargier (1983), Hurdle (1974), Henderson (1974, 1977) and Vickrey (1969). In order to solve the *DUE* problem for choice of departure time, they used the deterministic queuing model to estimate the travel delay, $w(t)$. The deterministic queuing model used is based on fluid approximation and is described in Newell (1971). It was shown, for this problem of departure time selection, that the departure rate from O when the first traveller departs is at least equal to the capacity of the bottleneck, i.e. $D(T_1) \geq \mu$. This is because, according to the fluid approximation technique, until arrival rate at the bottleneck exceeds the bottleneck capacity, at least once, no queuing is developed. Hence arrival rate at the bottleneck, at the arrival of the first group, is at least up to the capacity of the bottleneck. The delay is then estimated as:

$$w(t) = \frac{1}{\mu} \int_{T_1}^t (d(x) - \mu) dx \quad \dots\dots\dots (4.7)$$

Henderson (1977) considered a non-queuing situation and did not use a queuing model. He used the fundamental speed-flow relationship to obtain travel time. Mahmassani and Herman (1983) also used the speed-flow relationship to evaluate travel time to their work that expanded the model of Hendrickson and Kocur (1981) to include route selection in an idealised network.

All the *DUE* approaches discussed above used linear functional relationships for each of the cost functions; $C_w(w(t))$, $C_a(t_D - t_a(t))^+$ and $C_l(t_a(t) - t_D)^+$. Henderson (1981), however, used a general functional relationship for each of the costs in his study of staggered work hours. Hendrickson and Kocur (1981) and Fargier (1983) extended their works to study the effects of staggered work hours.

In this study, the approach of Hendrickson and Kocur (1981) is reviewed in some detail after it has been adapted to include work start time flexibility.

Hendrickson and Kocur (1981) considered a set of travellers commuting to a single destination with all congestion occurring at a single bottleneck facility (as in Figure (4.1)). They assumed that all trips occur in private automobiles and the bottleneck facility can serve a constant number of travellers per unit time. For simplicity, the average automobile occupancy is assumed to be constant throughout the discussion, so the facility service rate, μ , can be expressed in terms of the number of travellers served rather than vehicles per unit of time. Instead of assuming that all travellers begin their work at the same time, B , the approach is adapted to include work start time flexibility, i.e. by considering the time interval $[B-\Delta, B+\Delta]$, where $\Delta \geq 0$, is the desired time period for arriving at the destination D , and B denotes the centre of the period and Δ is a measure of work start time flexibility.

To characterise the demand for travel, it is assumed that the total number of travellers is a constant, Q , and that each traveller unilaterally attempts to minimise his own travel cost (or utility) $UC(t)$, expressed as linear combination of waiting (queuing) time at the bottleneck, $w(t)$, earliness in arrival at work, $s(t)$, lateness in arrival at work, $p(t)$, and any tolls, $f(t)$, all of which are function of user's arrival time at the bottleneck:

$$UC(t) = a_0 + a_1 w(t) + a_2 s(t) + a_3 p(t) + f(t) \quad (4.8)$$

Where a_0, a_1, a_2, a_3 are constant parameters, and $s(t)$ and $p(t)$ are given by:

$$s(t) = \begin{cases} (B - \Delta) - t_a(t) & \text{.....} t_a(t) < (B - \Delta) \\ 0 & \text{..... otherwise} \end{cases} \quad (4.9)$$

$$p(t) = \begin{cases} t_a(t) - (B + \Delta) & \text{.....} t_a(t) > (B + \Delta) \\ 0 & \text{..... otherwise} \end{cases} \quad (4.10)$$

Where $t_a(t)$ is the time at which the traveller who arrives at the bottleneck at time t reaches work (i.e. his destination) is given by:

$$t_a(t) = t + w(t) + t_c \quad (4.11)$$

No matter what departure time is chosen, each traveller incurs a fixed user cost associated with vehicle ownership, vehicle operating costs, and the uncongested travel to the bottleneck facility. These fixed costs are, in fact, included in the parameter a_0 .

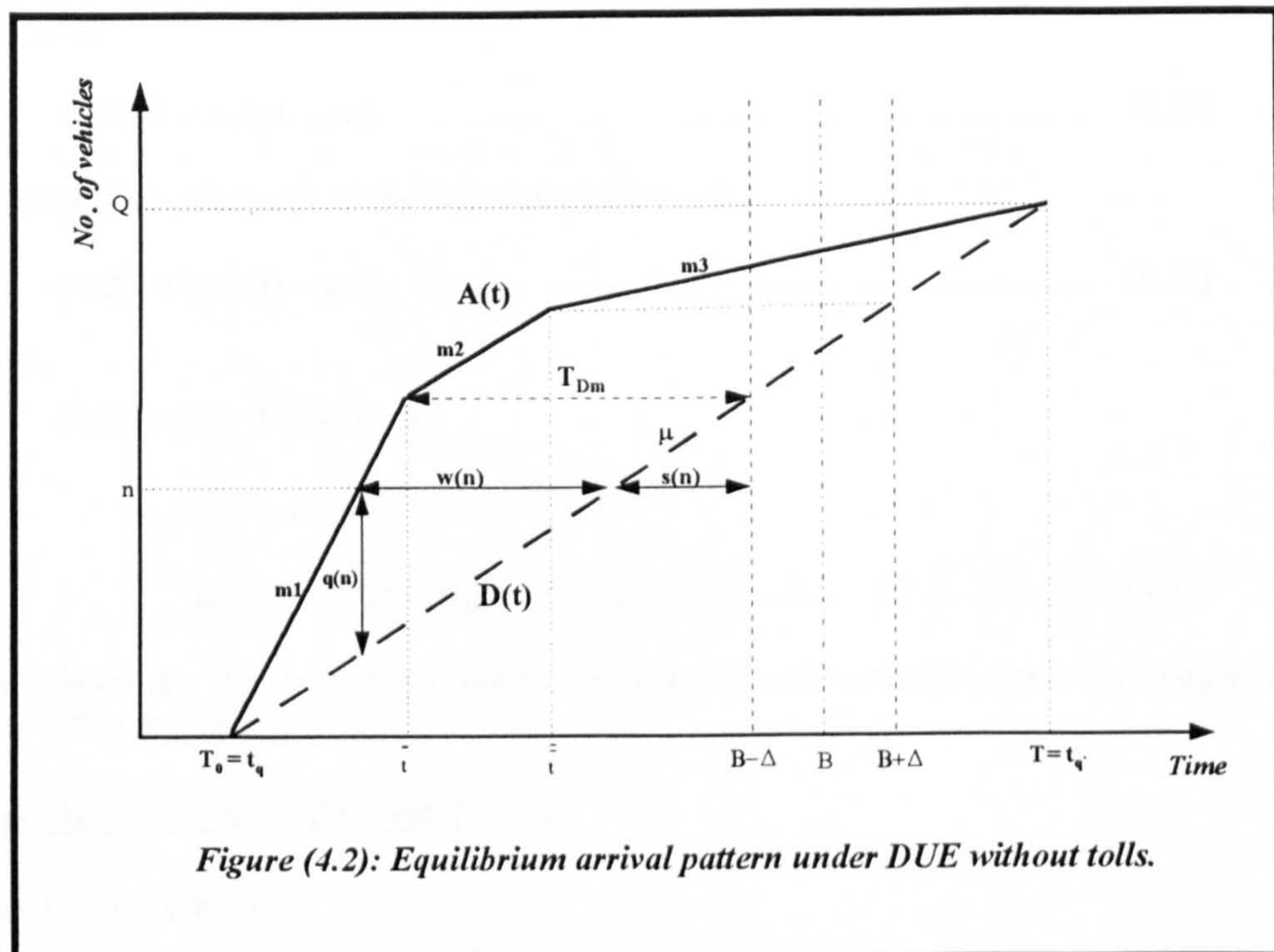
Several function are used in the discussion. The function $A(t)$ represents the cumulative number of travellers who arrived at the bottleneck facility by time t . Since travel time to the bottleneck facility is assumed to be constant and known to commuters, the function $A(t)$ is completely determined by departure time decisions of the travellers. While an actual arrival function such as $A(t)$ would be a step function with a step corresponding to each traveller's arrival, the fluid approximation is assumed for simplicity and it is assumed that $A(t)$ is then continuous and differentiable. The derivation of $A(t)$ is then the rate at which travellers arrive at the bottleneck facility as a function of time, denoted $m(t)$. The cumulative number of departures from the queue at the bottleneck is denoted $D(t)$; its derivative is the service rate μ .

Now suppose that late arrivals at work is permitted, although a penalty is incurred so that $a_3 > 0$; it is also assumed that no toll is charged (i.e. $f(t) = 0$) and a minute spent at the work-place prior to $(B-\Delta)$ is judged to be less onerous than a minute spent waiting in queue at the bottleneck (so $a_1 > a_2$). (If a minute spent at the work-place is more onerous than a minute in queues, no stable equilibrium develops, see Equation (4.18)).

The necessary condition for user equilibrium is that user costs are equal for all travellers:

$$UC(t) = a_0 + a_1 w(t) + a_2 s(t) + a_3 p(t) = C_{eq}, \quad \text{for all } t \dots\dots\dots (4.12)$$

Figure (4.2) illustrates the equilibrium pattern of arrivals in this case. The cumulative arrival time function at the bottleneck is represented by fluid approximation $A(t)$. For any particular arrival such as n in Figure (4.2), the vertical distance $q(n)$ represents the queue length encountered, the horizontal distance $w(n)$ represents the waiting time in the queue, and the horizontal distance $s(n)$ indicates the schedule delay incurred. A description of this type of deterministic queuing system appears in Newell (1971).



The cumulative arrival time function $A(t)$ is determined by the requirement that user cost be constant for all travellers. Therefore, during the early arrival period the queue is increasing to balance the reduction in early schedule delay, then it remains constant during the on-time arrival period as the schedule delay during this period is zero, then it starts falling down to balance the increase in late schedule delay during the late arrival period until it reaches zero at T (or t_q). The waiting time in the queue at any time t is :

$$w(t) = \frac{1}{\mu} \int_{T_0}^t (d(y) - \mu) dy \quad \dots\dots\dots (4.13)$$

Where $w(t)$ is the queuing time as a function of the arrival time t , and T_0 (or t_q) is the time the queue first forms. The early schedule delay at time t is then $s(t) = B - \Delta - t - w(t)$ and the late schedule delay at time t is $p(t) = t + w(t) - (B + \Delta)$. These are implying a user cost of:

$$UC(t) = a_0 + a_1 w(t) + a_2 (B - \Delta - t - w(t)) + a_3 (t + w(t) - B - \Delta) \quad (4.14)$$

The different slopes of $A(t)$ can be determined as follows:-

- For early arrival period $[T_0, \tilde{t}]$

The user cost is given by:

$$\begin{aligned} UC(t) &= a_0 + a_1 w(t) + a_2 (B - \Delta - t - w(t)) \\ &= a_0 + (a_1 - a_2) w(t) + a_2 (B - \Delta) - a_2 t \end{aligned} \quad (4.15)$$

For the user cost to be constant, its derivative with respect to time must be zero, thus

$$\partial UC(t)/\partial t = 0 = (a_1 - a_2) (\partial w(t)/\partial t) - a_2$$

which gives

$$\partial w(t)/\partial t = a_2/(a_1 - a_2) \quad (4.16)$$

Substituting in Equation (4.16) by Equation (4.13) yields:

$$\partial w(t)/\partial t = (m(t) - \mu)/\mu \quad (4.17)$$

Then

$$(m(t) - \mu)/\mu = a_2/(a_1 - a_2)$$

$$m(t) = \mu a_1/(a_1 - a_2)$$

$$m1 = \mu a_1/(a_1 - a_2) \quad (4.18)$$

which implies that the slope of $A(t)$ during the period of early arrival is constant and equal to $m1$ as shown in Figure (4.2).

- For the on-time arrival period $[\tilde{t}, \tilde{t}]$

The user cost is given by:

$$UC(t) = a_0 + a_1 w(t) \quad (4.19)$$

As before the derivative of user cost with respect to time must be zero:

$$\partial UC(t)/\partial t = 0 = a_1 \partial w(t)/\partial t$$

Then

$$\partial w(t)/\partial t = 0$$

Substituting by Equation (4.13) yields:

$$m(t) = \mu$$

or

$$m2 = \mu \quad (4.20)$$

which implies that the arrival rate at the bottleneck during the period of on-time arrival $m2$ is equal to the service rate μ .

- For late arrival period $[\tilde{t}, T]$

For a late arrival at work the user cost becomes:

$$\begin{aligned}
 UC(t) &= a_0 + a_1 w(t) + a_3 (t + w(t) - B - \Delta) \\
 &= a_0 + (a_1 + a_3) w(t) + a_3 t - a_3 (B + \Delta) \dots\dots\dots (4.21)
 \end{aligned}$$

The queuing time $r(t)$ is given by Equation (4.12) as before and the derivation of user cost with respect to time must equal to zero:

$$\partial UC(t)/\partial t = 0 = (a_1 + a_3) \{(m(t) - \mu)/\mu\} + a_3 \dots\dots\dots (4.22)$$

The slope of the arrival function for late arrivals is then:

$$\begin{aligned}
 m(t) &= a_1 \mu / (a_1 + a_3) \\
 \text{or} \\
 m_3 &= a_1 \mu / (a_1 + a_3) \dots\dots\dots (4.23)
 \end{aligned}$$

The initial arrival time can be obtained by solving Equation (4.14) with the constraint that, at equilibrium, costs of the first and last arrivals must be equal. This is given by Equation (4.24) below:

$$t_q = B + (\Delta(a_3 - a_2) - \frac{Q}{\mu} a_3)/(a_2 + a_3) \dots\dots\dots (4.24)$$

The queue vanishes at t_q which is given by:

$$t_q = B + (\Delta(a_3 - a_2) + \frac{Q}{\mu} a_2)/(a_2 + a_3) \dots\dots\dots (4.25)$$

The pattern shown in Figure (4.2) represents a stable user equilibrium since no traveller can unilaterally reduce his cost. The user cost, at equilibrium, in this case is

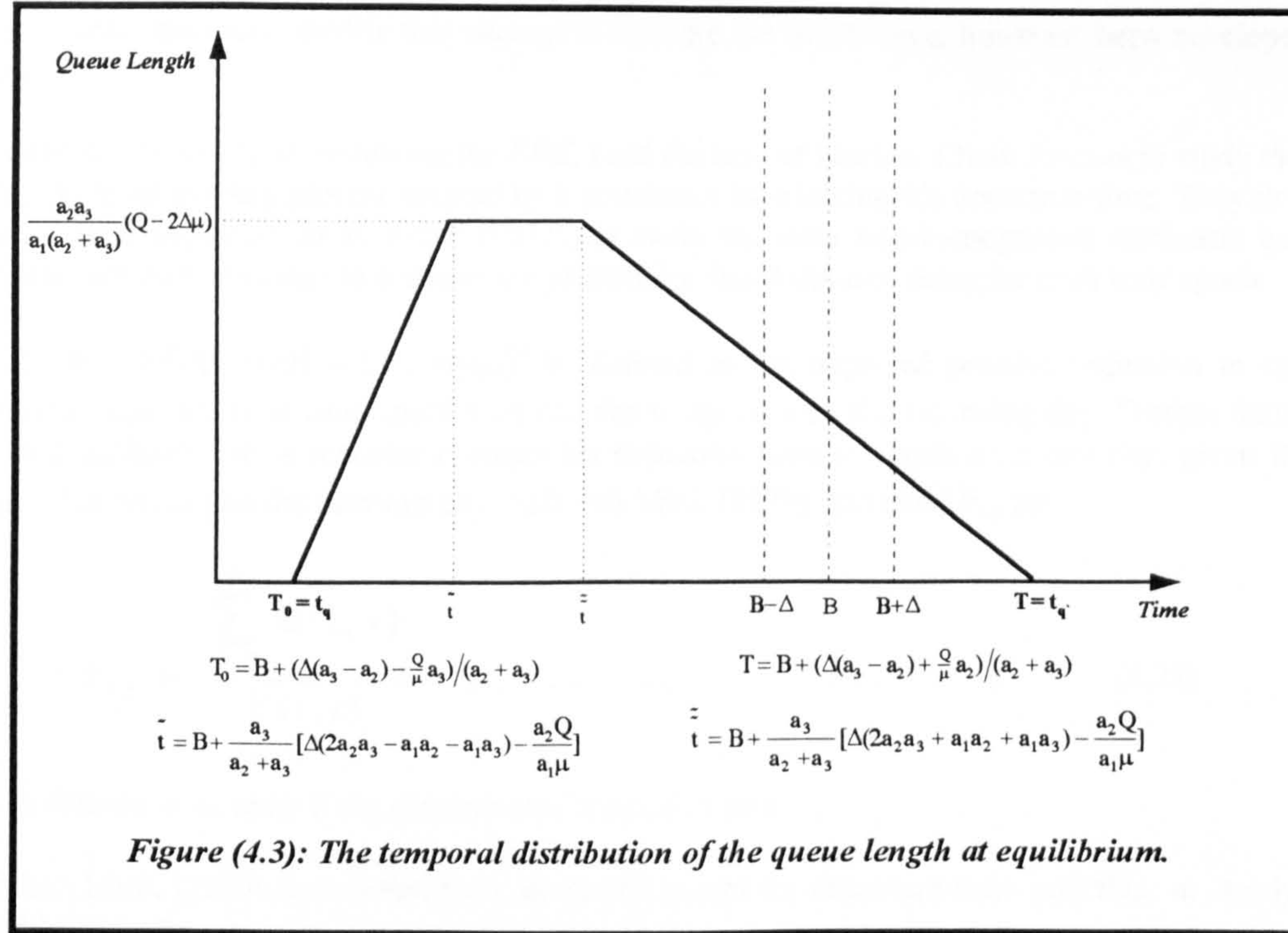
$$a_0 + a_2 a_3 \left(\frac{Q}{\mu} - 2\Delta \right) / (a_2 + a_3),$$

the maximum queue length (in vehicles) is

$$a_2 a_3 (Q - 2\mu\Delta) / a_1 (a_2 + a_3)$$

and the average waiting time is one-half this amount. The distribution of the queue length over time is illustrated in Figure (4.3). The commuters between $\bar{t}$ and $\bar{t}$ are incurring maximum queuing delay, T_{Dm} , as shown in Figure (4.2). These particular commuters must exit the queue just at their desired time (i.e. $B - \Delta$, $B + \Delta$ respectively), since otherwise they could reduce both queuing delay and schedule delay by shifting their schedules. Commuters entering before $\bar{t}$ exist earlier than they desire, and those entering after $\bar{t}$ exit later than they desire.

To summarise, *DUE* models require and assume that users have perfect knowledge of the system costs. In this case, it is assumed that the commuters have perfect knowledge of delay in the system at each time during the peak period. This assumption is not necessarily valid due to traveller's inaccurate and distorted perception of travel time throughout the peak period and also due to the stochastic variation in travel time over many different days for the same time of day. In addition, the use of deterministic queuing model for estimating delay ignores the statistical fluctuations in the arrival of travellers and service at the bottleneck, therefore predicting zero delays during under-saturation periods (Alfa (1986)). Whereas the non-queuing model used by Henderson (1977) for estimating the travel time is free of the latter problem, its weakness is that when estimating travel time it considers only the flow at that particular time; it does not consider, in addition, the flows before that time.



4.2.2 Stochastic User Equilibrium Approach (SUE)

Stochastic User Equilibrium (SUE) models recognize the inaccuracies and distortions in travellers' perceptions of travel times or delays, $w(t)$, and hence cost, $C(t)$. Similar to the definition of SUE for traffic assignment stated by Daganzo and Sheffi (1977), the SUE for departure time selection can be generally stated as follows: "in an SUE departure time selection no traveller believes he can improve his total cost by unilaterally changing his departure time." Thus $w(t)$ is assumed to be a random variable. It assumes values between a minimum $\underline{w}(t)$ and a maximum $\overline{w}(t)$. $C(t)$ is thus a stochastic variable controlled by the random variable $w(t)$.

Let $C(t)$ be restated as $C(t, w(t))$, where

$$C(t, w(t)) = C_w(w(t)) + C_e(t_D - t_a(t))^+ + C_l(t_a(t) - t_D)^+ \quad (4.26)$$

Where $t_a(t)$ and t_D are the actual and the desired arrival times at destination D respectively.

This definition is to emphasise the dependence of $C(t)$ or $C(t, w(t))$ on $w(t)$. Let $P(t)$ be defined as the probability that a traveller chooses to depart from O at time t . The SUE model is stated as:

$$P(t) = \Pr\{C(t, w(t)) = \min(C(s, w(s))), \quad \forall s \in [T_0, T], \quad T_0 \leq t \leq T\},$$

$$\underline{w}(s) \leq w(s) \leq \overline{w}(s), \quad \underline{w}(t) \leq w(t) \leq \overline{w}(t),$$

$$T_0 \leq t \leq T; \quad T_0 \leq s \leq T. \quad (4.27)$$

For very large Q , the relationship $d(t) \cong P(t) \times Q$ can be assumed.

The interdependence of $w(t)$ on the cumulative distribution of $d(t)$ makes solving the SUE problem, stated above, cumbersome not to mention establishing existence and uniqueness of a solution. One of the problems is the difficulty in solving time non-homogenous queuing models, and the other is the evaluation of Equation (4.27). The SUE model for departure time selection has thus not been solved as

such. Several stochastic models that attempt to emulate the *SUE* have, however, been developed and solved.

Alfa and Minh (1979), in emulating the *SUE*, used the idea of Markov Chain Process to study the day-to-day decision making process adopted by a commuter in selecting his departure time. They used the discrete time approach, as in Minh (1977), to study the time non-homogenous stochastic queuing problem and were thus able to evaluate the probability distribution of delay for each time epoch.

Let $E(t,s) = E(C(t, w(t)) - C(s, w(s)))^+$ be defined as the expected positive reduction in cost for changing departure time from epoch t on one day to epoch s on the following day. Further define $P_{t,s}$ as the probability that a traveller changes his departure time to epoch s on one day, given that he departed at epoch t on the previous day. Alfa and Minh (1979) estimated $P_{t,s}$ as:

$$P_{t,s} = \frac{\sum_{v=T_1}^{T_2} E(t, v)}{E(t, s)} \quad \dots\dots\dots (4.28)$$

$P_{t,s}$ is defined to be unity if the denominator is equal to zero.

Alfa and Minh (1979) then stated their stochastic model for departure time selection, at steady-state (equilibrium), as

$$P(s) = \sum_{t=T_1}^{T_2} P(t) \cdot P_{t,s}, \quad T_1 \leq s \leq T_2 \quad \dots\dots\dots (4.29)$$

The model, as set up, is dynamic and could be used for evaluating the day-to-day decisions, but the ultimate interest was in the decisions at equilibrium. In the dynamic set up of their model the expected reduction in cost $E(t,s)$ varies from day to day, because $w(t)$ did not have the same distribution each day at the transient stage. A suffix r was added to define $E_r(t,s)$ as the $E(t,s)$ for the base day r . By approximately adding the "day" suffix to the other parameters in equations (4.28) and (4.29) the dynamic model was set up.

De Palma et al. (1983), on the other hand, in emulating the *SUE*, used the logit model to study the departure time selection by a traveller. They used the deterministic queuing model to estimate delay. They also translated the cost function to a utility function. This model is discussed in more detail below.

Stochastic User Equilibrium Model

The basic model consists of the following elements: a waiting time model, a queue length equation, an arrival rate model and a utility function. Each element is considered in turn below:—

(a) The waiting time model

Let $L(t)$ be the number of cars in the queue at time t . The waiting time for an arrival at B_1 at time t is derived from a model of a deterministic queue as follows:

$$w(t) = L(t)/\mu \quad \dots\dots\dots (4.30)$$

With no congestion $L(t)=0$, and $w(t)=0$. This assumption can be regarded as an approximation of a stochastic queue for $Q \gg 1$. It is the same as the "fluid approximation" developed by Newell (1971).

He assumed that for large queues the discrete and stochastic arrival and departure processes can be approximated by non-stochastic and continuous variables.

(b) The queue length equation

Let $L(t)$ satisfy the following differential equation based on the flow conservation relationship at the bottleneck:

$$dL(t) / dt = \dot{L}(t) = \text{arrival flow} - \text{outflow} \quad (4.31)$$

Let $r(t)$ be the arrival rate at point B_1 (see Figure (4.1)). Without congestion, $L(t)=0$ and the outflow is equal to the arrival rate. Thus

$$\dot{L}(t) = 0, \text{ for } L(t) = 0 \quad (4.32)$$

With congestion $L(t) > 0$ and the outflow is equal to the capacity. Thus,

$$\dot{L}(t) = r(t) - \mu, \text{ for } L(t) > 0 \quad (4.33)$$

De Palma et al. (1983) demonstrated that there can exist at most one congestion period. Denote by t_q and t_q' the beginning and the end times of a congestion period respectively. Then, an equivalent formulation of Equations (4.32) and (4.33) is given by:

$$\dot{L}(t) = 0, \quad \text{for } t \leq t_q \text{ and } t \geq t_q' \quad (4.34)$$

and

$$\dot{L}(t) = r(t) - \mu, \quad \text{for } t_q < t < t_q' \quad (4.35)$$

(c) The departure time choice model

A driver's choice of a departure time from O can be equivalently expressed as a choice of an arrival time at the bottleneck when an arrival time t is the same as a departure time $t-t_1$. Let $U(t)$ be unobservable random utility for an arrival at B_1 at time t and

$$U(t) = V(t) + \eta \varepsilon(t) \quad (4.36)$$

Where $V(t)$ is a deterministic utility component; $\varepsilon(t)$ is a disturbance term (it is a random variable whose mean is usually assumed to be zero. It represents the effects of all factors that caused perceived and measured costs to differ). η is a scale parameter, where $\eta=0$ corresponds to the deterministic choice and $\eta=\infty$ corresponds to a pure random choice ($\eta = \sigma\sqrt{6}/\pi$, where σ is the standard deviation from the mean, see Ben-Akiva and Lerman (1985)). Every Individual selects the time t which maximises his utility. With an additional assumption about the distribution of the disturbances, the probability of unit time interval t being chosen is given by the following continuous logit model (see Ben-Akiva and Watanatada (1981) and Litinas and Ben-Akiva (1979)):-

$$P(t) = \exp[V(t)/\eta] / \int_{T_0}^T \exp[V(u)/\eta] du$$

$$= (1/E) \exp [V(t)/\eta] \quad (4.37)$$

Where T_0 and T are the first and last possible arrival times at B_1 respectively, and E is the logit denominator given by Equation (4.38):

$$E = \int_{T_0}^T \exp[V(u)/\eta] du \quad \dots\dots\dots (4.38)$$

The natural order of alternative departure times calls into question the validity of the logit model structure. Therefore, Small (1982) and Abokowitz (1980) performed a large number of tests that did not reject this assumption. They concluded that the logit model can serve as a reasonable model of departure time choice behaviour.

(d) The arrival rate equation

The Q cars arrive at B₁ between T₀ and T. Thus, the arrival rate at point B₁ at time t for a large Q can be written as:

$$r(t) = Q.P(t) = (Q/E). \exp [V(t)/\eta] \quad \dots\dots\dots (4.39)$$

This expression for an arrival rate can best be interpreted as the steady-state solution of a dynamic system as shown in de Palma and Lefever (1983) and employed by Ben-Akiva et al. (1984).

(e) Specification of the utility function

The two key variable among alternative departure times are the travel time and the schedule delay.

Disutility of travel time; the theoretical development assumes for simplicity a linear disutility as shown in Figure (4.4a). Other non-linear functional forms that might be more realistic can be used in simulation. From Figure (4.4a) the marginal disutility of an additional unit of travel time is $\alpha > 0$. Thus the utility associated with travel time is $-\alpha tt(t)$.

Disutility of schedule delay; let the time interval $[B-\Delta, B+\Delta]$ where $\Delta \geq 0$, be the desired time period for arrival at destination D, where B denotes the centre of the period and Δ is a measure of work start time flexibility. Let $t_a(t)$ be the arrival time at D of a car arriving at B₁ at time t:

$$t_a(t) = t + w(t) + t_0 + t_2 \quad \dots\dots\dots (4.40)$$

Let $\tilde{t}$ and $\tilde{\tilde{t}}$; be the arrival time at B₁ such that

$$t_a(\tilde{t}) = B - \Delta \quad \text{and} \quad t_a(\tilde{\tilde{t}}) = B + \Delta$$

In other words, arrivals at the bottleneck during the interval $[T_0, \tilde{t}]$ are early arrivals, during $]\tilde{t}, \tilde{\tilde{t}}[$ are on-time arrivals and during $[\tilde{\tilde{t}}, T]$ are late arrivals. It is assumed without loss of generality that $T > \tilde{\tilde{t}}$. Substitute in (4.40) to obtain

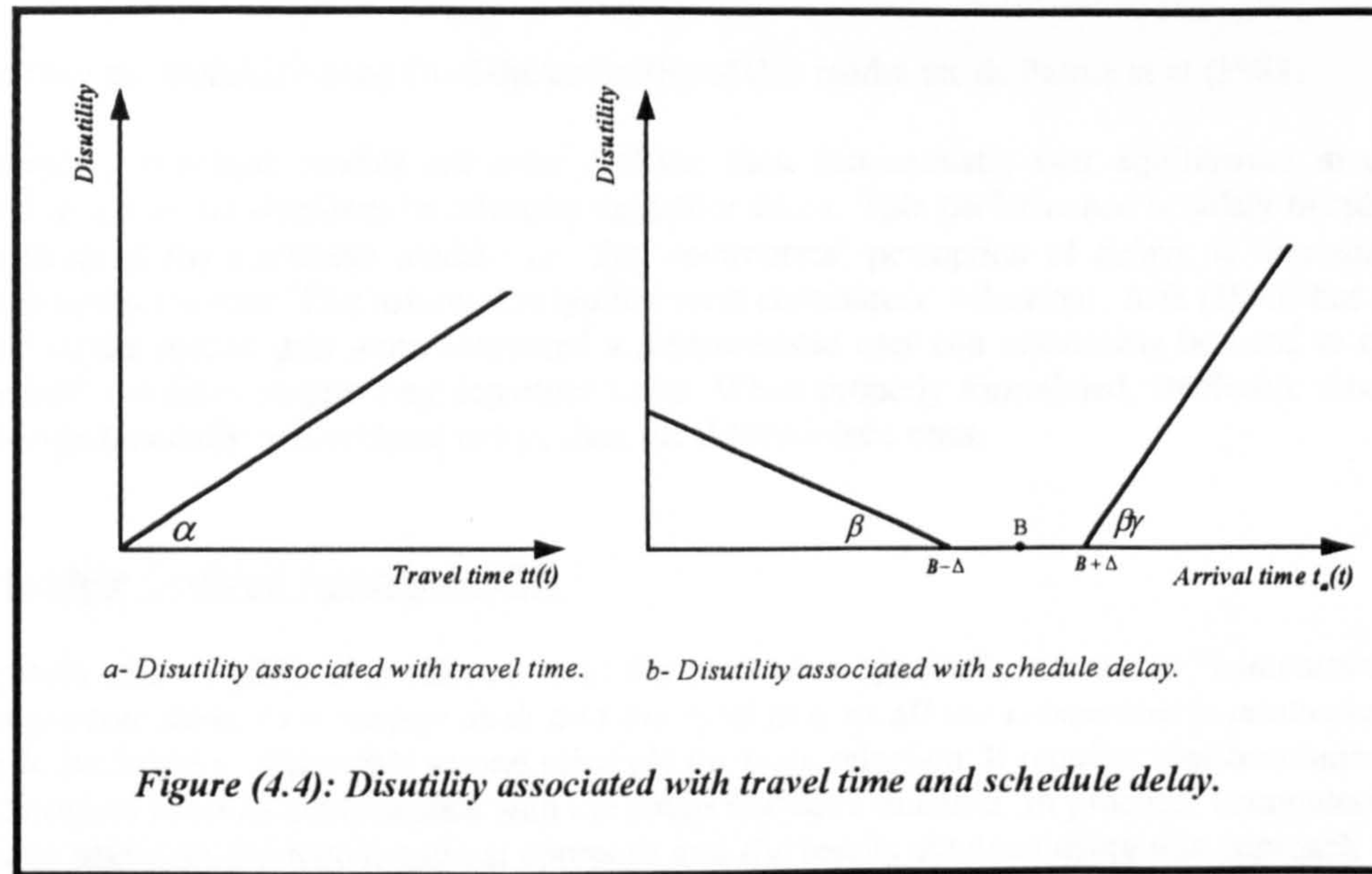
$$\tilde{t} = B - t_0 - t_2 - \Delta - w(\tilde{t}) \quad \dots\dots\dots (4.41)$$

$$\tilde{\tilde{t}} = B - t_0 - t_2 + \Delta - w(\tilde{\tilde{t}}) \quad \dots\dots\dots (4.42)$$

In no congestion case, Equations (4.41) and (4.42) reduce to

$$\tilde{t} = B - t_0 - t_2 - \Delta \quad \dots\dots\dots (4.43)$$

$$\tilde{\tilde{t}} = B - t_0 - t_2 + \Delta \quad \dots\dots\dots (4.44)$$



Assuming that the disutility of schedule delay decreases linearly at a rate $\beta > 0$ for $t \leq \tilde{t}$ and increases linearly at a rate $\beta\gamma$ for $t \geq \tilde{t}$, $\gamma > 0$; see Figure (4.4b). Let $\theta(t)$ be defined as follows:

$$\theta(t) = \begin{cases} 1 & \dots\dots\dots T_0 \leq t \leq \tilde{t} \\ 0 & \dots\dots\dots \tilde{t} < t < \tilde{t} \\ -\gamma & \dots\dots\dots \tilde{t} < t \leq T \end{cases} \dots\dots\dots (4.45)$$

The utility associated with schedule delay $US(t)$ is

$$US(t) = \begin{cases} -\beta(B - \Delta - t_a(t)) \\ = -\beta(B - t_0 - t_2 - \Delta - t - w(t)), \dots\dots\dots \theta(t) = 1 \\ 0 \dots\dots\dots \theta(t) = 0 \\ -\beta\gamma(t_a(t) - (B + \Delta)) \\ = -\beta\gamma(t + w(t) + t_0 + t_2 - B - \Delta) \dots\dots\dots \theta(t) = -\gamma \end{cases} \dots\dots\dots (4.46)$$

and for all values of $\theta(t)$, this utility can be expressed as follows:

$$US(t) = \beta\Delta|\theta| - \beta\theta(B - t_0 - t_2 - t - w(t)) \dots\dots\dots (4.47)$$

Thus, the utility associated with both travel time and schedule delay $UC(t)$ is

$$UC(t) = -\alpha(T_0 + w(t)) + \beta\Delta|\theta| - \beta\theta(B - t_0 - t_2 - t - w(t)) \dots\dots\dots (4.48)$$

The constant quantity $-\alpha T_0$ does not vary across departure times and can be omitted to obtain:

$$V(t) = \beta\theta t + (\beta\theta - \alpha).w(t) + \beta\Delta|\theta| - \beta\theta(B - t_0 - t_2) \dots\dots\dots (4.49)$$

For detailed mathematical analysis of the properties of this model see de Palma et al (1983).

In summary, stochastic models are more realistic than deterministic user equilibrium models for simulating commuter decisions in selecting departure times. This performance is solely based on the assumptions of the stochastic model— i.e. that commuters' perception of delays is inaccurate and distorted to some extent. This assumption typifies most commuters' behaviour. Alfa (1986) has pointed out that as the system gets more congested a deterministic user can reasonably be used to estimate commuters' decisions in selecting departure times. When properly formulated, stochastic models are more computationally cumbersome to use, than the deterministic ones.

4.2.3 System Optimal Approach (SO)

The system optimal (SO), also referred to as the normative approach, states that "*commuters select their departure times in a manner such that the total cost to all the commuters is minimum.*" This principle is similar to Wardrop's second principle for route selection. It requires that commuters select their departure times in conformance with the social objective function. In practice, commuters do not behave as suggested by system optimal approach and the results obtained using this approach will not represent commuters' behaviours. However, it would be worthwhile to examine the relationship between SO and DUE approaches to see if they are as related in the departure time selection as demonstrated in route selection process (see Alfa 1986).

However, Hendrickson and Kocur (1981) expanded their DUE model (discussed earlier) to consider the system optimal approach. This model is adapted to include the flexibility in work start times and discussed below.

Using the same parameter values used for DUE, Hendrickson and Kocur (1981) attempted to achieve a pattern of arrival which results in minimum average user cost, corresponding to a system optimal equilibrium. The arrival function which has this characteristic is illustrated in Figure (4.5); the rate of arrivals at the bottleneck equals the bottleneck service rate (i.e. $m(t) = \mu$), so that no queue develops. The initial arrival time which minimises total user cost may be obtained using calculus. Total cost, TC, is the sum (or integral, since a fluid approximation is used) of schedule delay over all travellers, Q. This could be given by:

$$TC = \frac{1}{2} a_2 \mu x^2 + \frac{1}{2} a_3 \mu [Q/\mu - (x + 2\Delta)]^2 \quad \dots\dots\dots (4.50)$$

Where x is the schedule delay incurred by the first arrival (at time T_0). Differentiating Equation (4.50) with respect to x yields:

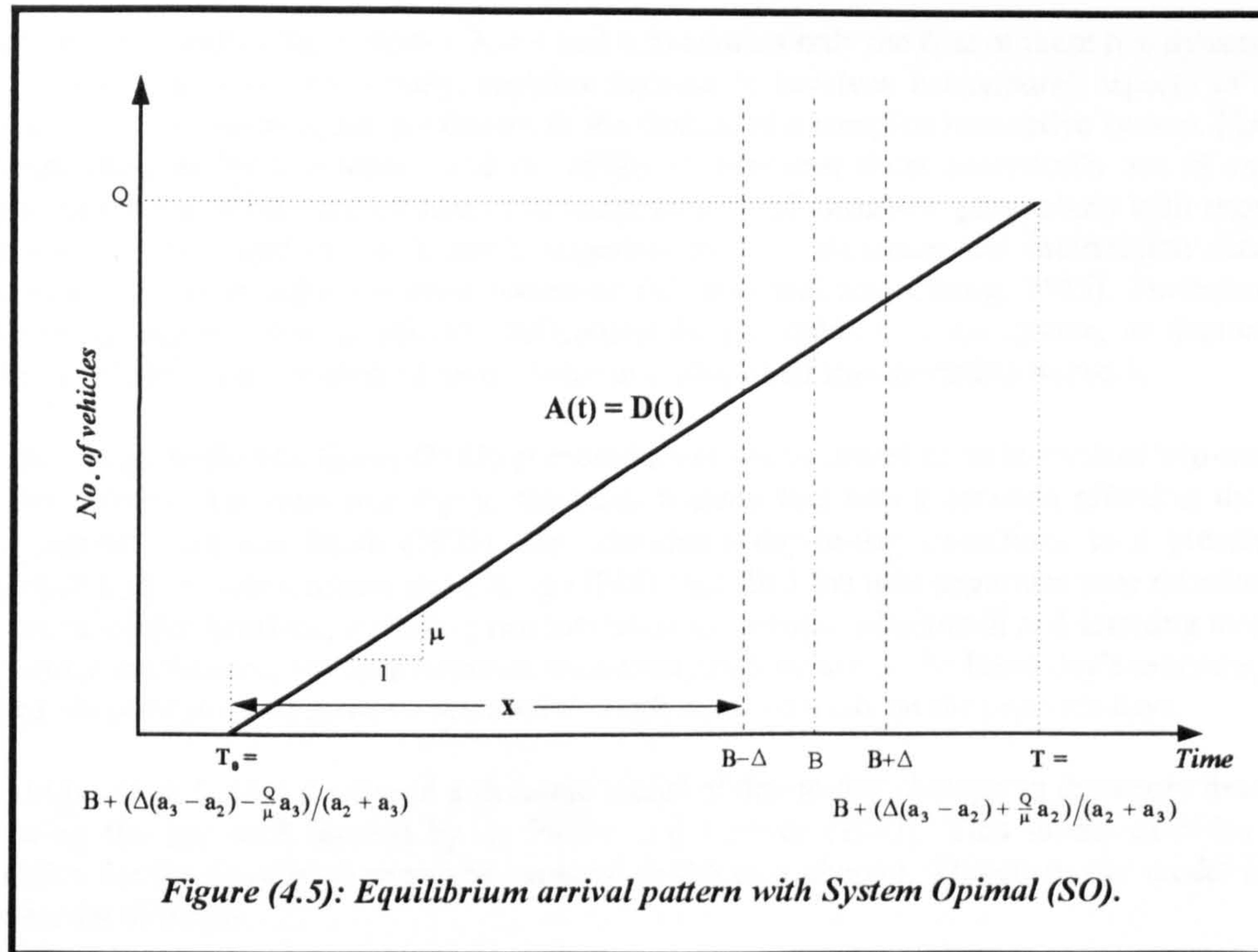
$$\partial TC / \partial x = a_2 \mu x - a_3 [Q/\mu - (x + 2\Delta)] \quad \dots\dots\dots (4.51)$$

which vanishes at the optimal value of x, i.e. x_{op} , given by:

$$x_{op} = a_3 / (a_2 + a_3) [Q/\mu - 2\Delta] \quad \dots\dots\dots (4.52)$$

(The second derivative of the total user cost given by Equation (4.50) is always positive; means the user cost is minimum at x_{op}).

This corresponds to an average user cost of $a_0 + a_2 a_3 / 2(a_2 + a_3) [Q/\mu - 2\Delta]$. The variable portion of user cost is exactly half that of DUE, and the initial and final times are equal to those of DUE.



To achieve the system optimal flow pattern a set of tolls (i.e. optimal tolls) must be established to create equal user cost for each arrival:

$$UC = a_0 + a_2 a_3 / (a_2 + a_3) [Q/\mu - 2\Delta] \quad (4.53)$$

Therefore, for early arrivals, schedule delay is decreasing with time, so optimal tolls must increase at a rate of a_2 pence per unit time to a maximum value of $a_2 a_3 / (a_2 + a_3) [Q/\mu - 2\Delta]$, which is the toll charged to any traveller arrives at the bottleneck at $[B-\Delta, B+\Delta]$. This value remains constant during this period, then, after time $(B+\Delta)$, the optimal tolls should decline at a constant rate of a_3 pence per minute until they reach zero at time T .

To summarise, the average user cost including the optimal tolls turns out to be equal to that of *DUE*; thus the tolls produce no internal benefits. In other words, as a result of imposing the optimal toll, there will be a cost transfer between the different components of the cost function for every individual such that no queue develops and without affecting the average user cost. All the dead-weight loss associated with the queue is thereby eliminated, resulting in a social saving equal to this dead-weight loss. The average optimal toll is equal to $a_2 a_3 / 2(a_2 + a_3) [Q/\mu - 2\Delta]$.

4.3 The day-to-day evolution of the departure-time decisions.

On a given day, the departure time decision of a trip-maker will yield a particular outcome (arrival time) which may or may not be acceptable to the user. Depending on the outcome, a readjustment may take place, whereby departure time is updated to reflect the user's current informational basis. Therefore, two dynamic aspects concerning the users departure time decisions and their interrelation with congestion should be considered: (1) the build-up and dissipation of system congestion on a given day, and (2) the day-to-day evolution of system conditions which depends on how users perceive and respond to congestion.

The above approaches (in sections 4.3, 4.4 and 4.5) address only the first of these two dynamic aspects. The second aspect is undoubtedly complex because it involves behavioural aspects of individual decision making, learning and judgement in the context of a complex interactive system. However, the understanding of these processes and the ability to represent them analytically are of considerable importance to the design and evaluation of congestion relief measures, particularly with regard to time lags that may be associated with user's responses to these measures and information dissemination programs that could influence these responses (Mahmassani and Chang, 1985). Furthermore, these behavioural aspects have significant implications for the stability of the system, as demonstrated by Horowitz (1984) in the context of route choice in a simplified transportation network.

An early related effort by Gaver (1967) proposed a normative model of an individual trip-maker trying to "beat" congestion from one day to the next, without this user's decision affecting the system's performance. Alfa and Minh (1979) also considered day-to-day transitions to a presumed stable departure pattern. Mahmassani and Chang (1986) modelled the user departure time decisions through the use of simple heuristic, including two mechanisms: myopic adjustment and learning model. Under the former mechanism, the user responds exclusively in function of the latest day's outcome, while the latter incorporates the experience acquired through repeated trials on the previous days.

Ben-Akiva et al. (1984) developed a dynamic model of day-to-day changes in departure time decisions following the approach studied by de Palma and Lefever (1983). This model provides a natural algorithm for the simulation approach explored in the next chapter. Therefore, the model is reviewed in some detail below.

Let $r(t, \omega)$ be the arrival rate at the bottleneck at time t on day ω . The number of individuals arriving during an interval of time $[t, t+\delta]$ on day ω is

$$x(t, \omega) = \int_t^{t+\delta} r(u, \omega) du \quad \dots\dots\dots (4.54)$$

Denote by $R(t, t', \omega) \Delta\omega$ the fraction of individuals who shift from an arrival during $[t, t+\delta]$ to an arrival during $[t', t'+\delta]$ during the time interval $[\omega, \omega+\Delta\omega]$. The rate of change of the number of individuals arriving during the interval $[t, t+\delta]$ can be expressed as the difference per unit time between the number of individuals shifting to $[t, t+\delta]$ and the number of individuals shifting from $[t, t+\delta]$, as follows,

$$\frac{\partial x(t, \omega)}{\partial \omega} = \sum_{t' \neq t} x(t', \omega) R(t', t, \omega) - x(t, \omega) \sum_{t' \neq t} R(t, t', \omega) \quad \dots\dots\dots (4.55)$$

Assume that the transition rates among different departure times can be modelled with the following simple dynamic logit model

$$R(t', t, \omega) = R1 \frac{e^{\bar{V}(t, \omega)/\eta}}{\sum_{t''} e^{\bar{V}(t'', \omega)/\eta}} \quad \dots\dots\dots (4.56)$$

Where $\bar{V}(t, \omega)$ is the observable utility of an arrival time during the interval $[t, t+\delta]$ on day ω and $R1$ is a constant transition rate of the current state. The structure of this dynamic model is a special case of a more general dynamic nested logit model developed by de Palma and Lefever (1983), de Palma and Ben-Akiva (1981) and Ben-Akiva and de Palma (1983). The simplifying assumption made here is that the utility of a shift to a new state is not dependent on the attributes of the current state.

Substitute (4.56) in (4.55) to get

$$\frac{\partial x(t, \omega)}{\partial \omega} = R 1. [Q \frac{e^{\bar{v}(t, \omega)/\eta}}{\sum_{t''} e^{\bar{v}(t'', \omega)/\eta}} - x(t, \omega)] \quad \dots\dots\dots (4.57)$$

Where Q as defined earlier is the potential number of travellers. Let $\delta \rightarrow 0$ to obtain the following continuous time limit of this dynamic system

$$\frac{\partial r(t, \omega)}{\partial \omega} = R 1. [Q \frac{e^{v(t, \omega)/\eta}}{\int e^{v(u, \omega)/\eta}} - r(t, \omega)] \quad \dots\dots\dots (4.58)$$

where the continuous logit model presented earlier replaces the discrete logit model. This last equation describes the evolution of the arrival rate at the bottleneck from day to day.

4.4 The stability of DUE under Time-Dependent Road Pricing (TDRP).

To examine the stability of *DUE* under *TDRP* the procedures under section 4.2.1 are employed and the toll function $f(t)=0$ is replaced with the *TDRP* function, $\tau(t) = T-t$, where the first derivative of this function is equal to -1 , i.e. $\tau'(t) = -1$, as derived in the previous chapter. The user cost function is then given by:

$$UC(t) = a_0 + a_1 w(t) + a_2 (B - \Delta - t - w(t)) + a_3 (t + w(t) - B - \Delta) + a_1 \tau(t) \quad \dots\dots\dots (4.59)$$

Then the different slopes of the cumulative arrival time function $A(t)$, at equilibrium, can be determined as follows:-

- For early arrival period $[T_0, \tilde{t}]$

The user cost is given by:

$$\begin{aligned} UC(t) &= a_0 + a_1 w(t) + a_2 (B - \Delta - t - w(t)) + a_1 \tau(t) \\ &= a_0 + (a_1 - a_2) w(t) + a_2 (B - \Delta) - a_2 t + a_1 \tau(t) \quad \dots\dots\dots (4.60) \end{aligned}$$

For the user cost to be constant, its derivative with respect to time must be zero, thus

$$\partial UC(t)/\partial t = 0 = (a_1 - a_2) (\partial w(t)/\partial t) - a_2 + a_1 \tau'(t)$$

Substituting with $\tau'(t) = -1$ gives

$$\partial w(t)/\partial t = (a_1 + a_2)/(a_1 - a_2) \quad \dots\dots\dots (4.61)$$

Substituting in Equation (4.56) by Equation (4.13) yields:

$$\partial w(t)/\partial t = (m(t) - \mu)/\mu \quad \dots\dots\dots (4.62)$$

Then

$$(m(t) - \mu)/\mu = (a_1 + a_2)/(a_1 - a_2)$$

$$m(t) = 2\mu a_1/(a_1 - a_2)$$

$$m1 = 2\mu a_1/(a_1 - a_2) \quad \dots\dots\dots (4.63)$$

which implies that the slope of $A(t)$ during the period of early arrival is constant and equal to $2\mu\alpha_1/(\alpha_1-\alpha_2)$.

- For the on-time arrival period $[\tilde{t}, \tilde{t}]$

The user cost is given by:

$$UC(t) = \alpha_0 + \alpha_1 w(t) + \alpha_1 \tau(t) \quad \dots\dots\dots (4.64)$$

As before the derivative of user cost with respect to time must be zero and Substituting by Equation (4.13) yields

$$m2 = 2\mu \quad \dots\dots\dots (4.65)$$

which implies that the arrival rate at the bottleneck during the period of on-time arrival $m2$ is as twice as the service rate μ .

- For late arrival period $[\tilde{t}, T]$

For a late arrival at work the user cost becomes:

$$UC(t) = \alpha_0 + \alpha_1 w(t) + \alpha_3 (t + w(t) - B - \Delta) + \alpha_1 \tau(t) \quad \dots\dots\dots (4.66)$$

Applying the same procedures used above for early and on-time arrivals, the slope of the arrival function for late arrivals is then:

$$m3 = 2\alpha_1 \mu / (\alpha_1 + \alpha_3) \quad \dots\dots\dots (4.67)$$

The arrival pattern in this case is similar to the one shown in Figure (4.2) with different values of $m1$, $m2$ and $m3$ which are as twice as their values before introducing the *TDRP*. This implies different changes in the lengths of early arrival, on-time arrival and late arrival periods in the produced pattern.

Solving Equation (4.59) with the constraint that, at equilibrium, costs of the first and last arrivals must be equal, the initial arrival time can be given by:

$$t_q = B + (\Delta(\alpha_3 - \alpha_2) - \frac{Q}{\mu} (\alpha_3 - \alpha_1)) / (\alpha_2 + \alpha_3) \quad \dots\dots\dots (4.68)$$

The queue vanishes at t_q which is given by:

$$t_q = B + (\Delta(\alpha_3 - \alpha_2) + \frac{Q}{\mu} (\alpha_1 + \alpha_2)) / (\alpha_2 + \alpha_3) \quad \dots\dots\dots (4.69)$$

The maximum queue length T_{Dm} is

$$(\alpha_3 - \alpha_1) (Q(\alpha_1 + \alpha_2) - 2 \alpha_2 \Delta \mu) / 2\alpha_1(\alpha_2 + \alpha_3)$$

Based on the values of α_1 , α_2 and α_3 given by Small (1982), this queue length is always less than the maximum queue length developed before *TDRP*.

The user cost, at equilibrium, including tolls is given by:

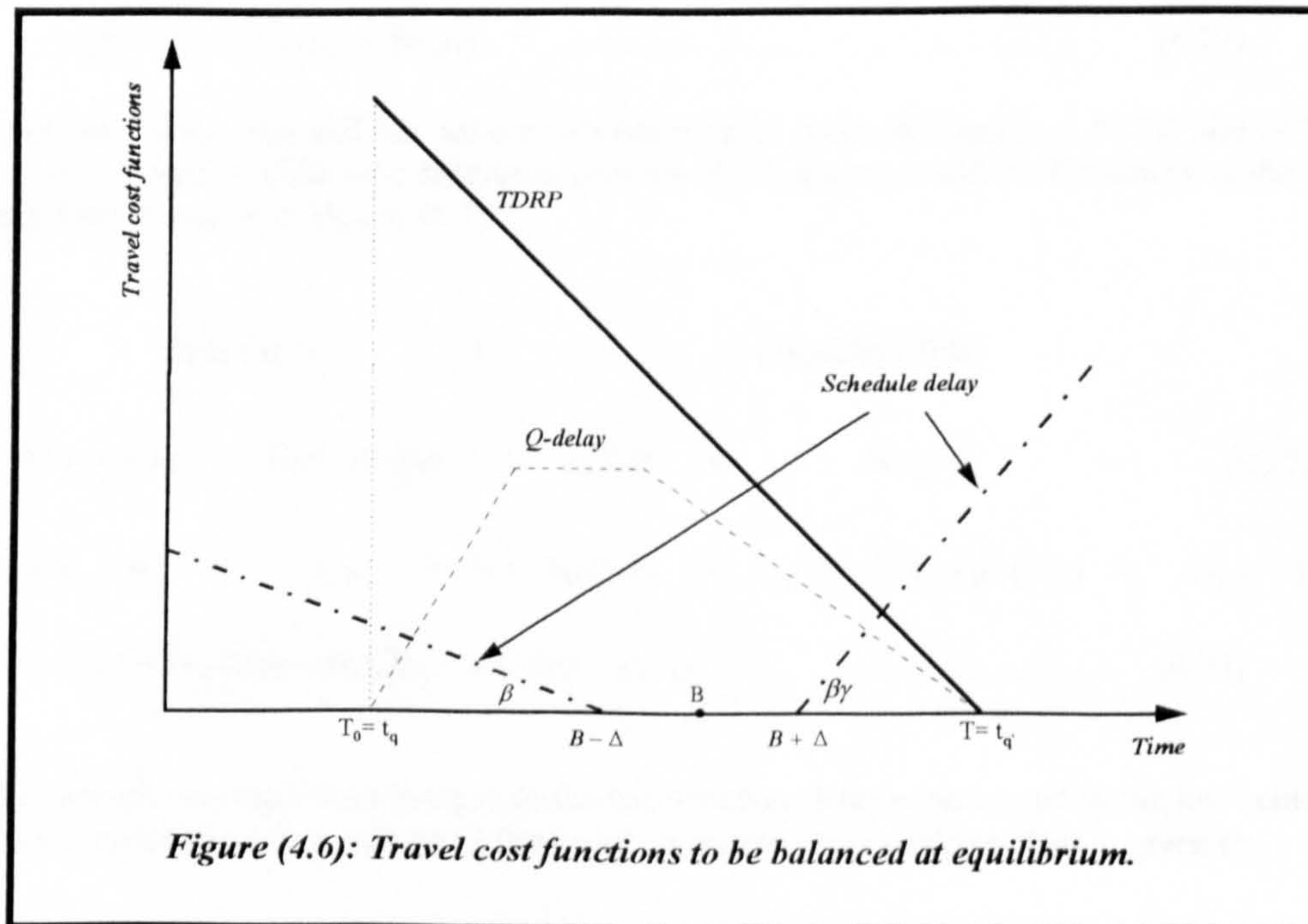
$$a_0 + a_3 \left(\frac{Q}{\mu} (a_1 + a_2) - 2 a_2 \Delta \right) / (a_2 + a_3)$$

and the average toll is $a_1 Q / 2\mu$ which is higher than the average optimal toll derived in section 4.2.3. The user cost excluding the average toll is then given by

$$a_0 + \left(\frac{Q}{2\mu} (a_1 a_3 + 2a_2 a_3 - a_1 a_2) - 2 a_2 a_3 \Delta \right) / (a_2 + a_3)$$

which turns out to be higher than the user cost before *TDRP* by $a_1 Q (a_3 - a_2) / 2\mu (a_2 + a_3)$. Paradoxically, however, instead of producing internal benefit, *TDRP*, in such a form, increases the user cost. The reason for that increase in user cost is revealed by comparing the initial arrival time in this case with the one given by Equation (4.24), i.e. the initial arrival time before *TDRP*. This shows that to get equilibrium under the three different functions shown in Figure (4.6), travellers are forced to shift their departure times to the right by $Q a_1 / \mu (a_2 + a_3)$ to balance the sum of the maximum toll and the early schedule delay cost at t_q with the schedule delay cost due to late arrival at t_q , since the delay cost function is equal to zero at the two ends, i.e. at t_q and t_q . This ends up with an equilibrium pattern which, in fact, is very unstable since any traveller can reduce his travel cost by changing his departure time to any time before t_q and face only the early arrival schedule delay. This is true as long as the schedule delay of his early arrival is less than the equilibrium user cost obtained under *TDRP*.

However, two different approaches are adopted in this study to modify *TDRP* in order to lead to a stable equilibrium: first, by modifying *TDRP* to include the schedule costs and benefits imposed by one another and second, using the day-to-day adjustment process. These two approaches are discussed in some detail in the following two sections.



4.4.1 Modifying TDRP to include the schedule delay changes to other vehicles

In modelling *TDRP* in Chapter (3) only the queuing delay imposed by the arriving vehicle at any time on other vehicles behind was considered. As a matter of fact, any additional vehicle joining the queue at any time causes some changes in the schedule delay to all (or most of) the vehicles joining the queue afterwards in addition to the queuing delay imposed on these vehicles. In other words, any additional vehicle joining the queue at any time causes a shift to the arrival times of every individual vehicle behind by $1/\mu$. Thus every vehicle arriving afterwards and experiencing late arrival (in the absence of this vehicle) will incur an additional late schedule delay cost of a_3/μ , while every vehicle arriving afterwards and experiencing an early arrival will experience an additional benefit of a_2/μ as a result of reducing its early schedule delay by $1/\mu$. Also some of the vehicles experiencing zero schedule delay, in the absence of the additional vehicle, will incur a late schedule delay cost of a_3/μ as a result of the arriving vehicle. Therefore, in this section, the *TDRP* function is modified to include the schedule delay changes to other vehicles in addition to the queuing delay costs.

The total additional schedule delay cost imposed by the arriving vehicle at time t on others is equal to a_3/μ times all the vehicles joining the queue afterwards and experiencing late arrival. This is given by:

$$Sc_l(t) = (a_3/\mu) \cdot \int_t^T \lambda(u) du \quad \dots\dots\dots \forall u \in [\tilde{t}, T] \quad \dots\dots\dots (4.70)$$

Similarly, the total additional schedule delay benefit as a result of the arriving vehicle at time t is given by:-

$$Sc_e(t) = (a_2/\mu) \cdot \int_t^{\tilde{t}} \lambda(u) du \quad \dots\dots\dots \forall u \in [T_0, \tilde{t}] \quad \dots\dots\dots (4.71)$$

The modified toll at any time t , $\tau_m(t)$, is then given by:

$$\tau_m(t) = \tau(t) + Sc_l(t) - Sc_e(t) \quad \dots\dots\dots (4.72)$$

The user cost for the first and the last arrivals can then be given by Equations (4.73) and (4.74) as a function of x , where x is the time difference between the first arrival and the beginning of the on-time arrival period as shown in Figure (4.7).

$$\begin{aligned} UC_{T_0} &= \text{(Real Cost)} + \text{(Modified Toll)} \\ UC_{T_0} &= (Q\text{-delay} + \text{Sch. delay}) + (TDRP + Sc_l(T_0) - Sc_e(T_0)) \\ &= (0 + a_2 x) + (a_1(Q/\mu) + a_3(Q - x - 2\Delta) \cdot \mu \cdot (1/\mu) - a_2(x, \mu) \cdot (1/\mu)) \\ &= a_2 x + (a_1(Q/\mu) + a_3(Q/\mu - x - 2\Delta) - a_2 x) \quad \dots\dots\dots (4.73) \end{aligned}$$

The user cost for the last arrival is equal to the late schedule delay experienced by the last vehicle since this vehicle experiences zero delay and has no effect on any other vehicles. This is given by:

$$UC_T = a_3(Q/\mu - x - 2\Delta) \quad \dots\dots\dots (4.74)$$

At equilibrium $UC_{T_0} = UC_T$, then $a_1(Q/\mu)$ must equal to zero which implies that no queue develops in this case. Thus, the modified toll at T_0 is then given by:

$$\tau_m(T_0) = a_3(Q/\mu - x - 2\Delta) - a_2 x \quad (4.75)$$

For a stable equilibrium, the toll at T_0 must equal to zero to avoid any sudden change in the user cost at that time. Thus, solving Equation (4.75) under this condition will give the value of x at equilibrium, i.e. x_{eq} .

$$x_{eq} = a_3(Q/\mu - 2\Delta)/(a_2 + a_3) \quad (4.76)$$

Consequently,

$$\begin{aligned} T_0 &= (B - \Delta) - x_{eq} \\ &= B - (a_3(Q/\mu) - (a_3 - a_2)\Delta)/(a_2 + a_3) \end{aligned} \quad (4.77)$$

Substituting in Equation (4.74) by Equation (4.76) gives the user cost at equilibrium:

$$UC_{eq} = a_2 a_3(Q/\mu - 2\Delta)/(a_2 + a_3) \quad (4.78)$$

The above results are similar to the results of the user cost and the initial arrival time obtained under the system optimal. Therefore, it is concluded that modifying *TDRP* to include the schedule delay changes to other vehicles will lead to a system optimal *SO*, i.e. queues will be eliminated and the optimal toll becomes equivalent to these schedule delay changes.

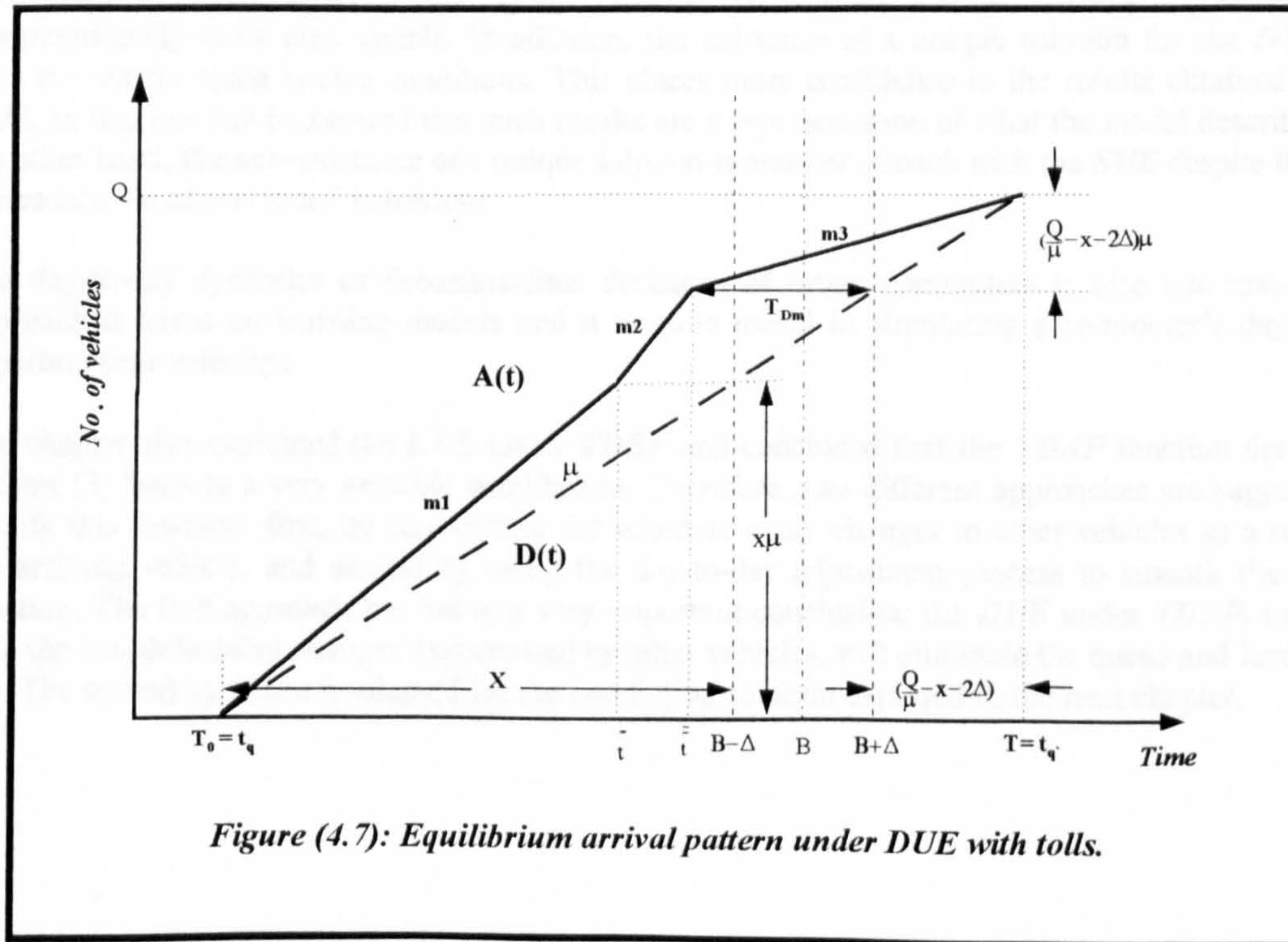


Figure (4.7): Equilibrium arrival pattern under DUE with tolls.

4.4.2 Modifying *TDRP* using the day-to-day adjustment process

To smooth the *TDRP* function in order to lead to a stable equilibrium, the day-to-day adjustment process adopted for the departure time decisions is employed. This process is used to adjust the value of the *TDRP* for every departure time from the bottleneck based on the amount of tolls at that particular time on the previous day. Thus, the *TDRP* at any departure time t from the bottleneck during day ω , $\tau(t, \omega)$, is given by

$$\tau(t, \omega) = R2.(T-t) + (1-R2). \tau(t, \omega-1) \dots\dots\dots (4.79)$$

Where T is the time at which the queue disappeared and $R2$ is the adjustment ratio and it is assumed to be constant for all travellers and from day-to-day.

This approach is adopted for the simulation solution explored in the next chapter.

4.5 Summary

Three main approaches, i.e. the Deterministic User Equilibrium *DUE*, Stochastic User Equilibrium *SUE* and System Optimal *SO*, have been used to study how travellers, who have particular DTTs, select their departure times from origin in a simple network. This simple network consists of one origin-destination pair connected by one major route with a bottleneck. Each of these approaches is capable of being used for simulating the effect of changes in DTT, demand or capacity on the temporal distribution of traffic demand.

Of the three approaches, the *SO* is the least realistic in terms of representing travellers' behaviour. The *DUE* and *SUE* approaches both have very important weaknesses making it very difficult to select which one should be considered for practical applications. The *SUE* is more realistic than the *DUE*, in that it represents a traveller's behaviour more closely. It does not assume that a traveller has perfect knowledge of travel costs in the system as the *DUE* does assume. However, the *DUE* is computationally more manageable. In addition, the existence of a unique solution for the *DUE* has been proved for some known conditions. This places more confidence in the results obtained by the *DUE*, in that one can be assured that such results are a representation of what the model describes. On the other hand, the non-existence of a unique solution is another setback with the *SUE* despite its close representation of travellers' behaviour.

The day-to-day dynamics of departure-time decisions of urban commuters is also addressed. This approach is based on learning models and it is quite useful in simulating a commuter's day-to-day departure time selection.

The chapter also examined the *DUE* under *TDRP* and concluded that the *TDRP* function derived in chapter (3) leads to a very unstable equilibrium. Therefore, two different approaches are suggested to modify this function: first, by considering the schedule delay changes to other vehicles as a result of the arriving vehicle, and second by using the day-to-day adjustment process to smooth the *TDRP* function. The first approach has led to a very important conclusion: the *DUE* under *TDRP*, modified with the schedule delay changes experienced by other vehicles, will eliminate the queue and lead to the *SO*. The second approach is adopted for the simulation solution explored in the next chapter.

CHAPTER FIVE

5. Simulation Solution for a Single traffic Bottleneck

5.1 Introduction

5.2 The dynamic simulation model

- 5.2.1 Structure of the simulation model.
- 5.2.2 Deterministic departure-time selection process.
- 5.2.3 Stochastic departure-time selection process.
- 5.2.4 Other features of the simulation model.

5.3 Simulation experiments (A numerical example)

- 5.3.1 Base values of the model parameters.
- 5.3.2 Convergence and stability of the base case results.
- 5.3.3 Base case departure time choices and Generalized Cost results.
- 5.3.4 The different impacts of *TDRP* on the base case results.

5.4 Summary and concluding remarks

5.1 Introduction

The dynamic simulation is the numerical solution of the analytical models described in the previous chapter. However, in moving towards more practically-oriented models, the elegance and obvious limitations of a closed form analytical solution have been abandoned in favour of a simulation-based approach, which allows considerably more flexibility and realism. For instance, highway performance follows well-established traffic-flow theoretic relationships underlying a special purpose traffic simulation model, which takes as input the aggregated departure time decisions of users. These decisions are modelled using heuristic behavioural rules governing departure time selection on a given day, in conjunction with explicit models of the users' learning through repetitive daily use of the facility. The overall modelling system is exercised over a sequence of days, with user decisions and system congestion fully interdependent over time. The effect of alternative choices rules and learning models on the system's performance and its dynamic properties (particularly convergence and stability) can thus be explored (Mahmassani and Chang, 1986).

Therefore, this chapter describes a framework for the simulation solution used for a single traffic bottleneck (this framework could also be used on a large scale for a typical network as explored in the next chapter). This framework is then used to perform a set of simulation experiments (using a numerical example) to investigate the base case results and the various effects of *TDRP* on these results. A detailed description of the dynamic simulation model is given in Section (5.2). The simulation experiments performed are described in Sections (5.3.1), followed by a presentation and critical discussion of key results in Sections (5.3.2–5.3.4). A summary and concluding remarks are presented in Section (5.4).

5.2 The dynamic simulation model

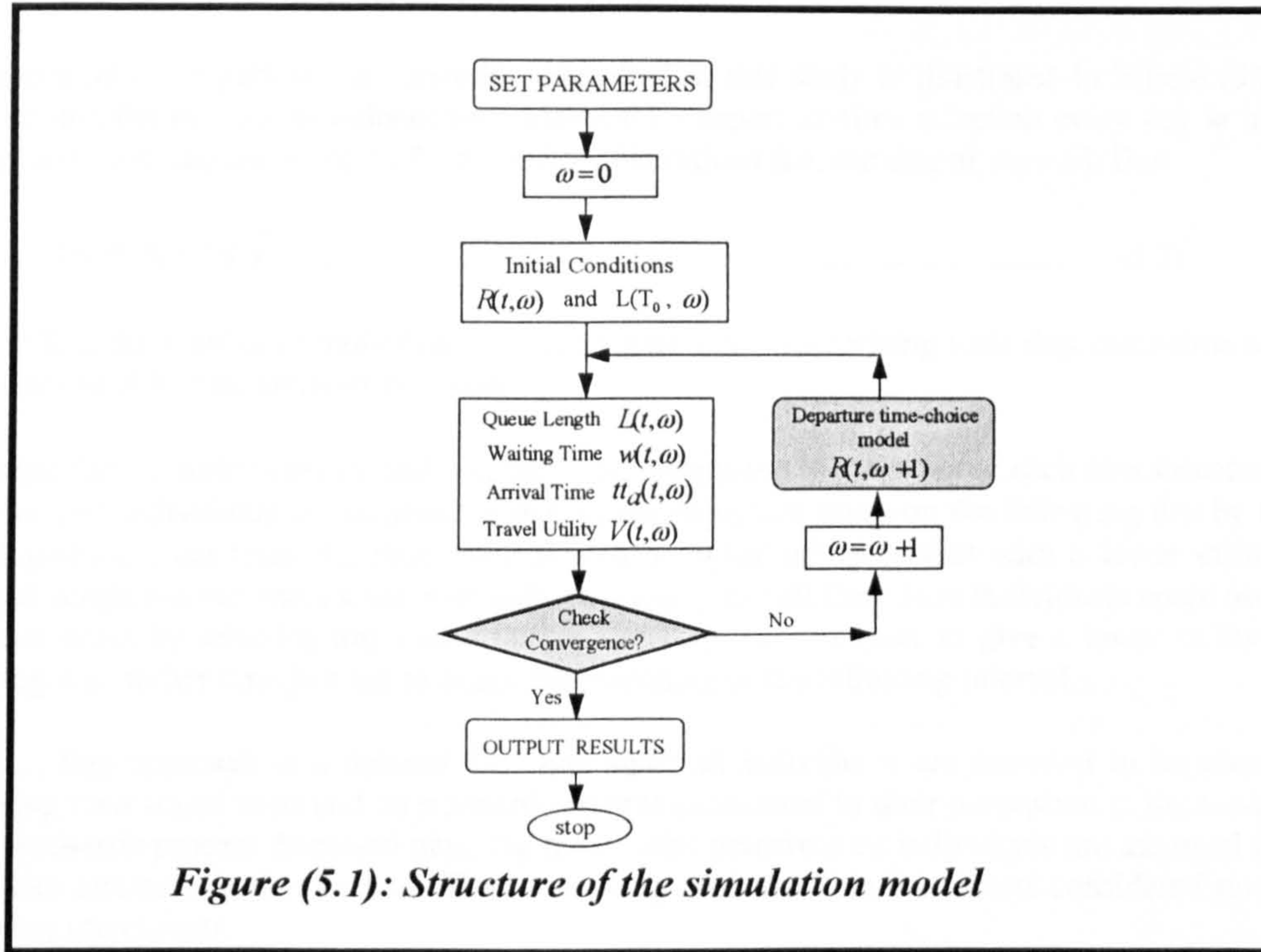
This section describes the general structure of the simulation solution for the dynamic models of peak period and summarises its main features.

5.2.1 Structure of the simulation model

To simulate the dynamic models of peak period, the study period $[T_0, T]$ is divided into equal time intervals of length δ . Define $N(t, \omega)$ to be the number of users choosing the departure time interval $[t, t+\delta]$ on day ω . The parameter δ could be interpreted as a measure of the ability of individuals to discriminate among alternative departure times. This view is supported by Mahmassani et al. (1984) who developed an experimental procedure to study the choice of departure time and found that the participants adjusted their departure times by multiples of 5 minutes, with a minimum adjustment interval of 5 minutes. However, Ben-Akiva et al. (1986a) concluded that if δ is small enough (of the order of 5 to 10 minutes), the results are extremely stable. In the following, $R(t, \omega)$ will denote the departure rate per unit time that is equal to $N(t, \omega)/\delta$.

The general structure of the simulation algorithm is shown in Figure (5.1). It begins with a specification of the parameters of the simulation model and the initial conditions given by the departure rates on day $\omega=0$, and the initial queue length at the beginning of the study period. Subsequently, an iterative process is used with an iteration corresponding to a day. This process is terminated when either the last specified day of the study (i.e. number of iterations required) is reached or convergence occurs. The condition for convergence is when the coefficient of variation of the generalized cost (or utility) at any day ω is less than a pre-specified tolerance factor, i.e.

$$\frac{\sqrt{\sum_{t=1}^K N(t, \omega) \cdot (\bar{GC}(\omega) - GC(t, \omega))^2}}{\bar{GC}(\omega)} \bigg/ \sum_{t=1}^K N(t, \omega) \leq \varepsilon \quad t=1, 2, 3, \dots, K \quad \dots (5.1)$$



where $GC(t, \omega)$ is the generalized cost for any individual selecting his departure time during the period $[t, t+\delta]$ on day ω and $\overline{GC}(\omega)$ is the average generalized cost for all users on day ω . K is the total number of time intervals and ε represents a prescribed tolerance factor for the deviation from the mean. This condition of equilibrium is set to fulfil the definition of Wardrop's user equilibrium. The dynamic deterministic user equilibrium is defined as the state at which *no traveller can reduce his travel cost by unilaterally changing his departure time (or route)*. In this case, it is assumed that travellers have identical tastes, and that there are no errors involved in the perception of the actual trip characteristics; therefore travellers' perceived times and schedule delays are equal to the actual ones. Therefore at the equilibrium state, every traveller will have the same travel cost. Likewise, Vythoulkas (1991) defined the dynamic stochastic user equilibrium as the state at which *no traveller believes he can reduce his perceived travel cost by unilaterally changing his departure time (or route)*. In this case, travel cost (or utility) is assumed to be a random variable, reflecting the inaccuracies and distortions in travellers' perceptions of trip characteristics.

In every iteration, the following quantities are computed in sequence for every time interval using the deterministic queuing model discussed in chapter (3):—

- Queue Length $L(t, \omega)$;
- Waiting Time $w(t, \omega)$;
- Arrival Time $tt_a(t, \omega)$; and
- Utility Function $V(t, \omega)$ or Generalized Cost $GC(t, \omega)$.

Convergence is checked at the end of each iteration and as long as the condition for convergence is not reached, individuals use information gathered from the last day ω in making their choices on the following day $\omega+1$. In this study, two different approaches are adopted for simulating the process used by individuals to revise their departure time choices from one day to another: first, using a deterministic model based on heuristic assumptions, and second using a stochastic model. These two approaches are described below.

5.2.2 Deterministic departure-time selection process

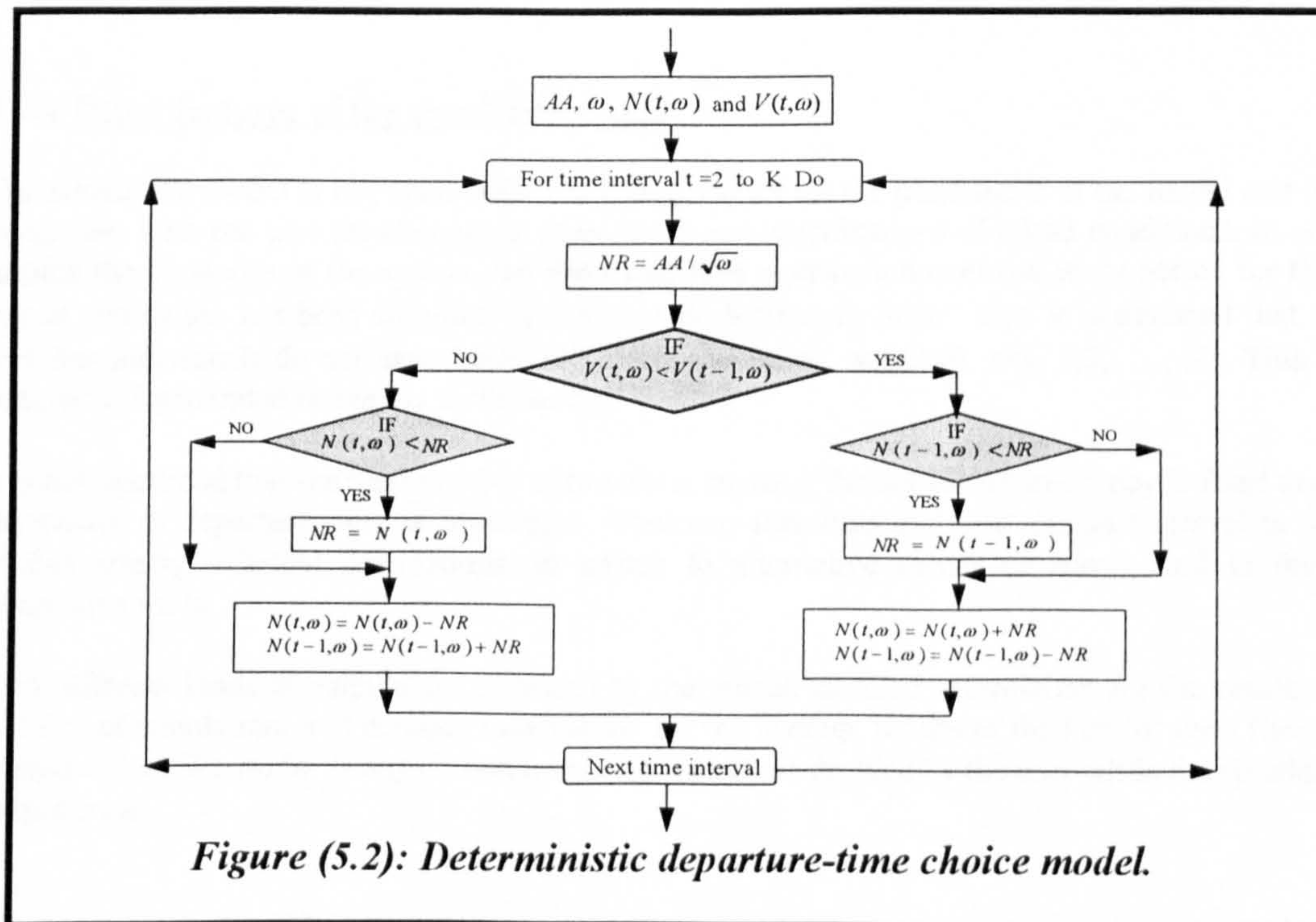
The deterministic departure time selection suggested in this study is illustrated in Figure (5.2). It is assumed that the number of individuals revising their departure-time selection every day is inversely proportional with the square root of the number of iterations (i.e. number of days ω), thus

$$NR = AA / \sqrt{\omega} \quad \dots\dots\dots (5.2)$$

Where NR is the number of individuals (for every time interval) revising their departure-time selection every day and AA is an arbitrary constant.

This algorithm is quite heuristic and it is based on comparing the utilities of each two successive time intervals, and individuals are assumed to select their departure times on the following day by altering their departure times from the time interval with a higher utility to that with a lower utility. This approach might not represent what is actually happening in real life, since individuals could alter their departure times by selecting any time interval that they might expect to give a lower utility on the following day, rather than just opt to either the preceding or the following interval.

However, this approach is a deterministic one since all individuals are assumed to be identical in perceiving their travel costs and no stochastic error is considered in their perception to the travel costs. In the stochastic process discussed next, the travel costs perceived by individuals are assumed to differ by random amounts, owing to the differences among individuals in the factors considered relevant to evaluating travel costs.



5.2.3 Stochastic departure-time selection process

Following de Palma and Lefever (1983) and Ben-Akiva et al. (1984, 1986a and 86b), the stochastic departure time selection used in the simulation solution in this study, is given by:

$$R(t, \omega) = R1 \frac{Q}{E(\omega)} \exp\left[\frac{1}{\eta} GC(t, \omega)\right] + (1 - R1) \cdot R(t, \omega - 1) \dots\dots\dots (5.3)$$

where $R(t, \omega)$ is the departure rate during the time interval $[t, t+\delta]$ on day ω . Q as defined earlier is the total number of users during the study period. $R1$ is the fraction of users who review their departure time choices every day and it is assumed to be constant. $GC(t, \omega)$ is the generalized cost for time interval $[t, t+\delta]$ on day ω . η is a scale parameter representing the degree of heterogeneity of preferences among individuals. $E(\omega)$ is the denominator of the logit model and is approximated by the following discrete form:

$$E(\omega) = \frac{1}{2} \left[\exp\left(\frac{1}{\eta} GC(0, \omega)\right) + \exp\left(\frac{1}{\eta} GC(K, \omega)\right) \right] + \sum_{t=1}^{K-1} \exp\left(\frac{1}{\eta} GC(t, \omega)\right) \quad (5.4)$$

where K is the total number of intervals ($K=(T-T_0)/\delta$) and δ is the time increment (or time interval).

This adjustment process is derived from a dynamic Markovian model, for more details see Ben-Akiva et al. (1986b). It is assumed that the trip decisions on day $(\omega+1)$ are based on the distribution of travel times and delays on the previous day ω . That is, road users are assumed to be perfectly informed about the road conditions during the previous day. Ben-Akiva et al. (1986a) pointed out that it would be useful to explore situations in which individuals are not perfectly informed, differ in their level of information and ability to form expectations, and base their decisions on their past experience. Some of these issues have been explored by Horowitz (1984) who investigated the consequences of alternative assumptions on the choice between two parallel roads with constant demand. Mahmassani et al. (1984) considered an adjustment process based on a satisfaction criterion (as mentioned in the previous chapter).

5.2.4 Other features of the simulation model

The simulation model is implemented with default values for the parameters of the model and is fully interactive with the user for changes in parameters and specifications of initial conditions. In order to initiate the dynamics of the system, the departure time distribution over the study period for the first day of simulation has been specified by the "Pseudo-Stationary State" (i.e. it is assumed that on the first day individuals do not experience any congestion delay, $w(t,0)=0$ $t=1, 2,3,\dots, K$). This initial condition is generated internally by the model.

It is also assumed that the total number of travellers crossing the bottleneck every day is fixed and only the choice of departure time is considered. However, travellers can also decide to travel or not, to choose among different destinations, to switch to alternative modes of travel, and to divert to alternative route.

Two different kinds of outputs are produced by the model: detailed information for the results of the last day of simulation, and detailed information for the average results of the last 'n' days (iterations) of simulation. Where 'n' is any chosen number of iterations decided by the user while the simulation is on progress.

5.3 Simulation experiments (A numerical example)

The numerical example used for the analysis in this chapter is assuming that every morning a fixed number, Q , of commuters travel between home (A) and work (F). It is also assumed that there is only one single route between A and F, and travel along this route is uncongested except at a single bottleneck (between A and F) with a limited capacity μ .

In this section, the base case which is the starting point of the subsequent analysis is described. Then the convergence and the stability of the results are examined under the deterministic and stochastic

departure time selection processes discussed above. The sensitivity of the results to different values of the scale parameters " η " and the reallocation factor "R1" is also examined. Finally the various impacts of *TDRP* on the departure time pattern and the generalized cost of the base case results are investigated.

5.3.1 Base values of the model parameters

The following values are used to represent the base case for the simulation experiments:

Total number of vehicles (Q) = 18000 vehicles
 Capacity of the bottleneck (μ) = 7200 vehicles/hour
 Scale parameter of the Stochastic Choice Model (η) = 1
 On-time arrival period (2Δ) = 30 minutes

The study period of the day considered (defined by T_0 and T) is [5:00 a.m., 10:00 a.m.], and the desired arrival period at F is [8:45 a.m., 9:15 a.m.], that is the centre of this period, B , is 9:00 a.m. and $\Delta = 15$ minutes. The time interval, δ , is 1/12 hr. (i.e. 5 minutes), this implies a total of 60 time intervals throughout the study period. ϵ , the tolerance parameter is 0.01.

The shadow values of travel time, early arrival time and late arrival time (α , β , and $\beta\gamma$) are \$6.40/hour, \$3.90/hour, and \$15.20/hour respectively. These values are derived from Small (1982) and imply that a minute of late schedule delay is valued the same as 3.9 minutes of early schedule delay, and an additional minute of an early schedule delay is valued the same as 0.6 minutes of extra travel time.

5.3.2 Convergence and stability of the base case results

a. Deterministic departure time process

In examining the convergence and stability of the results under the deterministic approach for departure time choices, different values for the arbitrary constant, AA , are tested. The simulation results for these tests are summarised in Table (5.1) below.

Table (5.1): Base case convergence results under different values for the constant AA .

Constant ' AA '	No. of Iterations ' ω '	$NR = AA / \sqrt{\omega}$	Coeff. of Variation	Av. G. Cost
50	3000	0.913	0.489	4.552
100	3000	1.825	0.349	6.895
200	1828	4.678	0.009	5.959
300	882	10.102	0.009	5.991
400	524	17.470	0.008	5.962
500	552	21.280	0.009	5.979
600	722	22.330	0.011	6.055
700	582	29.020	0.010	6.045
800	1652	19.680	0.011	6.035
900	1280	25.160	0.009	6.038
1000	2114	21.750	0.012	5.955
2000	3254	35.060	0.009	6.070

The above table shows that not all the values of the constant AA could lead to convergence: for example, after 3000 iterations, no convergence is reached for AA equal to 50 or 100. However, any value between 200 and 2000 could lead to a convergence after different number of iterations. Table

(5.1) also shows that there is no fixed relation between the value of AA and the number of iterations required for convergence, but as the value of AA increases (above 900) the number of iterations required for convergence increases dramatically. In this particular example, the optimal value of AA (in terms of the number of iterations required for convergence) is 400. Therefore, this value is used in producing all the subsequent results under the deterministic departure time process.

Figure (5.3) below shows the fluctuation of the coefficient of variation of the generalized cost over the number of iterations. The figure depicts that the coefficient of variation is falling sharply after 200 iterations and the system converged to a stationary state (with some insignificant oscillations). The average coefficient of variation for the last 25 iterations is 0.02.

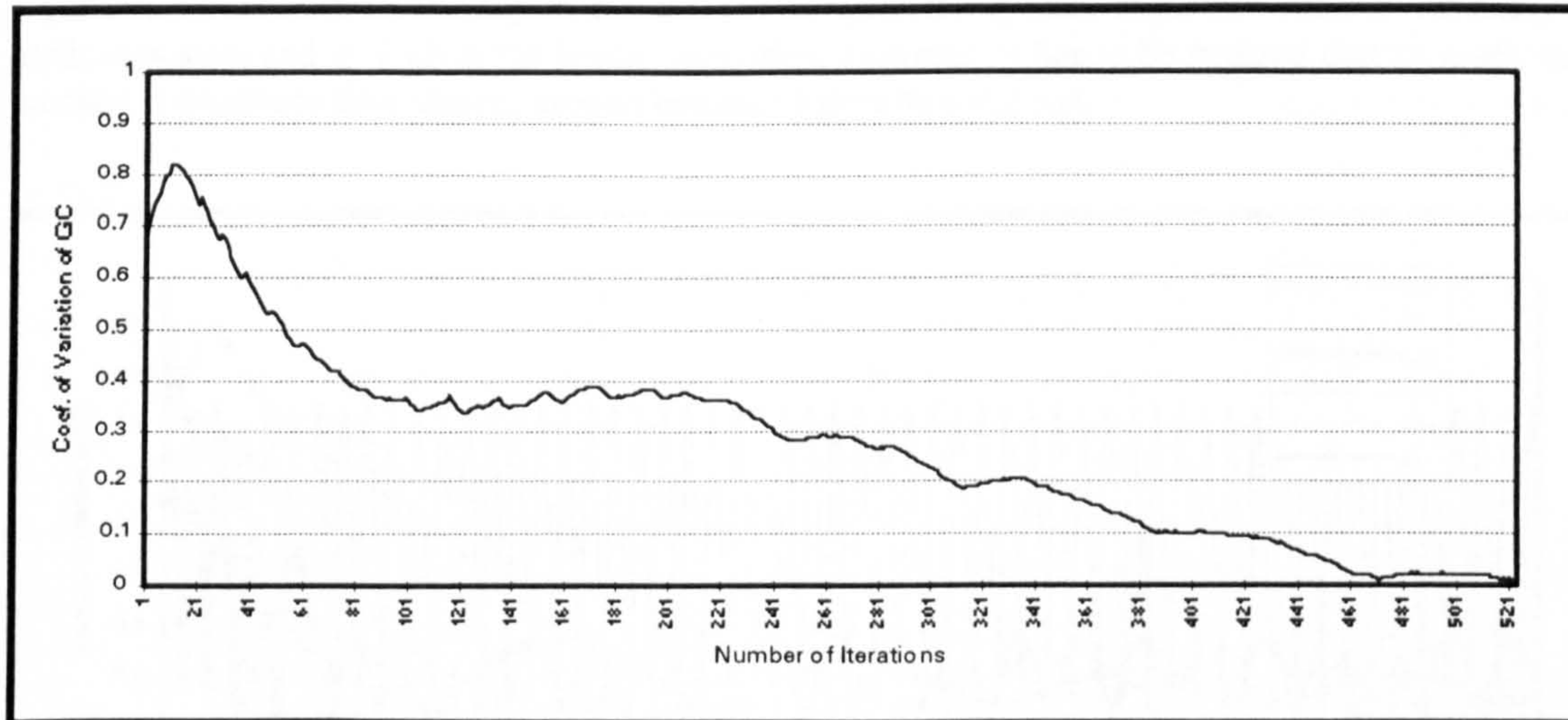


Figure (5.3): The fluctuation of the coefficient of variation of Generalized Cost under the Deterministic Model.

b. Stochastic departure time process

The convergence and the stability of the results under the stochastic departure time process is examined under different values of the reallocation factor R1 and the scalar (or randomness) parameter η . The fluctuation of the coefficient of variation of the generalized cost under different values of R1 and $\eta=1$ is shown in Figure (5.4). The different values of R1 demonstrated in this figure are: 0.2, 0.1, 0.05, 0.01 and R1* that is differentiated according to the number of iterations ω as given by Equation (5.5) below.

$$R1^* = \begin{cases} 0.20 & \text{---} \omega \leq 50 \\ 0.10 & \text{---} 100 \geq \omega > 50 \\ 0.05 & \text{---} 150 \geq \omega > 100 \\ 0.025 & \text{---} 200 \geq \omega > 150 \\ 0.010 & \text{---} \omega > 200 \end{cases} \quad (5.5)$$

As a matter of fact, using a differentiated value of R1 is quite realistic because the percentage of travellers reviewing their departure times is expected to be higher at the beginning and declines eventually as the system approaches its final state. Also Ben-Akiva et al. (1986a) pointed out that the rate of convergence to a stationary state is dependent on the value of the reviewing factor, R1, and for a high value, a convergence to a stationary state is not guaranteed, while for small values convergence occurs towards a unique stationary state.

Returning to Figure (5.4), it is shown that after a certain number of iterations the fluctuation of the coefficient of variation started to oscillate (following a COS-Function) for all the values of $R1$. Increasing the value of $R1$ increases the fluctuation of the coefficient of variation and decreases the cycle length of the oscillation. It also takes longer for smaller values of $R1$ (i.e. 0.02 and 0.01) to reach this state of oscillation. However, the differentiated value of $R1$ given by Equation (5.5) helps to overcome this problem and leads to a convergence with the lowest oscillation after 1000 iterations. Therefore, $R1^*$ is used in all the subsequent results in this chapter.

Figure (5.5) illustrates the fluctuation of the coefficient of variation of the generalized cost for $R1^*$ and different values of the randomness parameter η (i.e. 0.1, 0.5, 1.0, 1.5, 2.0 and 3.0). As η increases the variations in the travel costs perceived by travellers also increase, and this will consequently lead to a very stable convergence with a very high coefficient of variation between travellers (as shown in the figure for $\eta > 1$). For $\eta \leq 1$, the system oscillates without reaching a stationary state and $\eta = 1$ gives the lowest oscillation. However, it has to be realised that as $\eta \rightarrow 0$, the stochastic departure time choices process becomes a deterministic one.

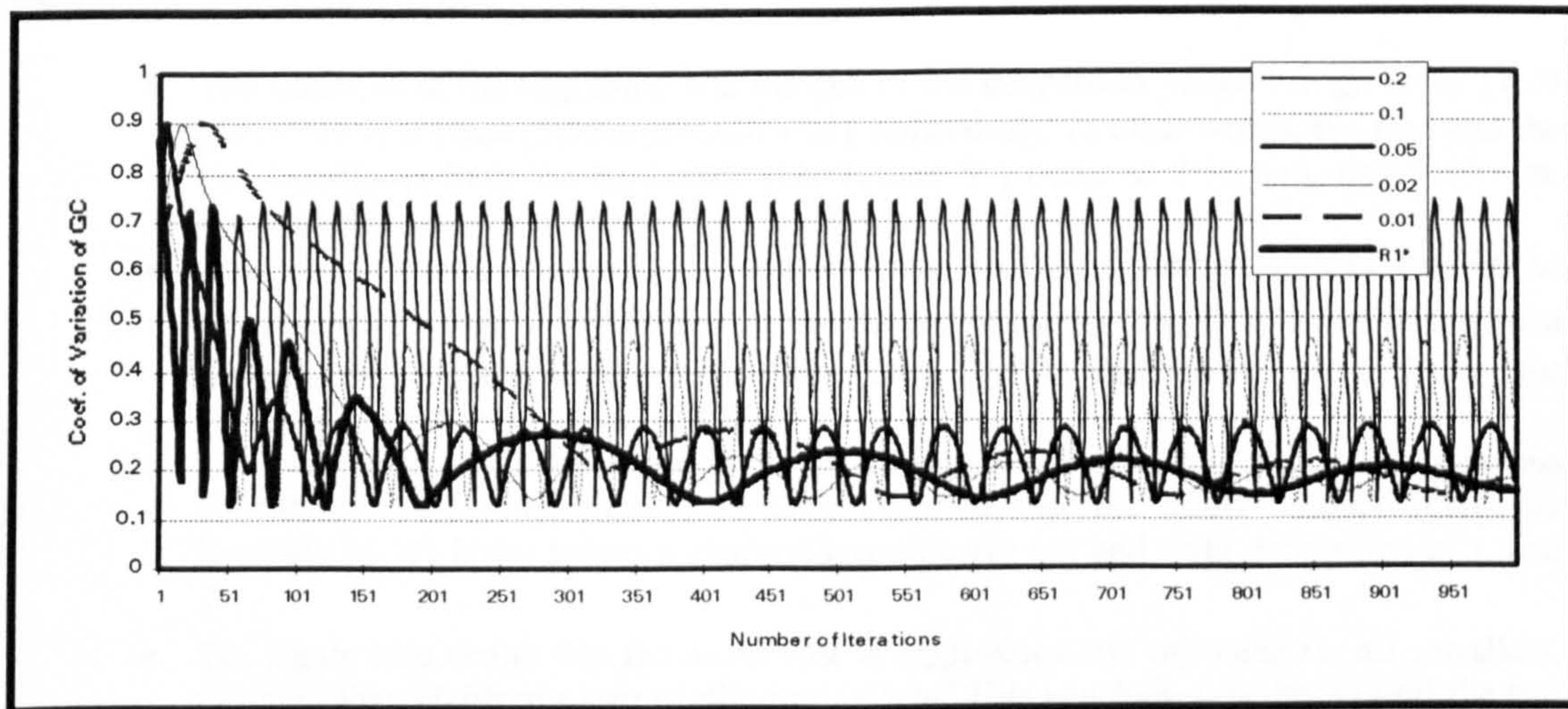


Figure (5.4): The fluctuation of the coefficient of variation of Generalized Cost under different values of the re-allocation factor ' $R1$ ' (Stochastic Model).

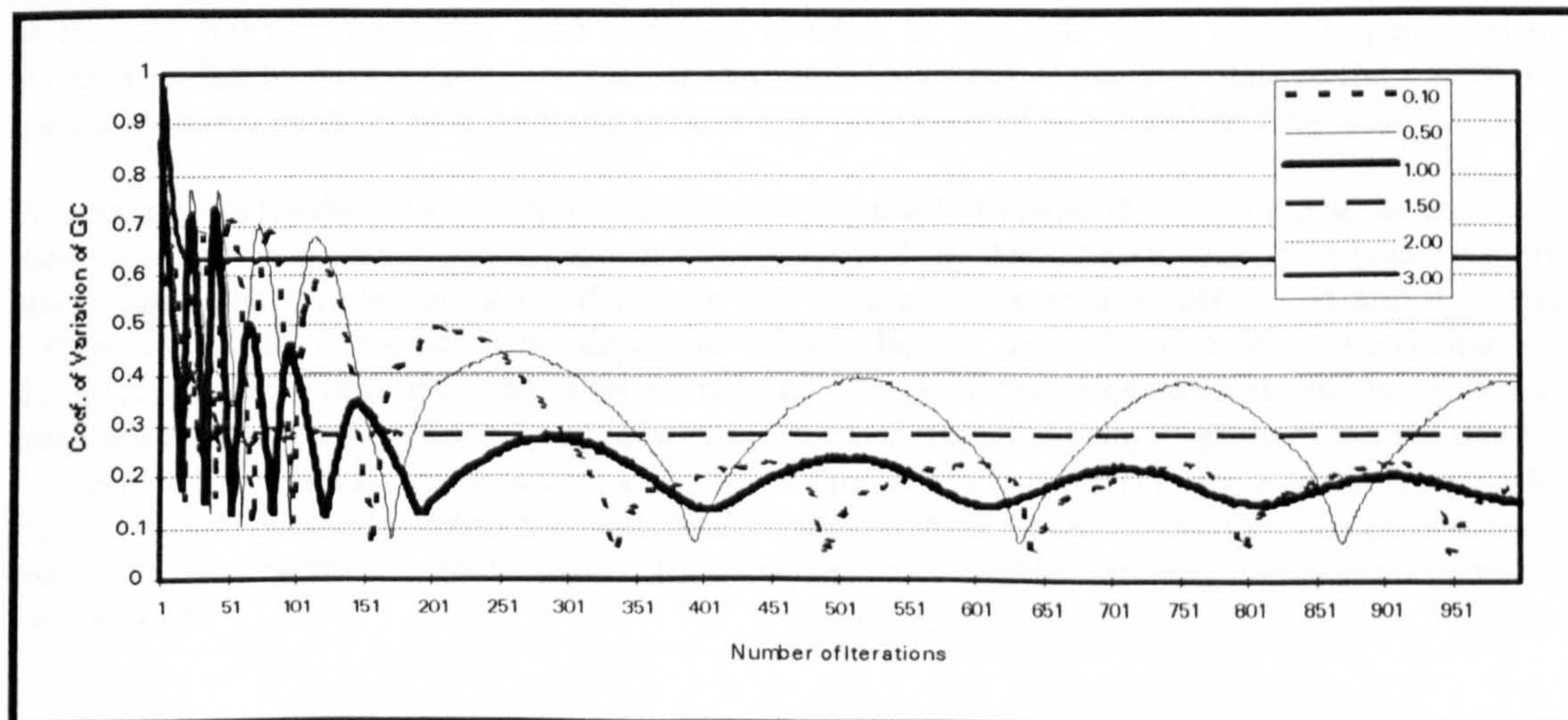


Figure (5.5): The fluctuation of the coefficient of variation of Generalized Cost under different values of the Scalar Factor ' η ' (Stochastic Model).

In conclusion, the system under the stochastic departure time choices process keeps oscillating without reaching a stationary state, and $\eta=1$ with $R1^*$ exhibit the lowest coefficient of variation of this oscillation. Therefore, these two values are used in producing all the subsequent results for the stochastic departure time choices model. However, the value of η equal to unity is the empirical value that has the implications of closely representing the actual travellers' behaviour when logit model is adopted for departure time choices (see Ben-Akiva et al. (1986a, b), Mahmassani et al. (1984) and Mahmassani and Chang (1986)).

5.3.3 Base case departure time choices and Generalized Cost results.

The results, which correspond to the base case when the system has reached the final state starting from the Pseudo-stationary state and using the deterministic departure time choices model are illustrated in Figure (5.6). These results represent the average of the last 25 iterations after 524 iterations in total. The figure shows the stationary distribution of departure rates, generalized cost and its components (i.e. travel time and delay cost and schedule delay cost). The following key properties are depicted from Figure (5.6):

- The intervals of the beginning and the end of the congestion period are given by [7:10 a.m.–7:15 a.m.] and [9:35 a.m.–9:40 a.m.] respectively. In other words, the first and the last departures from the bottleneck (i.e. T_0 and T) occur at 7:10 a.m. and 9:40 a.m. respectively.
- The period of on-time arrival $[\tilde{t}, \tilde{t}]$ is [7:50 a.m.–8:20 a.m.], implying that the period of late arrival begins at $\tilde{t} = 8:20$ a.m. before the end of congestion, $T=9:40$ a.m. It has been demonstrated that this is always true for a single route by de Palma et al. (1983) and for two routes by Ben-Akiva et al (1986a).
- The distribution of departure times is continuous but does not have continuous derivatives. The abrupt changes in slope occur because the utility function (given by Equation (4.12) in the previous chapter) has different left and right derivatives at $\tilde{t}$ and $\tilde{t}$.
- The figure also shows that the travel cost is (approximately) the same for all travellers, meaning that Wardrop's user equilibrium is held. This is achieved by trading-off the two components of the generalized cost (travel time and delay cost and schedule delay cost) such that their sum is constant for any departure time between T_0 and T .

Similarly, Figure(5.7) shows the distribution of the departure rates, generalized cost and generalized cost components using the stochastic departure time choices model. The results represent the average of the last 100 iterations after 1000 iterations in total. In this case, since the costs perceived by travellers differ by random amounts, the travel costs for all users at the final state of the system are not equal (they also differ by a random amount with a coefficient of variation less than 0.2).

Comparing this results with the deterministic one as depicted in Figure (5.8) shows that the departure time period given by the stochastic model extends beyond the deterministic one from both sides. In other words, the distribution of the departure rate is spreading over a longer period and it is also shifted to the right with a maximum departure rate less than the deterministic one. As a consequence, the perceived travel time and delay costs are reduced while both the perceived early and late schedule delay costs are increased. Since the shadow value of travel time and delay is greater than the that for the early schedule delay and less than that for the late schedule delay (i.e. $\beta > \alpha < \beta\gamma$), then the perceived costs during the periods of early and on-time arrivals would be less than the deterministic costs, while during the late arrival period the perceived costs would be slightly higher (as depicted in Figure (5.8)).

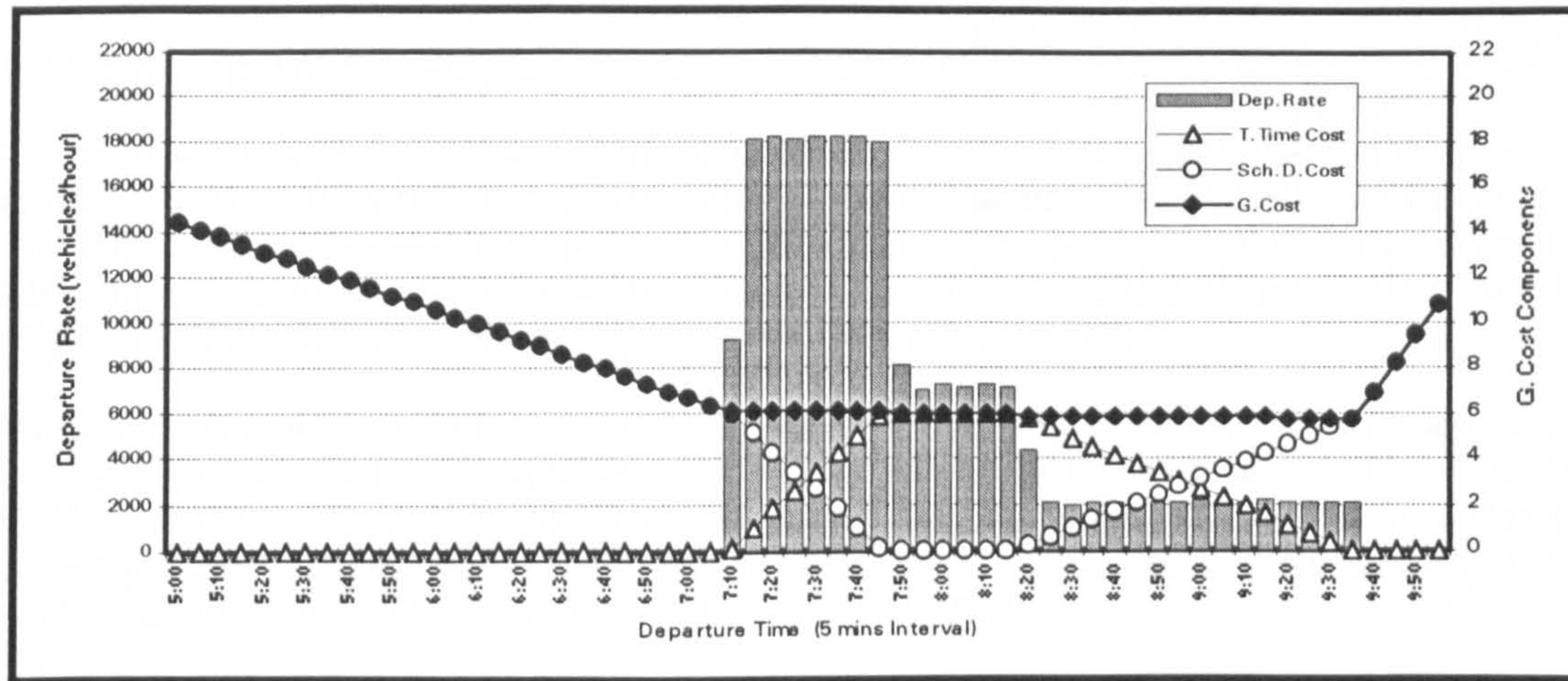


Figure (5.6): The Departure Pattern vs. the G. Cost and the GC Components under the Deterministic Model (No RP).

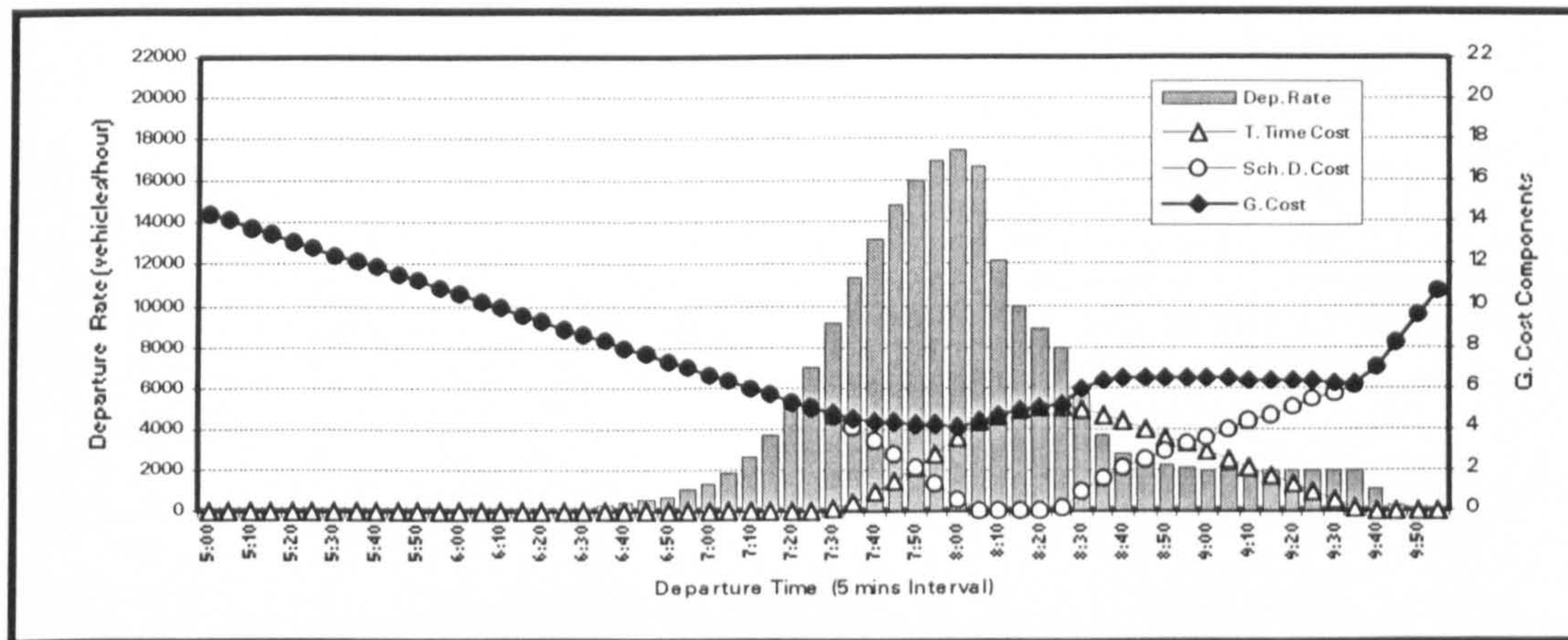


Figure (5.7): The Departure Pattern vs. the G. Cost and the GC Components under the Stochastic Model (No RP).

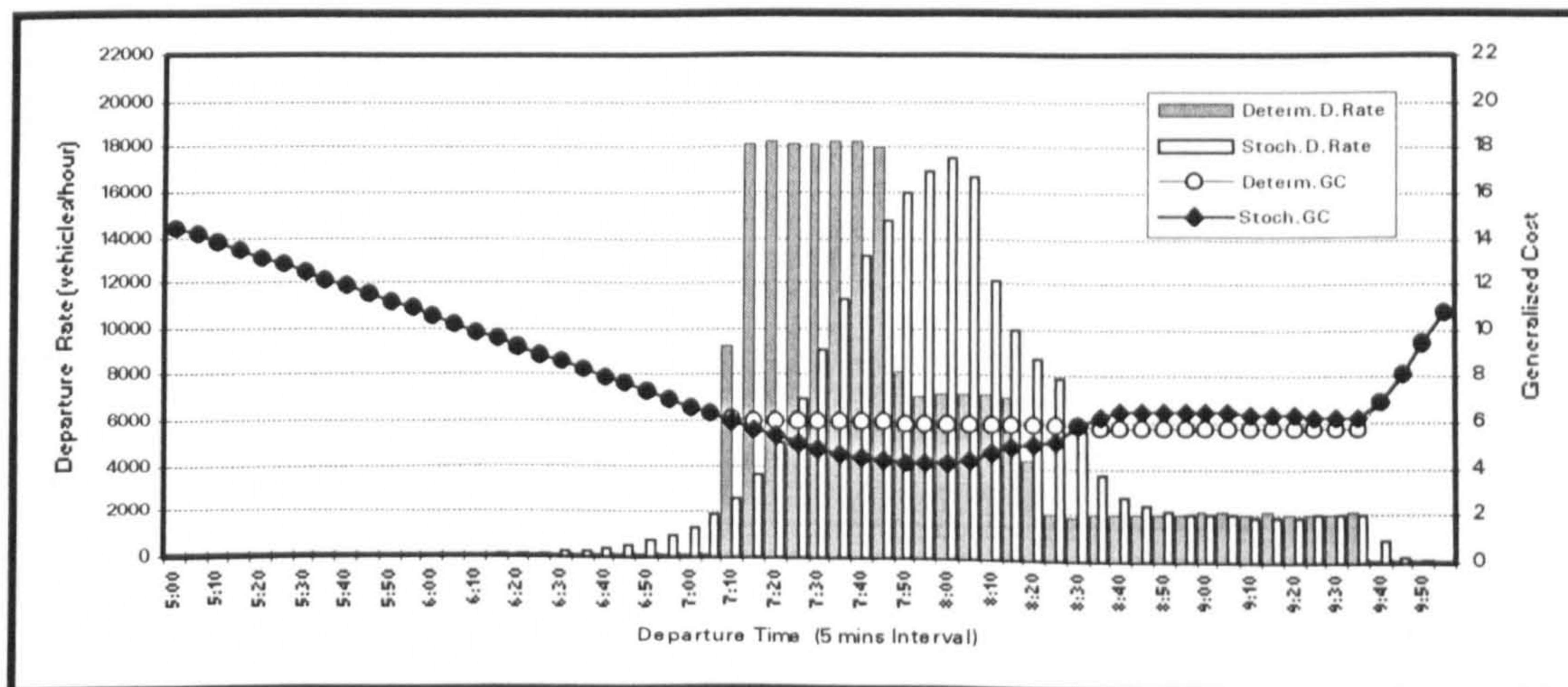


Figure (5.8): The Departure Pattern vs. the G. Cost under the Deterministic and the Stochastic Models (No RP).

5.3.4 The different impacts of *TDRP* on the base case results.

This section discusses the various impacts of *TDRP* on the base case results under the deterministic and the stochastic departure time choices models. The two approaches for modelling the original *TDRP* function (as discussed in the previous chapter) are considered throughout: i.e. by modifying the *TDRP* to include the schedule delay costs imposed on other vehicles "called Modified *TDRP*" and by adjusting *TDRP* using the day-to-day adjustment process "called *TDRP*".

The discussion in this section is based on the results of the second approach since the first approach is proved to be equivalent to the optimal toll. However, the results of the "modified *TDRP*" is also highlighted throughout the discussion to demonstrate this conclusion. The reviewing factor for adjusting *TDRP*, R_2 in Equation (4.79) in the previous chapter, is set to 2 per cent in all the subsequent simulation results with *TDRP* and the modified *TDRP*.

The impacts of *TDRP* on the following are considered each in turn:

- a.* Convergence and stability of the results;
- b.* Generalized cost and departure time pattern;
- c.* Queuing pattern; and
- d.* *TDRP* pattern at equilibrium.

a. Convergence and stability of the results

The fluctuation of the coefficient of variation of the generalized cost before and after **TDRP** and the modified **TDRP** for the deterministic and the stochastic models are illustrated in Figures (5.9) and (5.10) respectively. The results are produced for a total number of iterations equal to 500. The two figures exhibit a very significant improvement in the convergence and stability of the results as a result of imposing the **TDRP**. The system reaches its stationary state under the stochastic model after only 50 iterations (compared with 1000 before **TDRP**) with an average coefficient of variation equal to 0.1. In the deterministic model, the system reaches its stationary state with little oscillations after 140 iterations (compared with 524 before **TDRP**) and the average coefficient of variation is 0.05.

The two figures also exhibit that the convergence and stability of the results under **TDRP** and the modified **TDRP** are quite similar.

Different values of R_2 are also examined during the simulation (for example 5 and 10 per cent) and they also lead to a similar convergence with slightly higher oscillations.

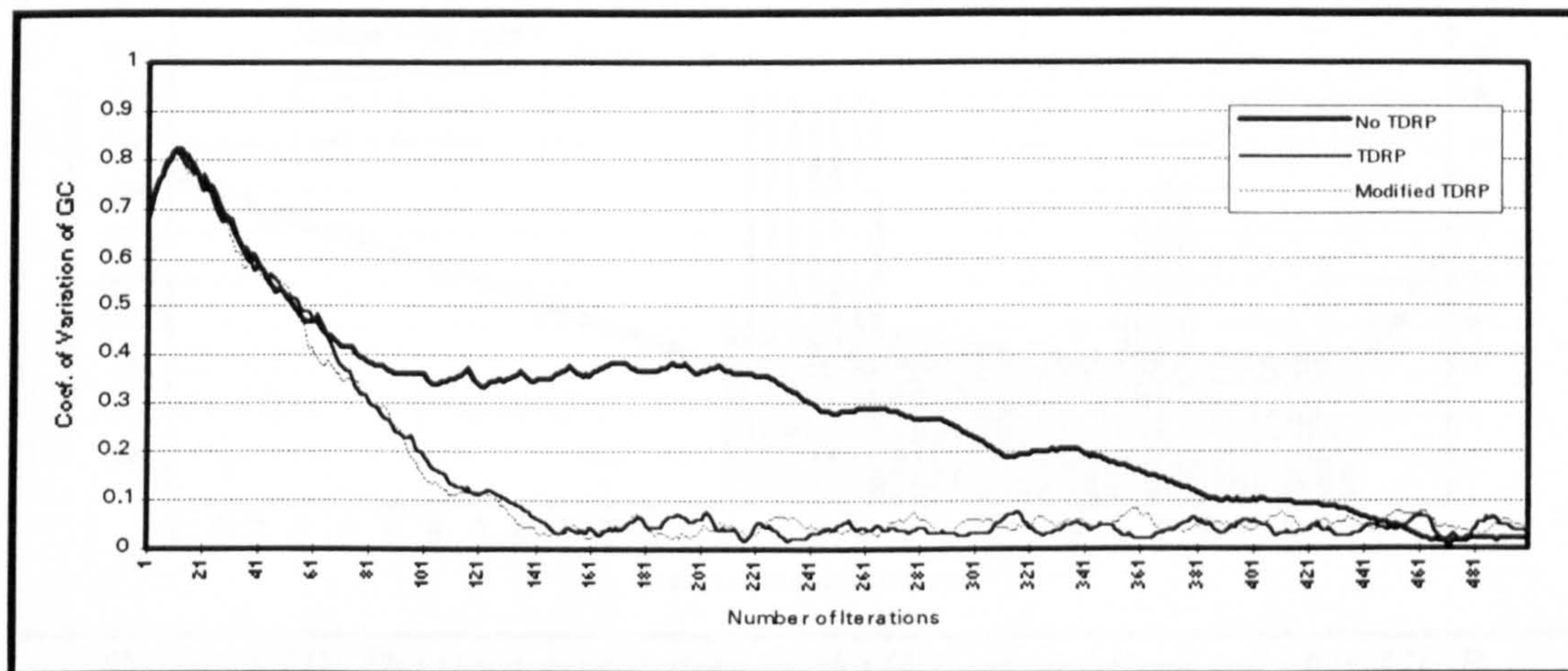


Figure (5.9): The fluctuation of the coefficient of variation of Generalized Cost before and after **TDRP** and the Modified **TDRP** (Deterministic Model).

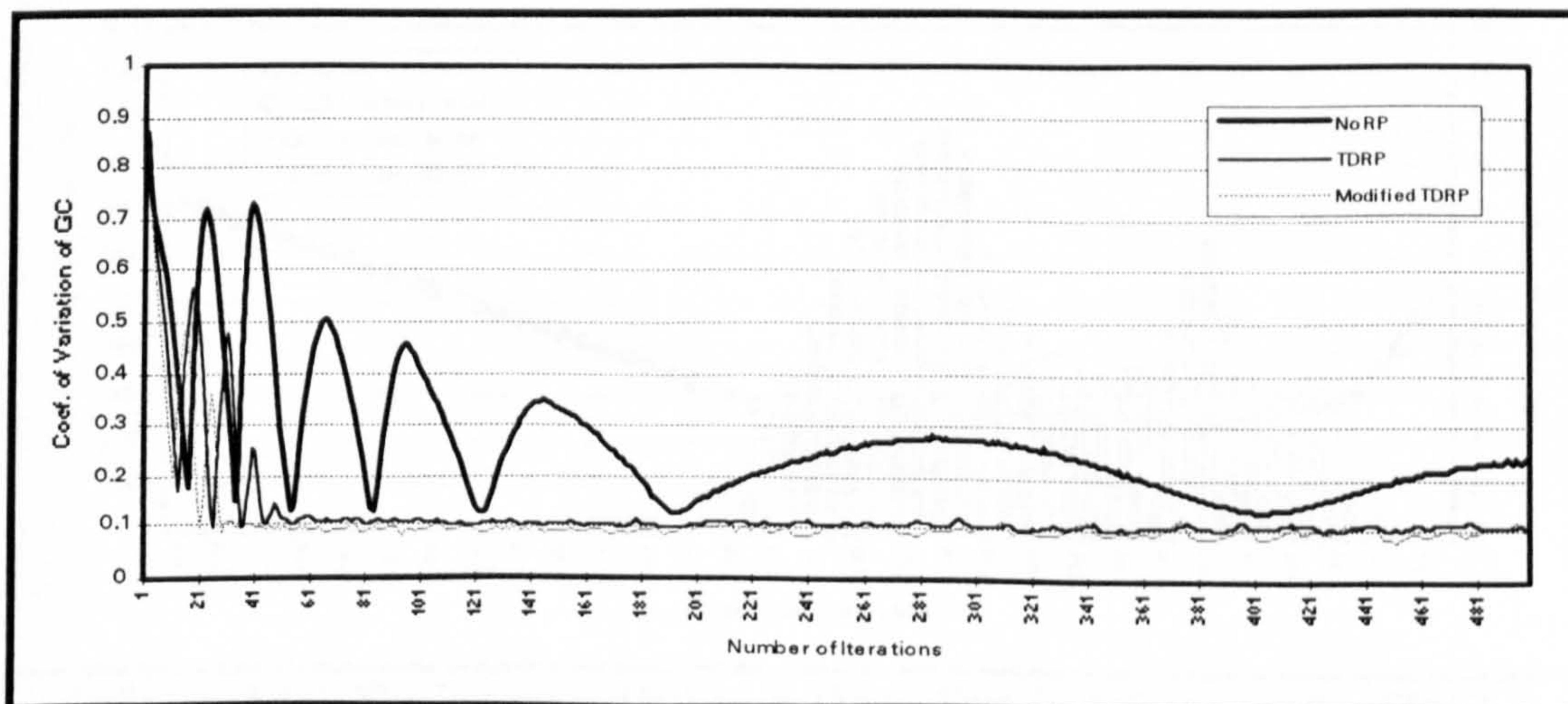


Figure (5.10): The fluctuation of the coefficient of variation of Generalized Cost before and after **TDRP** and the Modified **TDRP** (Stochastic Model).

b. Generalized cost and departure time pattern under *TDRP*

Before discussing the different impacts of *TDRP* on the departure pattern and the generalized cost, it has to be pointed out that the *TDRP* (adjusted by the day-to-day adjustment process) cannot eliminate queuing. This is because *TDRP* is a function of the queue duration (i.e. the starting and vanishing times of the queue, but not the queue length) and if the queue were eliminated the value of toll would be zero and this consequently leads to the same equilibrium as the one before *TDRP*.

Figures (5.11), (5.12) and (5.13) show the departure pattern versus the generalized cost under the deterministic, stochastic, and both deterministic and stochastic departure time choices models respectively. Generally speaking, the average user cost after *TDRP* is slightly higher than before, which implies that the total user cost (including tolls) increases by imposing the *TDRP*. These three figures also illustrate that the starting and ending times of the congested period do not change after *TDRP*, i.e. there is no spreading for the congestion period as a result of imposing *TDRP*. Furthermore, the maximum departure rate is shifted to the right over a shorter time period and its value is less than the maximum departure rate before the toll. This pattern results in a very substantial

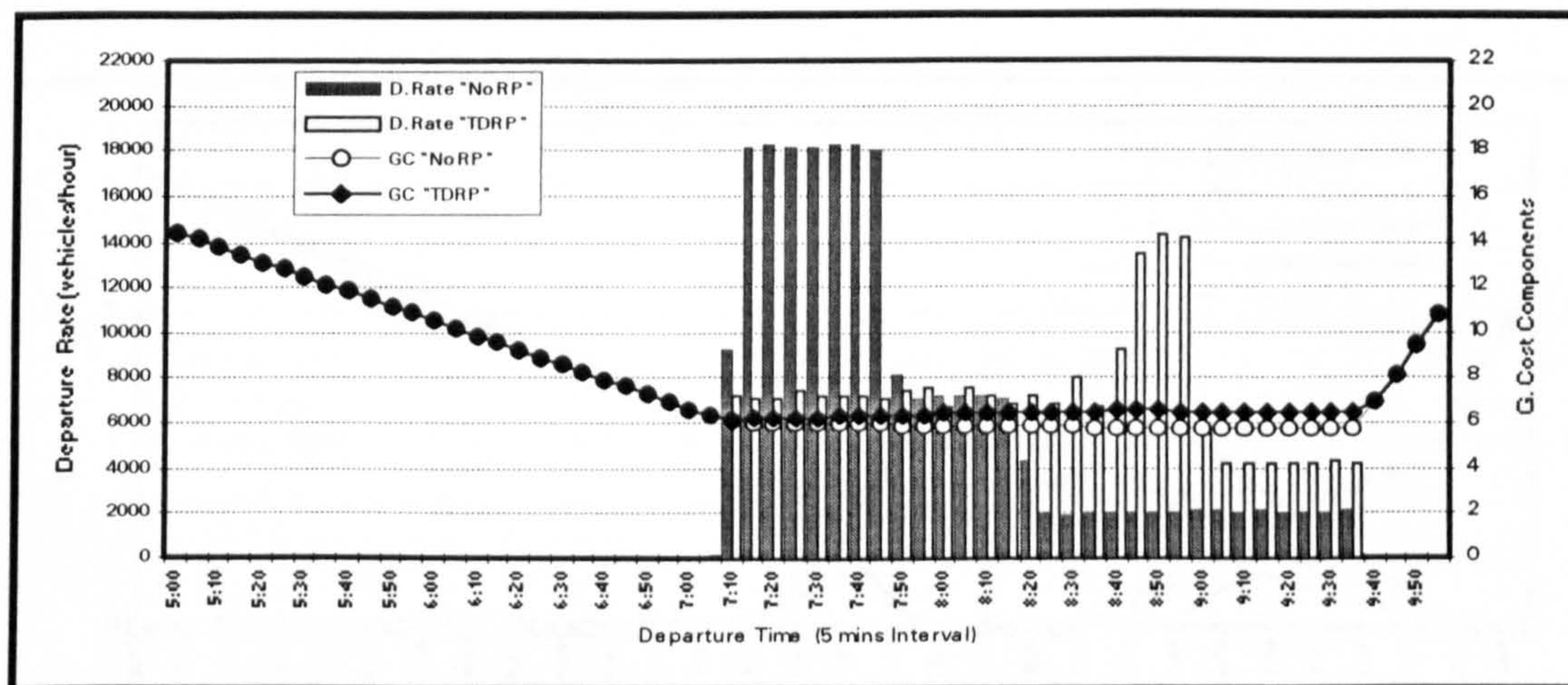


Figure (5.11): The Departure Pattern vs. the G. Cost for before and after *TDRP* (Deterministic Model).

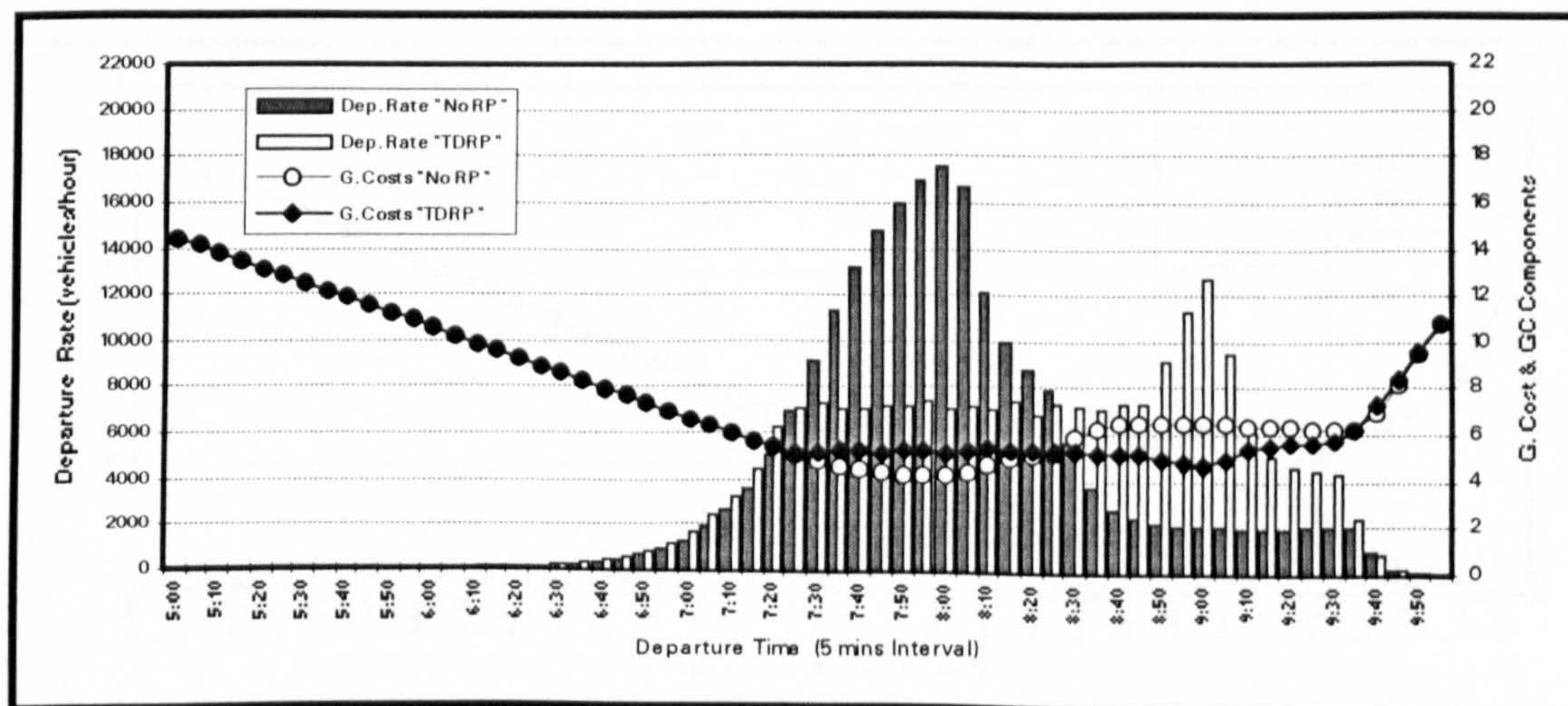


Figure (5.12): The Departure Pattern vs. the G. Cost for before and after *TDRP* (Stochastic Model).

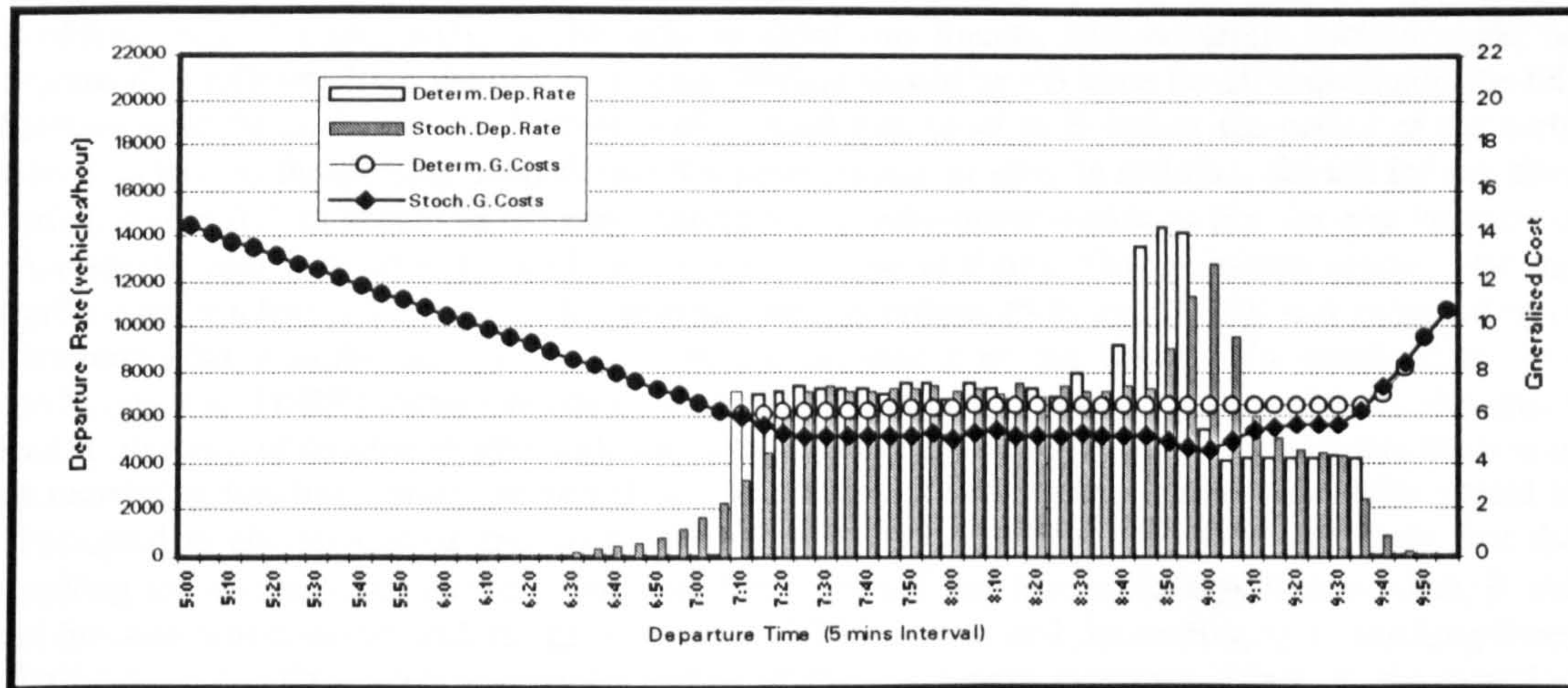


Figure (5.13): The Departure Pattern vs. the G. Cost for Deterministic and Stochastic Models (with TDRP).

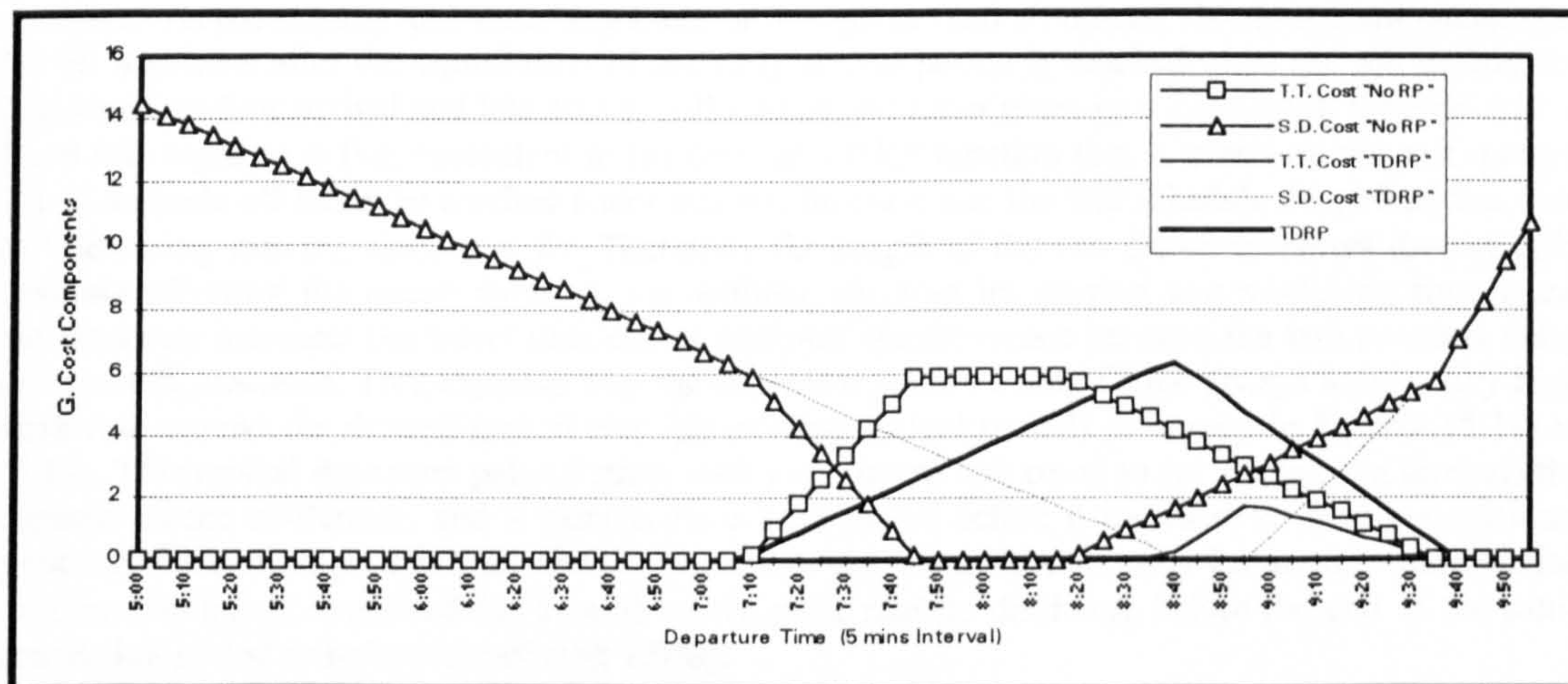


Figure (5.14): The Generalized Cost Components for before and after TDRP (Deterministic Model).

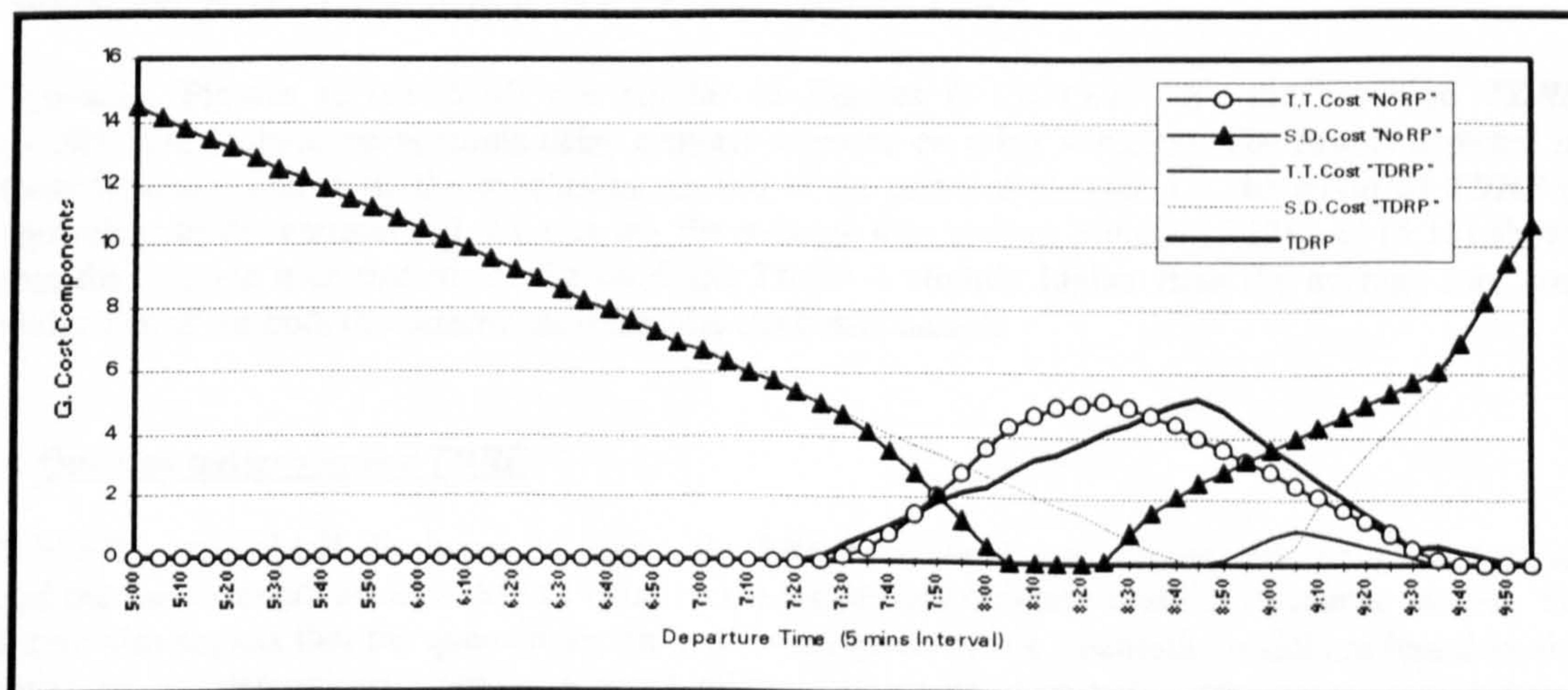


Figure (5.15): The Generalized Cost Components for before and after TDRP (Stochastic Model).

improvement in the travel time as depicted in Figures (5.14) and (5.15) for the deterministic and the stochastic models respectively. In the light of these two figures, this departure pattern could be explained as follows. Since the user cost at equilibrium should be the same for all individuals, the toll function must be an increasing function with a fixed rate equal to β during the period of the early arrival as long as the queuing delay during this period is almost zero. In addition, the toll for any time period at equilibrium should be the same during all the subsequent iterations (i.e. for any iteration ω after equilibrium $\tau(t, \omega) = \tau(t, \omega+1) = \tau(t, \omega+2) \dots$ and so forth). This condition assures that the equilibrium reached is a stable one (as demonstrated in Figures (5.9) and (5.10)) and progressing in iterations after equilibrium will not disrupt the influences of the results. To satisfy these two conditions, the *TDRP* function which is originally a decreasing function independent of the queue length (in terms of number of vehicles), should be changed by the day-to-day readjustment process to an increasing function during the period of early arrival. Therefore, the queue during this period is interrupted to shorter queues and every queue should be longer than its proceeding such that the resulting toll rate becomes an increasing one. When the toll rate reaches its equilibrium value, β , the toll function will trade-off with the early schedule delay function and the starting and vanishing times of all queues should remain unchanged during all the subsequent iterations. However the resulting queues should have a very short length such that they have no (or very little) influence on travel costs during the period of early arrival.

After the end of the early arrival period the toll will be given by the last (and longest) queue in that sequence. As the starting and vanishing times of this queue (and all others) should remain unchanged for all iterations after the equilibrium of the early arrival period is reached, then the toll during the periods of on-time arrival and late arrival will also be fixed and given by a decreasing function with a fixed rate equal to α (i.e. equivalent to the original *TDRP* function that is based on the last queue). Thus, no trade-off could be reached under this toll function and the late schedule delay function with an increasing rate $\beta\gamma$, since $\alpha < \beta\gamma$. Therefore, the length of the last queue increases dramatically (without affecting the queue duration, i.e. without affecting its starting and vanishing times) and consequently increases the travel time cost to trade-off the difference between the late schedule delay and the toll functions. This explains why the departure pattern under *TDRP* always have a very high peak rate around the desired arrival time (or on-time arrival period) as shown in Figures (5.11) to (5.13). This typical departure pattern starts with a departure rate equal to (or fluctuating around) the capacity of the bottleneck, and it extends for a long period before it increases to a very significant peak rate for a short period. Then the departure rate goes below the capacity for another short period to allow the queue developed (as a result of the peak rate) to discharge before the end of the same congestion period as before introducing *TDRP*.

In addition, Figures (5.14) and (5.15) illustrate that re-timing of journeys to take this typical pattern results in shifting (to the right) and shortening the period $[\tilde{t}, \bar{t}]$ and consequently increases the early schedule delay and its time period, and decreases the late schedule delay and its time period since the earliest and the latest departures remain the same as before *TDRP*.

Typically, Figures (5.16)–(5.18) are similar to Figures (5.13)–(5.15) for the modified *TDRP* (modified to include the schedule delay changes imposed on other vehicles). The results depicted in these figures demonstrate the conclusion reached in the previous chapter: i.e. the modified *TDRP* is equivalent to the optimal toll. Comparing the average user cost in Figures (5.13) and (5.16) shows that the average user cost under the modified *TDRP* is slightly higher than the average user cost under *TDRP* for both the deterministic and the stochastic models.

c. Queuing pattern under *TDRP*

While the optimal toll eliminates the queue, the *TDRP* does not but it reduces both the queue duration and the queue length as depicted in Figure (5.19) for the deterministic and the stochastic models. The figure also depicts that the queue duration and length given by the stochastic model are less than that given by the deterministic one as a result of the random error in travel cost perceptions between travellers. The queue depicted in this figure after *TDRP* starts at the end of the early arrival period and disappears at the same time as the queue before *TDRP*. Furthermore, the queue developed does not

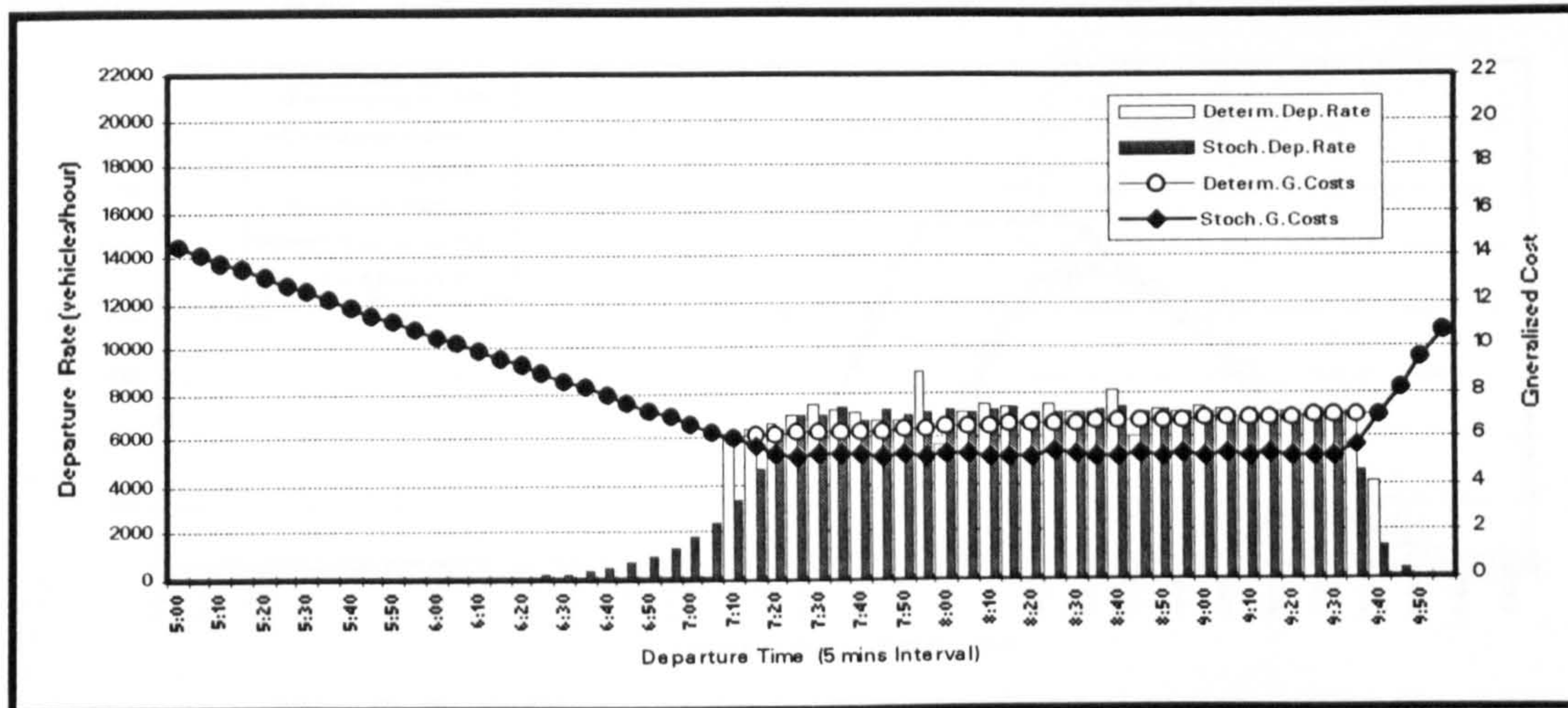


Figure (5.16): The Departure Pattern vs. the G. Cost for Deterministic and Stochastic Models (with the Modified TDRP).

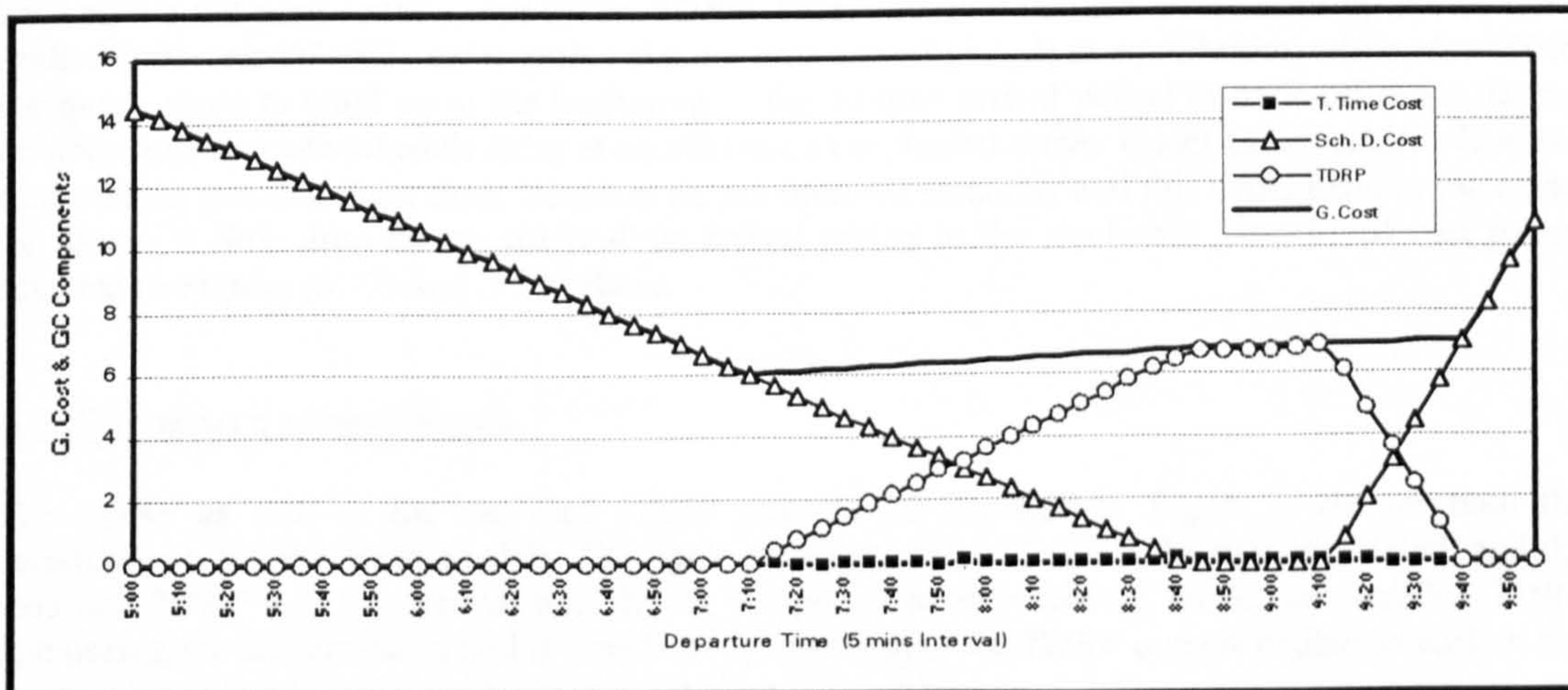


Figure (5.17): The Generalized Cost Components after the Modified TDRP (Deterministic Model).

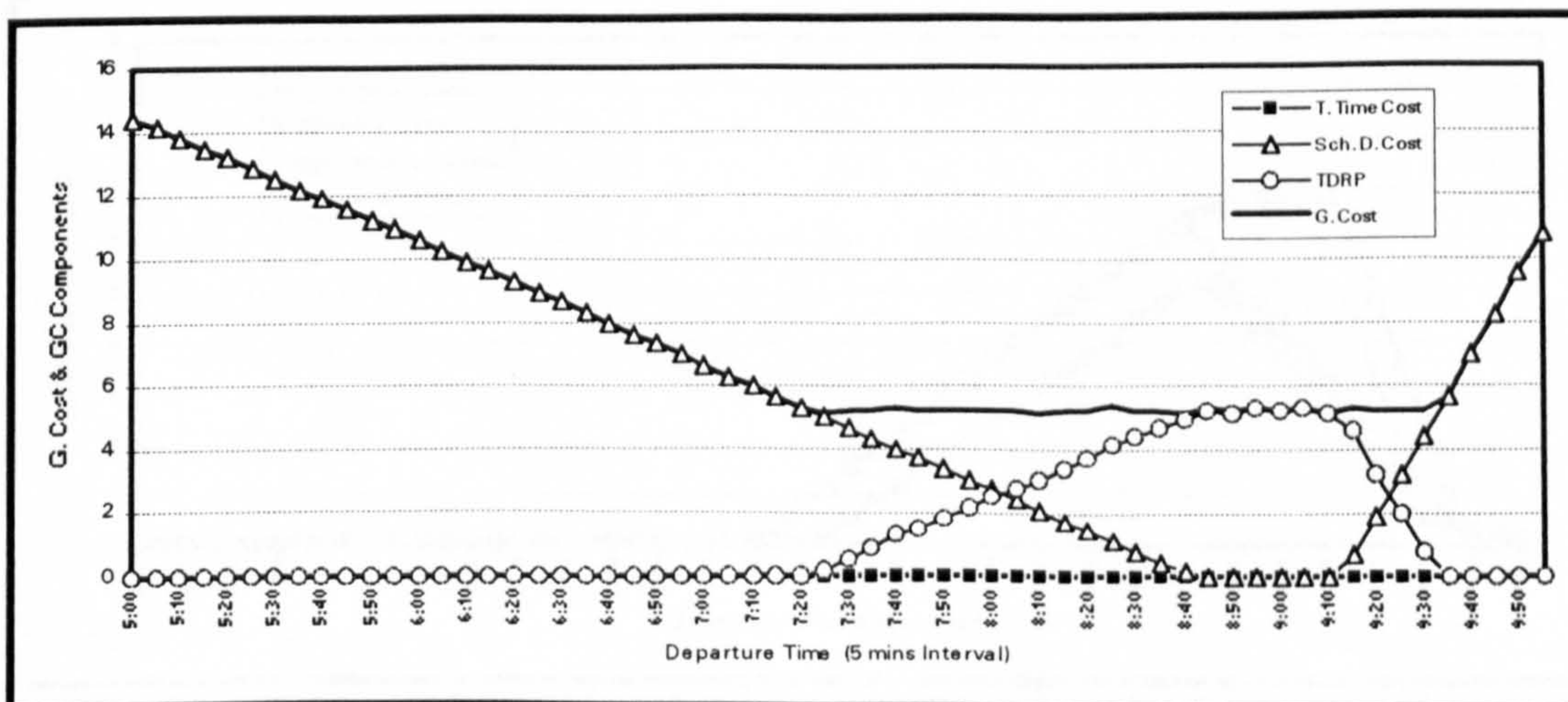


Figure (5.18): The Generalized Cost Components after the Modified TDRP (Stochastic Model).

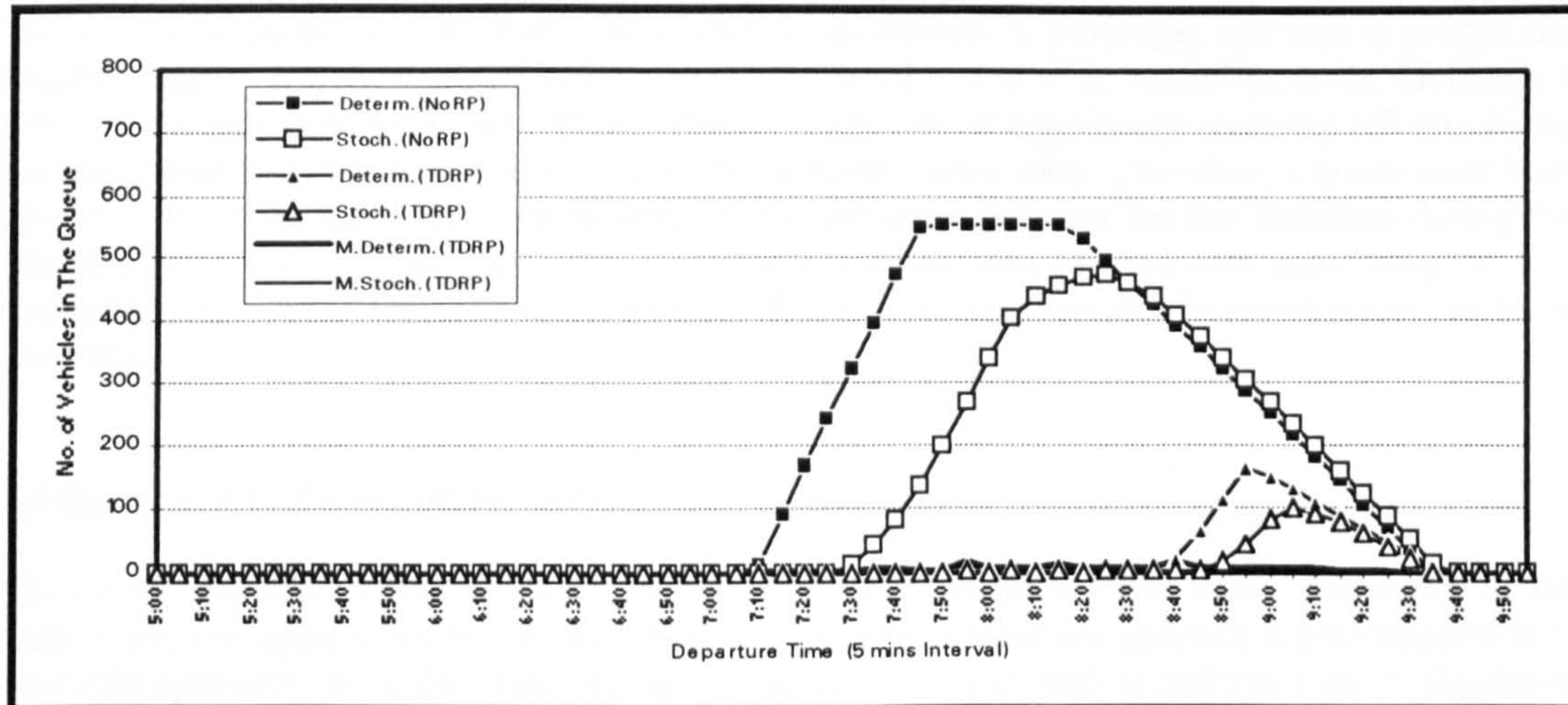


Figure (5.19): The Queue duration and Queue length before and after TDRP and the Modified TDRP for the Deterministic and the Stochastic Models.

have such a flat peak as the queue before *TDRP*. This is because this flat peak builds up before toll to trade-off the zero schedule delay period (i.e. on-time arrival period) at equilibrium, while after *TDRP* the queue starts to build up at the beginning of the on-time arrival period to trade-off the difference between tolls and late schedule delay at equilibrium as explained above. Under the stochastic departure time choices process, these three functions do not trade-off smoothly and this could explain the reason for having a little drop during the on-time arrival period in the stochastic generalized cost pattern depicted in Figures (5.12) and (5.13) above.

d. *TDRP* pattern at equilibrium.

The *TDRP* as well as the modified *TDRP* patterns are depicted in Figure (5.20) for both the deterministic and stochastic models. The figure shows that the toll rate is the same for *TDRP* and the modified *TDRP* (i.e. the optimal toll) during the period of early arrival, while the modified *TDRP* rate during the late arrival period is significantly higher than the *TDRP* one (as explained earlier, the former is equal to $\beta\gamma$ while the latter is equal to α).

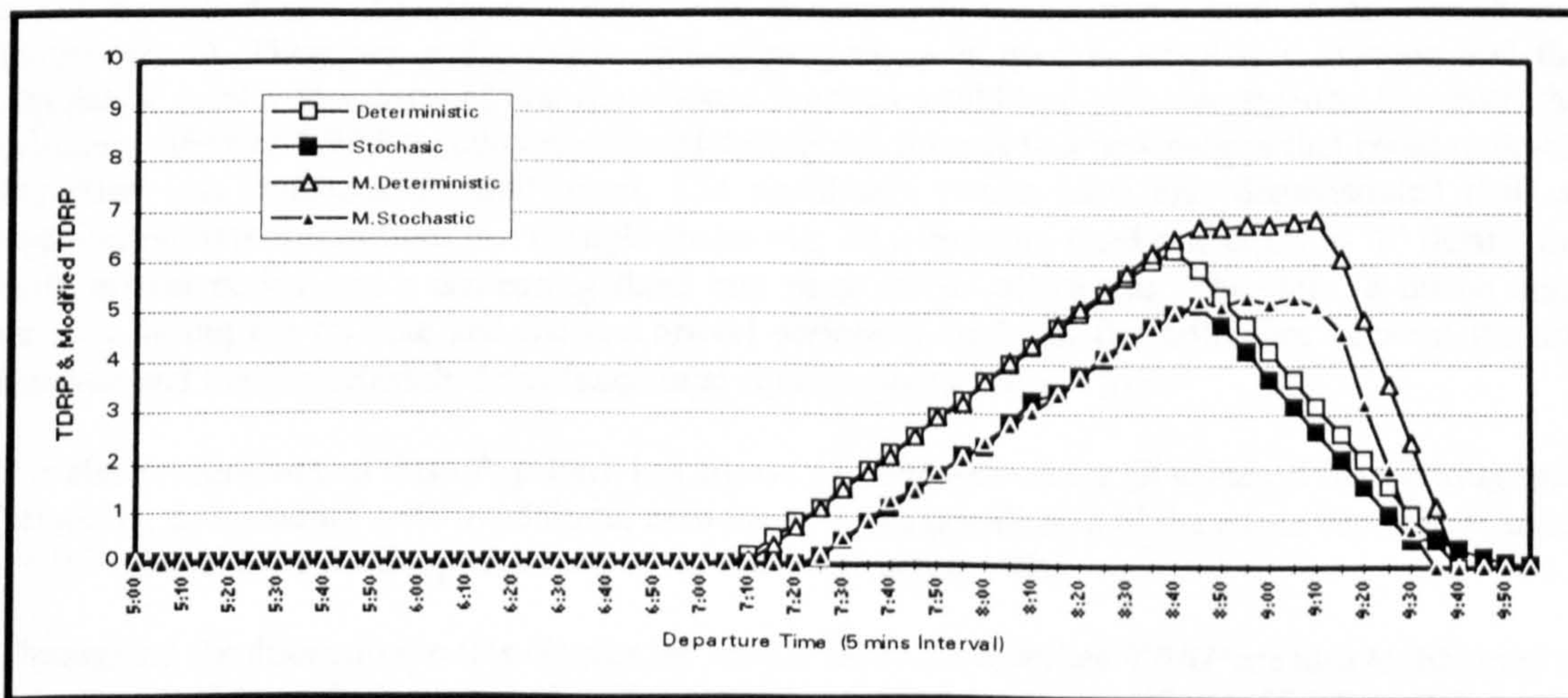


Figure (5.20): The *TDRP* patterns for the Deterministic and the Stochastic Models at equilibrium.

The typical shape of the toll pattern for *TDRP* at equilibrium is a triangle and this is always true whether there is a work time flexibility or not (i.e. for all values of Δ , including $\Delta=0$). However, if $\Delta=0$, the toll pattern will not trade-off the schedule delay costs at equilibrium since the toll rate during the late arrival period is less than the rate of late schedule delay costs. Therefore, a queue must build up and increase the travel time cost to trade-off the difference between the two functions during this period (exactly the same as with $\Delta \neq 0$). This conclusion has been verified with simulating the base case with $\Delta=0$, and the results demonstrated that *TDRP* does not eliminate the queue under any travel conditions.

5.4 Summary and concluding remarks

The chapter describes a framework for the simulation solution for a single traffic bottleneck. In this framework, two approaches for the departure time choices process are adopted: a deterministic (and heuristic) approach and a stochastic approach. This framework is used to perform a set of simulation experiments (using a numerical example) to examine the convergence and stability of the results before and after *TDRP*. Then it is used to investigate the different impacts of *TDRP* on the departure time pattern and the user generalized cost and its components (travel time and delay cost and schedule delay cost). The queue distribution under *TDRP* and the *TDRP* pattern at equilibrium are also investigated. Throughout the presentation, both the deterministic and the stochastic models results are considered for *TDRP* and the modified *TDRP*.

The results demonstrated that the convergence and stability of the results are significantly improved under *TDRP*. It is also concluded that the departure time pattern under *TDRP* has a typical shape starting with a departure rate equal to (or fluctuating around) the bottleneck capacity during the period of early arrival, then the departure rate increases dramatically for a short period before it goes below the bottleneck capacity sometimes before the end of the congestion period. Furthermore, this departure pattern takes place during the same period as before introducing *TDRP*, i.e. the congestion period does not spread over a longer period as a result of *TDRP*. An explanation of building up such a departure time pattern is also given.

It is concluded that although *TDRP* does not eliminate queuing, it generates some efficiency gains as a result of altering the frequency of distribution of the departure times to build up the typical departure pattern. For example, it results in a substantial reduction in queuing delay and consequently travel time costs. It also leads to shifting and shortening the period $[\tilde{t}, \tilde{t}^*]$ and consequently increases the early schedule delay and its time period and decreasing the late schedule delay and its time period.

Since the original *TDRP* function (with a wedge-shape and a minus 45° slope) can not trade-off the travel time cost and the schedule delay cost functions, it would never lead to an equilibrium (convergence). Therefore, a day-to-day adjustment process is used to adapt this function and the simulation results demonstrated that the adjusted function would lead to a convergence. However, the original pattern of *TDRP* function is changed during this process to a new pattern that could trade-off the other cost functions at equilibrium. The simulation results have also demonstrated that, at equilibrium, this new pattern is a triangle-shape with an increasing fixed rate equal to ' β ' during the early arrival period and a decreasing fixed rate equal to ' α ' afterwards. Therefore, a queue must develop during the on-time and the late arrival periods to trade-off the difference between the toll function and the late schedule delay function at equilibrium (since $\alpha < \beta\gamma$).

It is also recognised that this toll pattern has always a triangle-shape for all values of the working time flexibility, Δ , including $\Delta=0$. In addition, even for $\Delta>0$, the distribution of the queue developed under *TDRP* will not have the flat peak observed for the queue before tolls.

Throughout the discussion in this chapter the results under the modified *TDRP* are also highlighted to demonstrate the conclusion reached in the previous chapter: i.e. by modifying *TDRP* to include the schedule delay changes to other vehicles, it leads to the system optimal.

CHAPTER SIX

6. Network Traffic Simulation Model

6.1 Introduction

6.2 General description of the simulation model

6.3 Principles of the model

- 6.3.1 Overall Structure.
- 6.3.2 Time Variation/Time-Intervals.
- 6.3.3 Packets/Vehicles Generations.
- 6.3.4 Link Performance.
 - 6.3.4.1 Link Flow-Density Relationship.
 - 6.3.4.2 Car-Flow Behaviour.
- 6.3.5 Route Choice Model (Dial's Model).
- 6.3.6 Propagation of Packets/Vehicles from one link to another.
- 6.3.7 Model input/output.
- 6.3.8 Other features of the simulation model.

6.4 Expanding the original model to include *TDRP*

- 6.4.1 Overall Structure of the expanded model.
- 6.4.2 Conditions for building up queues.
- 6.4.3 Time-Dependent Road Pricing Model.
- 6.4.4 Generalized Cost Function.
- 6.4.5 Departure time choices model.
- 6.4.6 Checking for convergence.
- 6.4.7 Final output.

6.5 Interaction between nodes and its effect on the value of *TDRP*

- 6.5.1 Interaction between nodes (case with no merging/no diverging flow).
- 6.5.2 Introducing merging and diverging flows at nodes.
- 6.5.3 Procedures for adapting the charges at each individual node.
- 6.5.4 Suggested algorithm for a typical network.
- 6.5.5 Summary and implications for practical applications.

6.1 Introduction

This chapter describes the overall structure of the network traffic simulation model and outlines its main principles. It also sheds some light on how this model is expanded to include the departure time choice model and the time varying toll function (*TDRP*).

Perhaps the most striking merit of applying *TDRP* for a traffic network is the issue of node-interaction and its effect on the value of *TDRP* incurred by individual vehicles from one node to another. Therefore, this issue is investigated in this chapter under different circumstances and a general algorithm for solution is suggested. In addition, the implications of considering the interactions between nodes in practical applications is also addressed.

6.2 General description of the simulation model

The simulation model used in this study was developed by Yoshii and Kuwahara (1993). The model is basically developed to simulate congested traffic flow on urban expressway network (as all nodes are representing either merging or diverging nodes, i.e. no intersections are allowed between roads). It incorporates drivers' route choice behaviour and approximately reproduces time-dependent traffic conditions under time varying flow for each Origin-Destination movement. The bases of the model are the vehicle motion (vehicle assignment) model and the route choice model which are implemented alternatively as depicted in Figure (6.1) below.

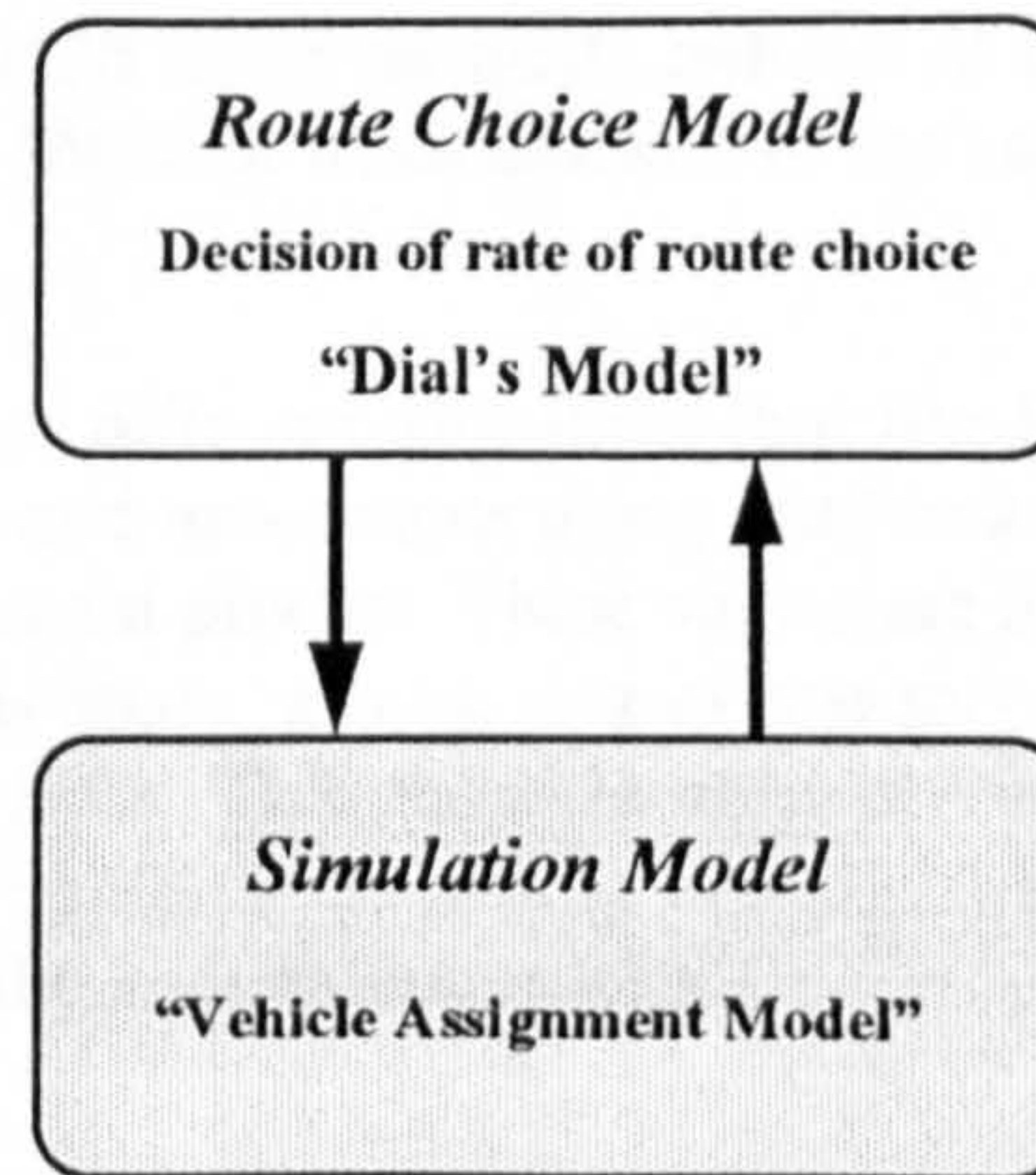


Figure (6.1): Route choice model and simulation model.

The vehicle motion is modelled based on the car-following theory in which vehicles follow each other on a highway (without passing), so that the character of total vehicular flow (i.e. the propagation distribution down a line of vehicles) is reproduced precisely. On the other hand, the route choice model is given by Dial's (1971) Model, in which traffic would not all flow on the minimum cost (or minimum travel time) route, but would be distributed among several reasonable routes connecting an origin and a destination. Thus, route choice is assumed to differ between individuals in a probabilistic way, being dependent on (variable) network attributes and personal variables. Therefore, the model assigns vehicles under the stochastic user equilibrium principles (see Vythoulkas (1991)).

The simulation model (including route choice model) was validated by the authors using a simple traffic network and they demonstrated that the results would reasonably satisfy the stochastic user equilibrium principles.

The next section outlines the overall structure of the model and describes its main principles.

6.3 Principles of the model

6.3.1 Overall Structure

The overall structure of the model with the different procedures is outlined in Figure (6.2). The main program controls the input/output data and serves to transfer the control from one (subroutine) procedure to another. The main principles of the model are briefly discussed in turn below.

6.3.2 Time variation/Time-intervals

The model has the ability to consider explicitly time variation in traffic conditions throughout the simulation period. This is achieved by subdividing the period to be modelled into a sequence of equal time intervals. These time intervals will also be the same for traffic demand departing from all origins.

Leonard et al. (1989) have pointed out that the selection of suitable time interval lengths is influenced by a number of factors: –

- Time intervals should be sufficiently short to represent the underlying time variations in traffic patterns adequately but not so short, say less than 10 minutes, that they reflect 'noise' in particular sets of flow measurements.
- The intervals should not be so long that average values for flows mask any changes in underlying variation of flow with time.
- In particular, there should be an adequate number of time intervals to reflect changes in flow patterns covering the time when demand in the network nears or exceeds capacity on some links.

The model also has the ability to use different time units (but fixed throughout the simulation period) for calculating vehicles generation and movements along road links and for calculating the minimum path routes and revising the assignment process. These values are defined by the user in the input data file. For all the applications in this study, a time unit of one second is used for calculating vehicles generation and movements along links. This value is normally used in modelling peak periods for urban areas (Leonard and Gower, 1982). Also a time unit of 10 seconds is used for calculating the minimum path route and revising the assignment process.

6.3.3 Packet entry time onto a network (Generation)

The traffic for every individual OD pair is handled in groups called 'packets', each packet consists of an integer number of vehicles (i.e. 1, 2, 5, or 10 vehicles). Essentially, packets are used rather than individual vehicles in order to reduce the computation time. The packets for a given OD pair enter the network at equally spaced intervals of time through each time interval– the number of packets for any OD pair during any time interval is determined by the level of demand flow for this OD pair during this interval and the packet size. The entry times are rounded to the nearest second (or to the nearest time unit used for calculating the vehicle or packet generation as mentioned in section (6.3.2) above). Therefore, the maximum number of packets generated from a certain origin (to all destinations) is 3600 per hour.

6.3.4 Link Performance

Flow, density, and space-mean-speed are of particular interest in describing the performance of vehicular flow along a link. These three variables are therefore defined in Table (6.1).

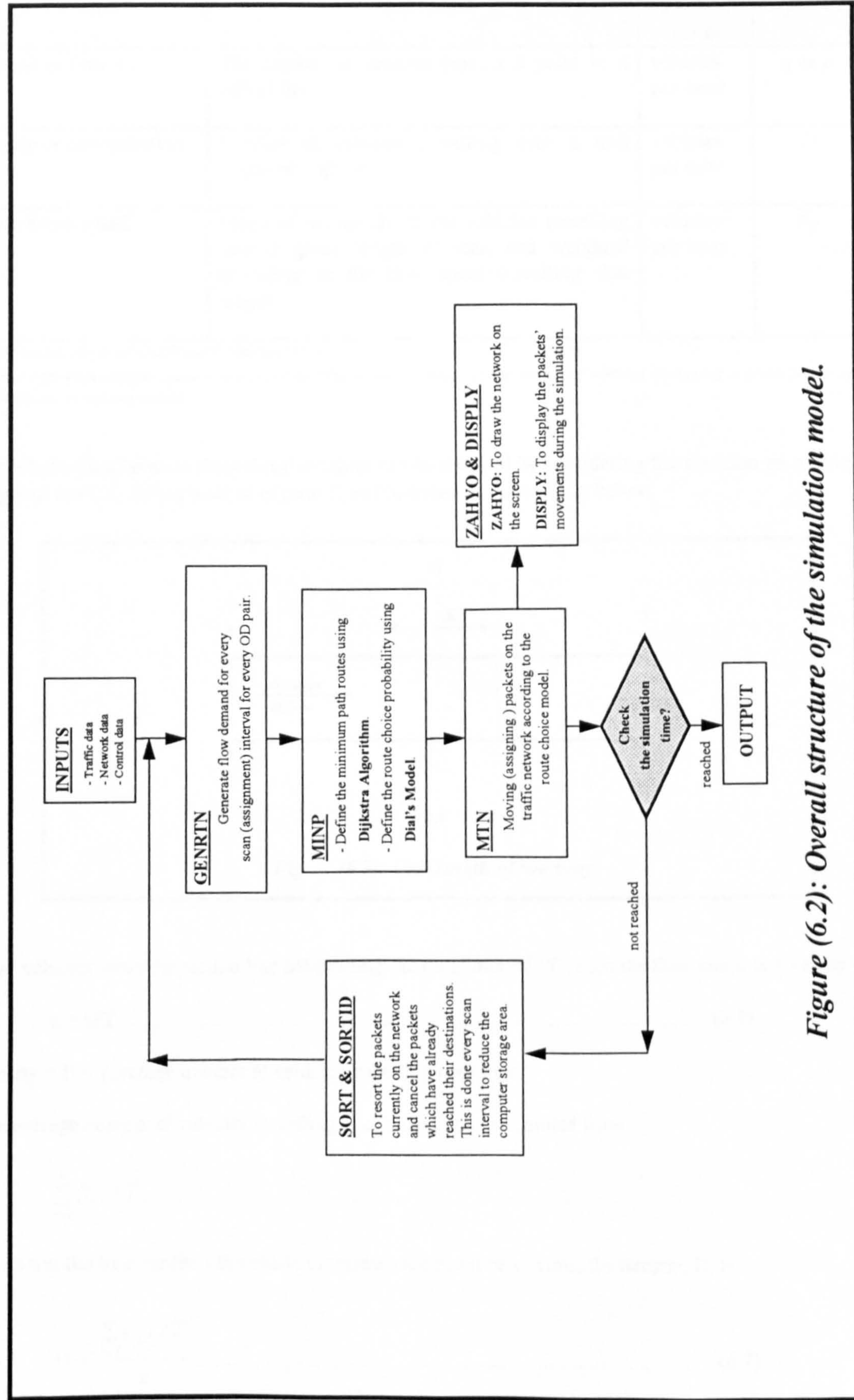


Figure (6.2): Overall structure of the simulation model.

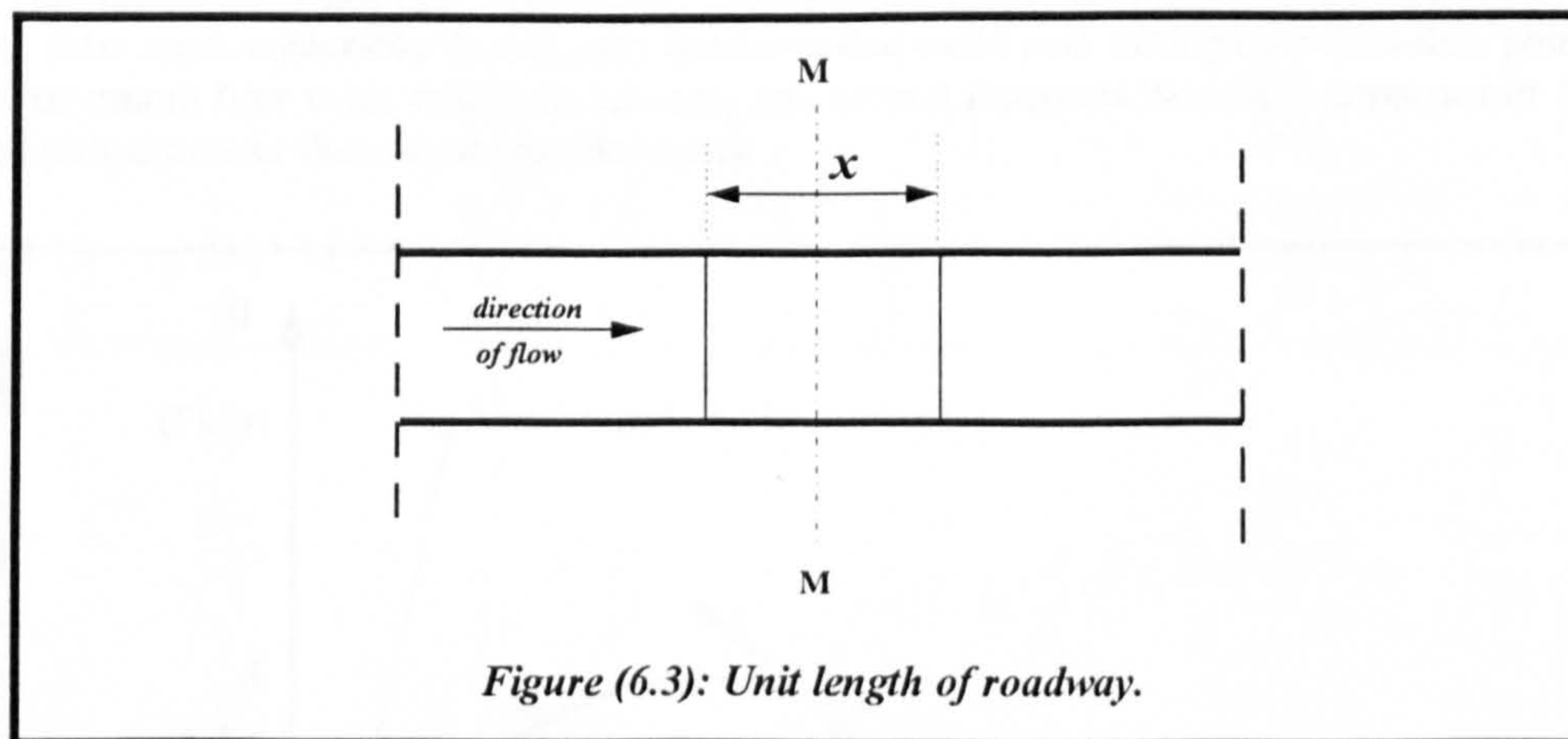
Table (6.1): Traffic Flow variables.

<i>Variable</i>	<i>Description</i>	<i>Typical units</i>	<i>Symbol</i>
Volume or flow rate	The number of vehicles passing a point in a unit of time.	vehicles per hour	q or μ
Density or concentration	Number of vehicles travelling over a unit length of highway.	vehicles per mile	D
Space mean speed*	Mean of the speeds of the vehicles travelling over a given length of road and weighted according to the time spent travelling that length.	vehicles per hour	$\bar{u}_s$

Source: Martin Whol and Brian V. Martin (1967).

* Other references define space mean speed as "the arithmetic mean of the speeds of vehicles occupying a given length of roadway at a given instant".

The relationship between these three variables can be obtained by considering the situation on a short length of road, x , for an interval of time T , as illustrated in Figure (6.3) below.

**Figure (6.3): Unit length of roadway.**

If 'n' vehicles cross the section line MM during the time interval 'T', then the flow rate μ is given by

$$\mu = n/T \quad \text{..... (6.1)}$$

Density = $D = (\text{average number of vehicles travelling over } x)/x$.

The average number of vehicles travelling over x could be determined from

$$\sum_{i=1}^n (t_i) / T,$$

where t_i is the time for the i th vehicle to traverse the distance x . Thus, the density, D , is

$$D = \frac{\sum_{i=1}^n (t_i) / T}{x} \quad \text{..... (6.2)}$$

Dividing μ by D , the space mean speed $\bar{u}_s$ is determined as follows:

$$\bar{u}_s = \frac{(n/T)x}{\sum_{i=1}^n (t_i)/T} = \frac{x}{\frac{1}{n} \sum_{i=1}^n t_i} \quad \dots\dots\dots (6.3)$$

Also,

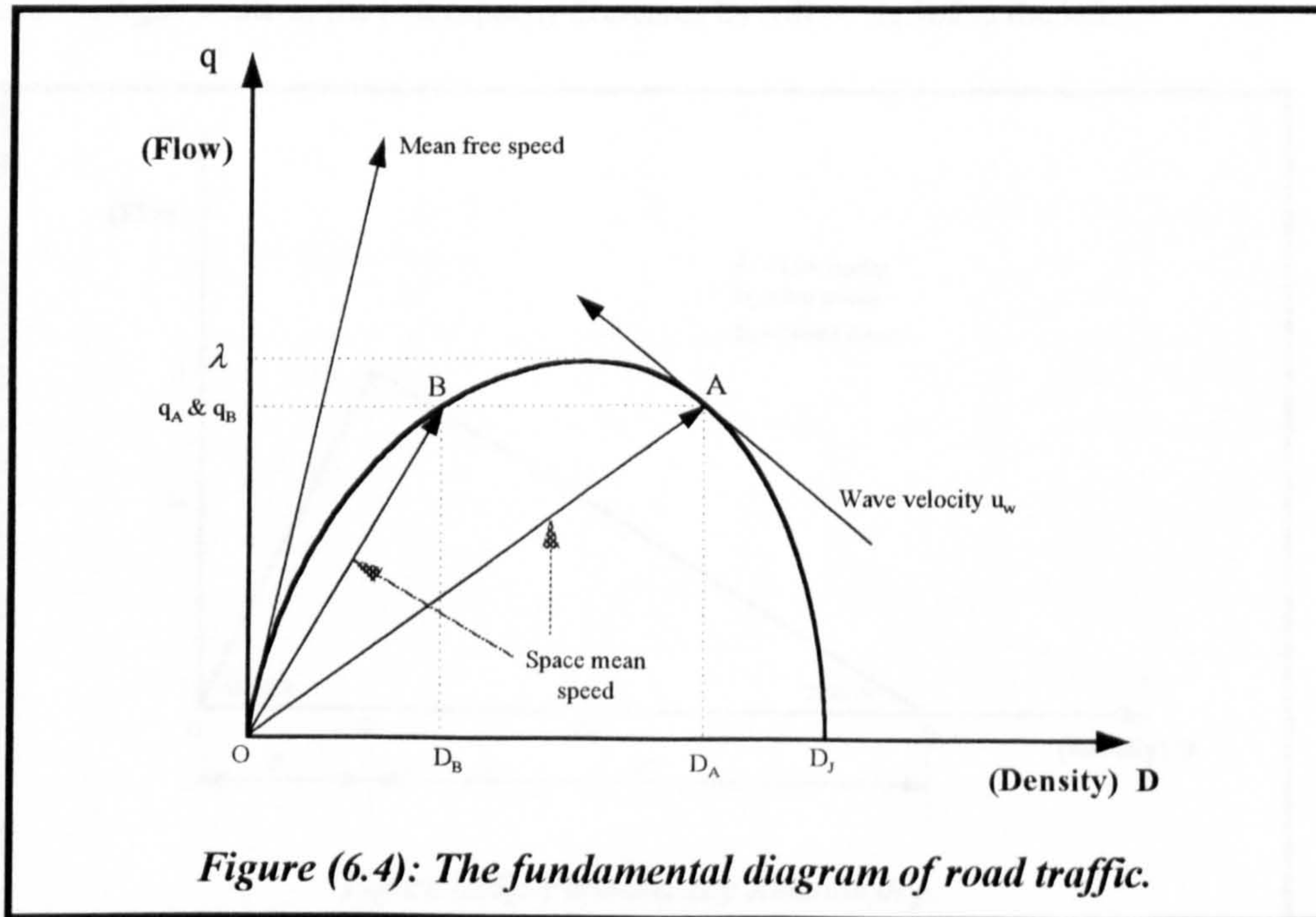
$$\mu = D \cdot \bar{u}_s \quad \dots\dots\dots (6.4)$$

Equation (6.4) is a common expression relating volume, density and space mean speed. However, this relation is of limited usefulness in its present form since each variable is dependent on a vast number of physical and psychological parameters which are functions of the particular sample of drivers, vehicle characteristics, roadway characteristics, and the weather conditions.

However, in this model, the packets' movements along a road link (i.e. link performance) is simulated based on the above link flow-density relationship and the car-following behaviour as discussed below.

6.3.4.1 Link flow-density relationship

The relationship between flow and density has a general form similar to that illustrated in Figure (6.4). This relationship has been termed the "fundamental diagram of road traffic" (Haight, 1963). Clearly, there are innumerable flow-density functions that could pass through the zero-flow points and have a maximum flow value falling in between, and several functions have been proposed or derived from empirical data or theoretical considerations.



Since space mean speed is flow divided by density, the slope of the line OA represents the space mean speed corresponding to flow q_A and density D_A . It is also clear that for the flow q_A there is another possible speed, defined by the slope of the line OB and corresponding to density D_B . Further, at the jam density D_J , speed is zero, and as the flow and density approach zero, the speed will be equal to the mean free flow speed (i.e. the speed that drivers will assume when free or virtually free from interference by other vehicles). The mean free flow speed will be a function of the drivers and their

vehicles characteristics, the highway characteristics (lane width, sight distance, etc.), and other factors such as lighting and weather.

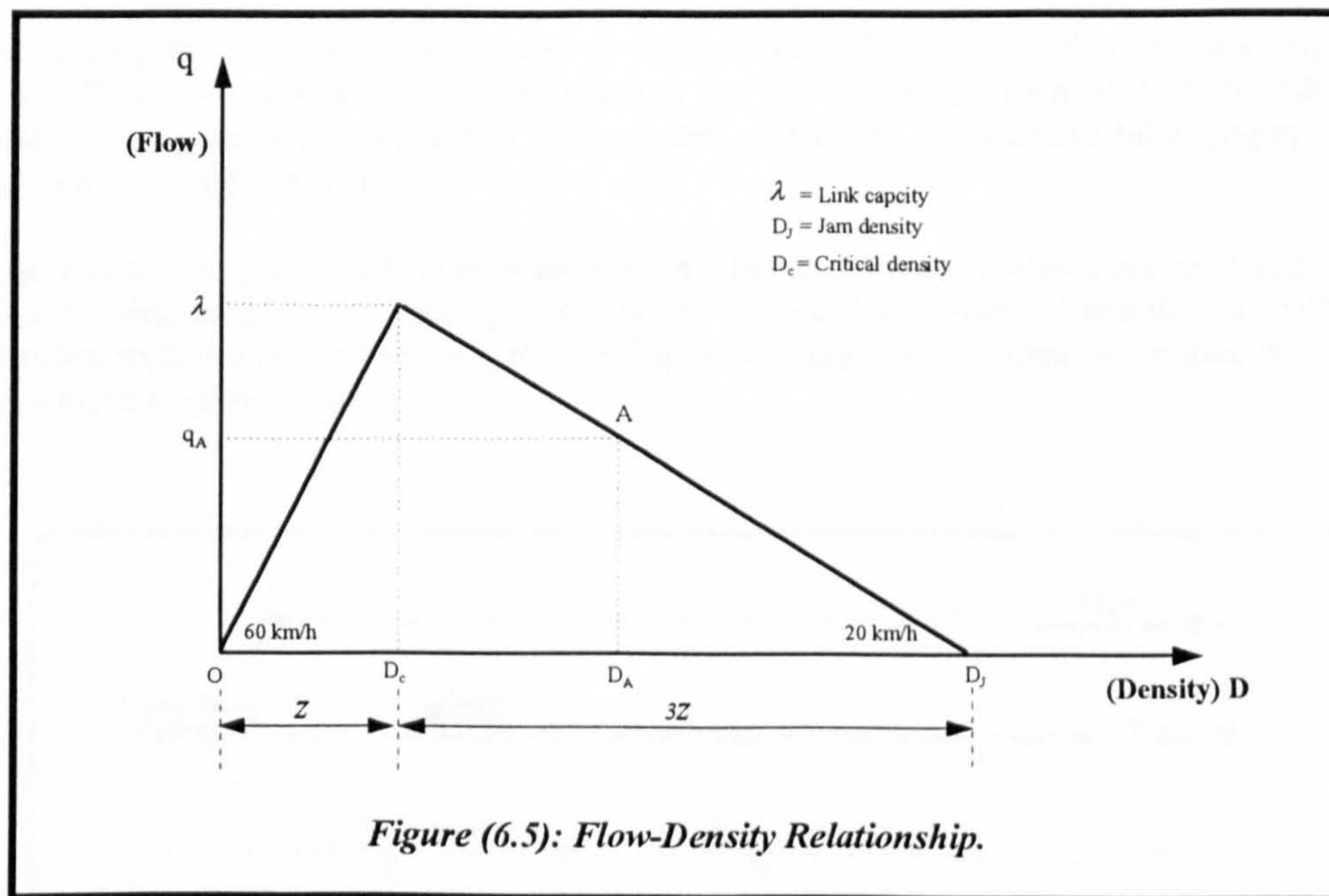
Lighthill and Whitham (1955) have shown that the speed of waves "carrying" continuous changes of volume through a vehicular flow is given by $d\mu/dD$ or the slope of the fundamental diagram; thus

$$u_w = \frac{d\mu}{dD} = \frac{d(\bar{u}_s \cdot D)}{dD} = \bar{u}_s + D \frac{d\bar{u}_s}{dD} \quad \dots\dots\dots (6.5)$$

Where u_w is the speed of the wave in miles per hour.

Figure (6.4) also shows that the space mean speed of traffic decreases with increasing density, and thus $d\bar{u}_s/dD$ will be negative. Also from Equation (6.5), the speed of the wave, u_w , will be less than the mean speed of traffic $\bar{u}_s$, and the wave will move backward relative to the mean vehicular flow. At low densities, when the interaction between vehicles is very small, $d\bar{u}_s/dD$ approaches zero, at that point the speed of the wave equals the vehicular speed. At densities above the point of maximum flow, the waves will move backward relative to the road, and under the conditions for maximum flow (i.e. at capacity), the wave will be stationary relative to the road. Although the flow and density are not linearly related, the model used in this study suggested a linear relationship between them as a workable approximation. The general form of this relation for all links is shown in Figure (6.5). The free flow speed is assumed 60 km/h and the speed of the wave at the jam density is equal to 20 km/h (the space mean speed at jam density is zero).

This relation differs from one link to another only according to the value of the capacity (maximum flow rate) corresponding to each link as shown in Figure (6.5). In other words, since the speed at the jam density and the free flow speed remain constant, then as the link capacity increases, D_j will be shifted to the right, while as the link capacity decreases D_j will be shifted to the left.



From the geometry of Figure (6.5), the jam spacing S_j is given by

$$S_j = \frac{1000}{4 \cdot \frac{\lambda}{60}} \quad (\text{in metres}) \quad \dots\dots\dots (6.6)$$

where λ is the link capacity in vehicles per hour.

Also the relation between flow and density at point 'A' is given by

$$D_J - D = \mu/20$$

Which could be rewritten in this form

$$D = \frac{D_J}{\left(\frac{\mu}{20D} + 1\right)}$$

Inverting this relation leads to

$$\frac{1}{D} = \left(\frac{v}{20} + 1\right) \cdot \frac{1}{D_J}$$

$$\text{i.e. } S = \left(\frac{v}{20} + 1\right) \cdot S_J \quad \dots\dots\dots (6.7)$$

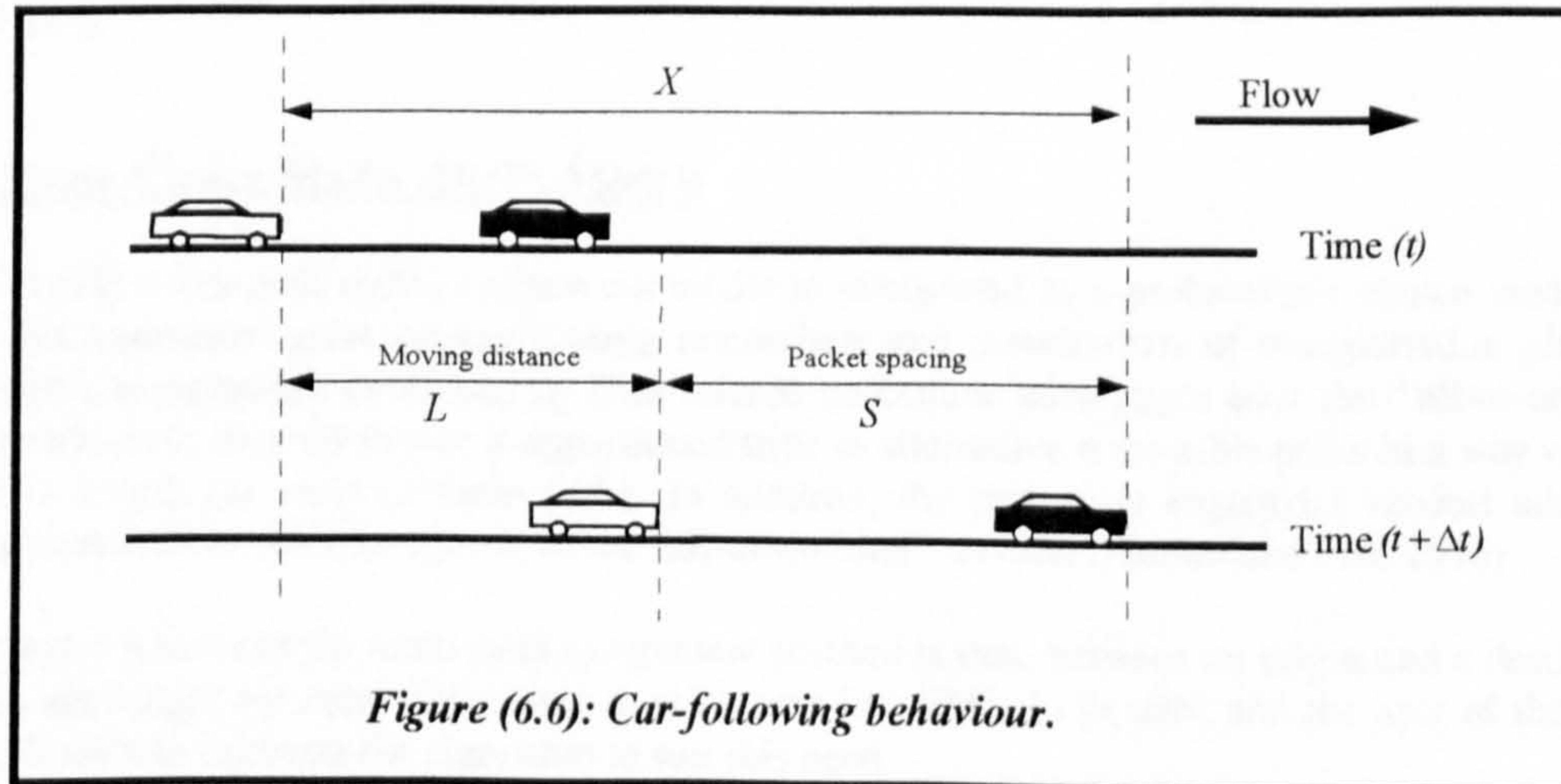
Where S and S_J are the current spacing and the jam spacing respectively (in metre), and v is the speed (in metre per second).

However, this last relation is used in the following section to formalise the relation between speed and density adopted for the simulation model to move a line of identical vehicles along a link.

6.3.4.2 Car-following behaviour

Car-following theory describes the manner in which vehicles follow each other on a highway without passing. There are several models (approaches) for car following, such as relative velocity and constant spacing control. A review of these models as well as an extensive bibliography, could be found in Martin and Brain (1967).

The approach used in this study is done by moving a line of identical vehicles (packets) each of which follows the one ahead under spacing and velocity control. The values of spacing and velocity are determined such that the flow-density relationship of that link is maintained. A detailed description of this approach is given below.



As shown in Figure (6.6), by moving the front vehicle, the distance between the head of the front vehicle and the next one becomes known ($S+L$). Then the moving distance of the next vehicle L is determined such that the new spacing S and the velocity v satisfy the predetermined flow-density relationship for that link as explained below.

From Figure (6.6), by moving the front vehicle, the distance x becomes known, and the values of spacing S and the moving distance L are given by

$$S+L = x \quad \text{.....} \quad (6.8)$$

The moving distance L during the time interval Δt is equal to $v \cdot \Delta t$, thus

$$S = x - v \cdot \Delta t \quad \text{.....} \quad (6.9)$$

Substituting by Equation (6.9) in Equation (6.7) yields

$$v = \frac{x - S_j}{\left(\frac{1}{20} S_j + \Delta t\right)}$$

Putting $\Delta t=1$ second, the above equation yields

$$v = \frac{x - S_j}{\left(\frac{3.6}{20} S_j + 1\right)} \quad (\text{m/sec}) \quad \text{.....} \quad (6.10)$$

The complete form of Equation (6.15) is given by

$$v = \begin{cases} 60 / 3.6 - - - - - S \geq S_c \\ \frac{x - S_j}{\left(\frac{3.6}{20} S_j + 1\right)} - - - - - S_c > S \geq S_j \end{cases} \quad (\text{m/sec}) \quad \text{.....} \quad (6.11)$$

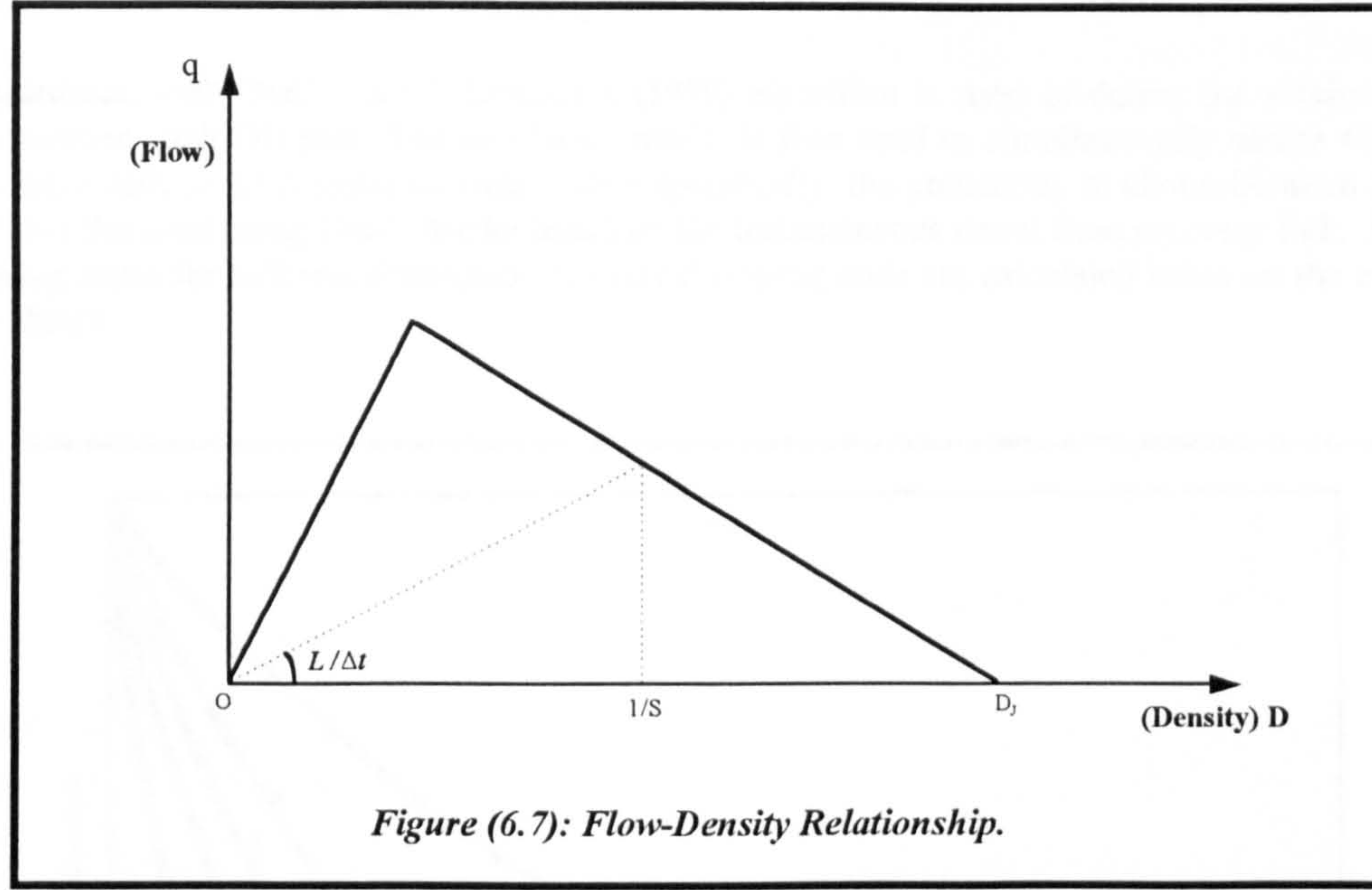
Where S_c is the critical spacing, i.e. the spacing that is corresponding to the link capacity as shown in Figure (6.5)

Since x is known and the value of S_j is given by Equation (6.6), then the speed v could be obtained from Equation (6.11). Substituting by the value of speed v and $\Delta t=1$ in Equations (6.8) and (6.9) to get the value of the spacing S and the moving distance L . This process is illustrated diagrammatically in Figure (6.7).

6.3.5 Route Choice Model (Dial's Model)

Dial's (1971) multi-path traffic assignment model is interpreted as a probabilistic choice model. The model has stimulated great interest among researchers and practitioners of transportation planning. The ingenious algorithm developed by Dial offered immediate advantages over the "all-or-nothing" traffic assignment method in that it apportioned trips to alternative reasonable paths in a way which is related to length (or cost) of these paths. In addition, the procedure required a modest additional computational effort when compared to the "all-or-nothing" method (Florian and Fox, 1976).

An attractive feature of the multi-path assignment method is that, between an origin and a destination, as paths get longer (or more expensive) they become less likely to be used and the user of the model has the facility to calibrate the algorithm to suit this need.



Dial's model is based on the hypothesis that there should be a non-zero probability of use of all 'efficient' paths. He proposed two definitions of efficient path, viz.:

1. a path in which every link has its initial node closer than its final node to the origin and has its final node closer than its initial node to the destination; and
2. a path in which every link has its initial node closer than its final node to the origin.

The first of these two definitions is the more attractive (Roy, 1991) but these dual constraints restrict the set of efficient path to those relating symmetrically to the origin and destination nodes. This duality requires that the assignment algorithm be executed once for each pair of nodes "o" and "d". While this is a reasonable requirement, it is time consuming in execution. First, the shortest path length from each node to "d" must be known, and second there are many o/d pairs. The second eliminates the need to know the shortest path distance from each node to the destination node "d", and all trips originating at the origin "o" to all destinations can be assigned simultaneously in a single execution of the algorithm (Roy, 1991).

Conceptually, consider an origin-destination pair (o, d) and the set of paths P that are admissible between them. The path flows are h_p , $p \in P$. Dial's mechanism achieves a subdivision of the total demand, q_{od} , for the pair (o, d) in the following proportions by means of a 'logit model'.

$$\text{Prob}(p) = k \exp(\theta(t^* - t_p))$$

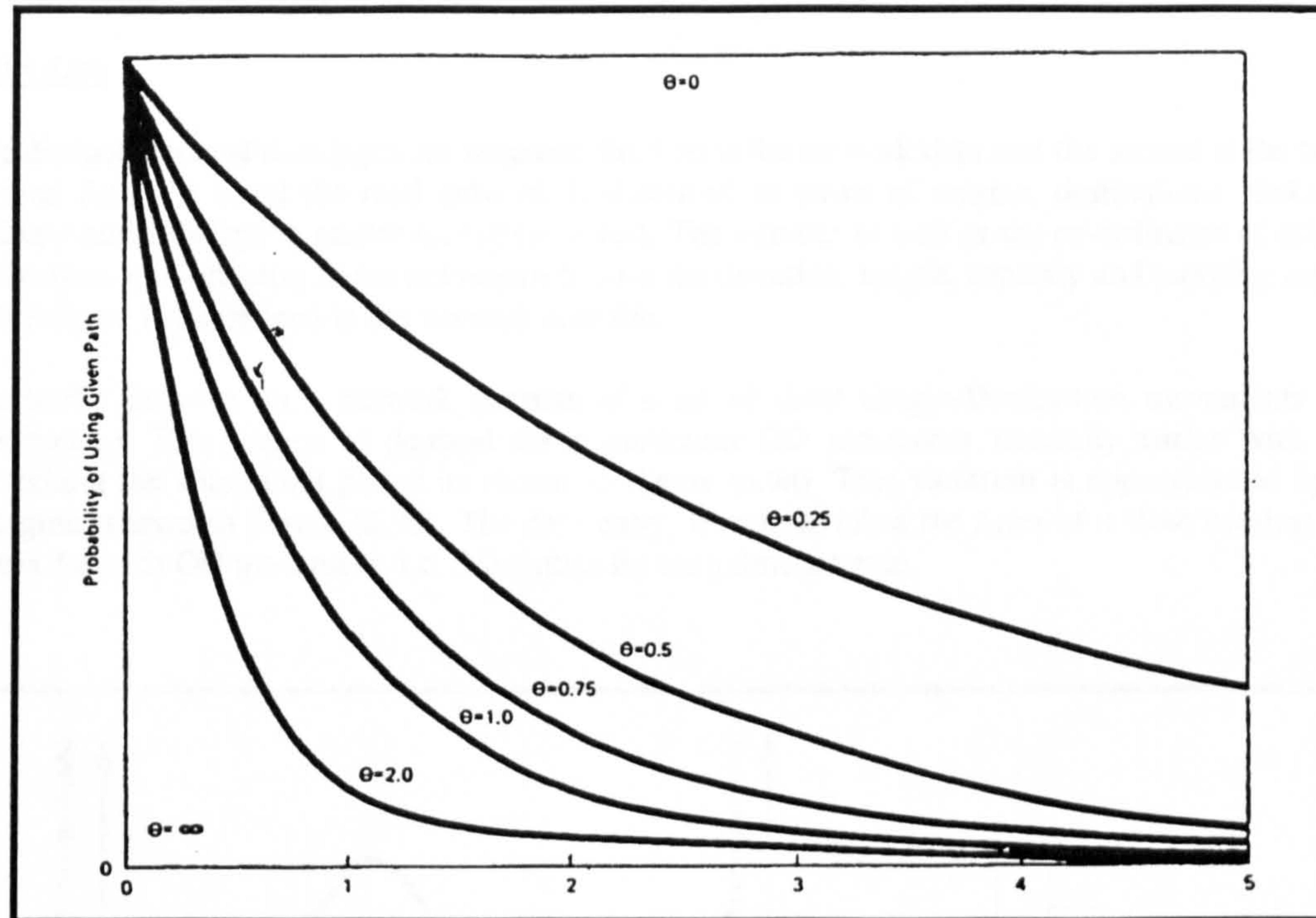
Where t^* is the length (or cost) of the shortest path between (o, d), t_p is the length (or cost) of path p and k is a constant such that

$$\sum_{p \in P} \text{Prob}(p) = 1$$

The parameter ' θ ' is a positive number which (can raise or lower division probabilities) determines the proportions of trips allocated to the efficient paths. As shown and graphed in Figure (6.8), as θ varies from zero to infinity, the probability of using a particular path which is Δt longer than the shortest path is directly proportional to $\exp(-\theta \Delta t)$. Thus as θ increases, the probability that a trip uses the shortest path also increases. When θ is zero, all efficient paths are considered equally likely. At the

other extreme, when θ is large, i.e. 10 or larger, the effect is a multiple shortest-path assignment, which assigns trips to all and only shortest paths.

In accordance with Dial's model, Dijkstra's (1959) algorithm is used to define the minimum path route between each OD pair. The simulation model is then used to simultaneously assign trips from each origin node to all destination nodes. More specifically, the probability of choice between different routes is calculated using Dial's Model based on the instantaneous travel time on every link. Then the diverging ratios for different destination at every diverging node are calculated based on the evaluated route choice.



Source: Dial (1971).

$$\Delta t = (\text{Length of shortest Path}) - (\text{Length of Given Path})$$

Figure (6.8): Probability of using a given path Δt longer than shortest path.

Note: The scale of the vertical axis depends on the number and length of competing efficient paths.

Since the route choice model is based on the instantaneous travel time along each link rather than the actual experienced one, it does not perfectly reproduce the stochastic user equilibrium (Yoshii and Kuwahara (1993)). Therefore, it is necessary to repeat these calculations in an iterative process with a feedback from the previous iteration (as explained later in this chapter under section (6.4)).

6.3.6 Propagation of Packets/Vehicles from one link to another

Since the basic model is developed to simulate traffic flow on urban expressway networks, no road intersections are considered and all the ordinary nodes are either merging or diverging nodes. For all links diverging at a certain node, the diverging ratios are determined by the route choice model (as explained above), while for all links merging at a certain node, the merging ratio for every link is defined in the input data file.

In this model, the maximum number of the merging links that could meet at one node is restricted to only two links. Thus, the priority of any two links i, j merging at a certain node is only decided when both of them reach the jam density. Then the priority will be given to the one with higher merging ratio (as defined in the data file), say link i . This priority is reversed and given to the other link j when

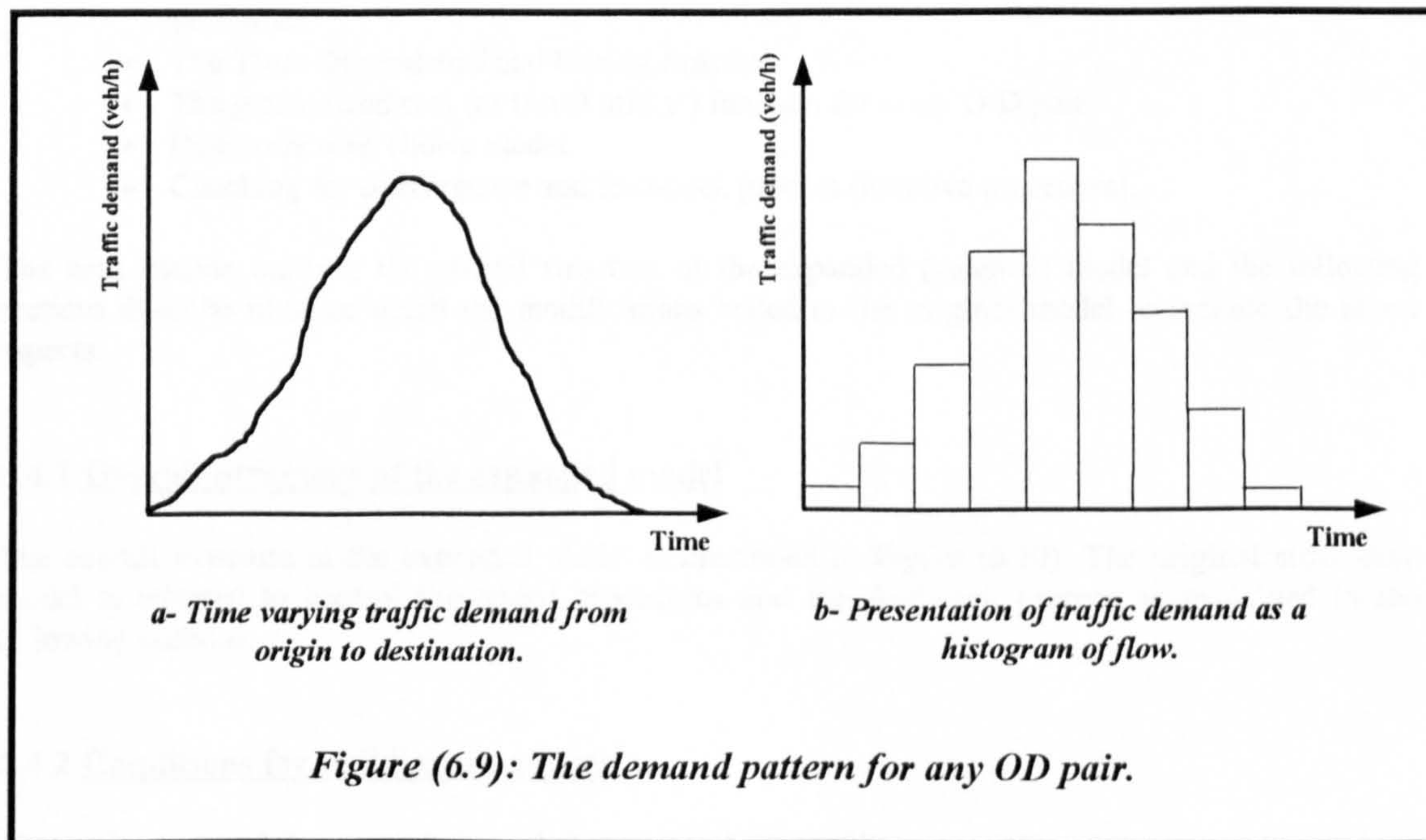
the ratio of the number of vehicles given the priority on link i over that for link j is greater than or equal to the ratio of the merging ratio for link i over that for link j and so forth. Thus, if this condition is satisfied, then the vehicles on link j will be given the priority and allowed to depart from this link to the following one. On the other hand, if this condition does not satisfy, the vehicles on link j will be moved forward but they will not be allowed to depart from the link during the current scan interval, and they must wait in this link until the above condition is satisfied for link i and the priority is reversed and given to link j . These procedures are carried out for every merging node during every scan interval throughout the study period.

6.3.7 Model input/output

Input data

Two distinct forms of data input are required: the first is the network data and the second is the traffic demand data. As usual the road network is described in terms of origins, destinations, links and ordinary nodes (merging and/or diverging nodes). The number as well as the co-ordinates of origins, destinations and ordinary nodes are required. Also the direction, length, capacity and merging ratio of each link are to be defined in the network data file.

The traffic demand on a network consists of a set of fixed Origin-Destination movements (OD movements). The pattern of demand for a particular OD movement normally varies with time throughout the simulation period as shown in Figure (6.9a). This variation is approximated by the histogram shown in Figure (6.9b). The data entry, therefore, takes the form of a time varying flow matrix for each OD movement, i.e. OD matrix for each time interval.



Output data

At this stage, the model output includes a detailed information about the packets' movements during the study period. For instance, the origin, destination, departure times and arrival times for each packet during the simulation period are given. The model also produces the cumulative number of vehicles (or packets) departing from each link every time interval throughout the simulation period.

6.3.8 Other features of the simulation model

The simulation model has some other features; for example, graphics display and the ability to simulate big traffic networks without having computer storage problems. These two features, however, are incorporated in the model by calling the following two pairs of subroutines.

ZAHYO and DISPLY subroutines

These two subroutines are responsible for graphics: while the function of ZAHYO subroutine is to draw the network on the screen, the DISPLY subroutine displays the vehicles (packets) movements between nodes on the screen during the simulation period.

SORT and SORTID subroutines

These two subroutines are used to resort the vehicles currently on the network and cancel the packets which have already reached their destinations. This process is done regularly every scan interval to reduce the storage area on the computer.

6.4 Expanding the original model to include TDRP

In order to demonstrate the different impacts of *TDRP* on road networks, the model outlined in section (6.3) is expanded to include the following important key aspects:–

- Queuing conditions downstream each link (i.e. starting and vanishing times for queues on each link).
- The Time-Dependent Road Pricing function.
- The generalized cost (or travel utility) function for every O-D pair.
- Departure time choice model.
- Checking for convergence and feed-back process (iterative procedure).

The next section outlines the overall structure of the expanded (adapted) model and the following sections describe in some detail the modifications added to the original model to include the above aspects.

6.4.1 Overall Structure of the expanded model

The overall structure of the expanded model is illustrated in Figure (6.10). The original simulation model is adapted to control the added procedures and the feed-back process as explained in the following sections.

6.4.2 Conditions for building up queues

The original model does not give any information about building up or decaying queues downstream links. On the other hand, queue starting and vanishing times are considered as a key element in estimating the value of *TDRP* (as explained in Chapter 3). It is thus necessary to modify the original model to check the queuing occurrence and to decide the starting and vanishing times of every queue occurring along any link during the study period.

Without losing generality, two assumptions are considered in order to carry out the above procedure:–

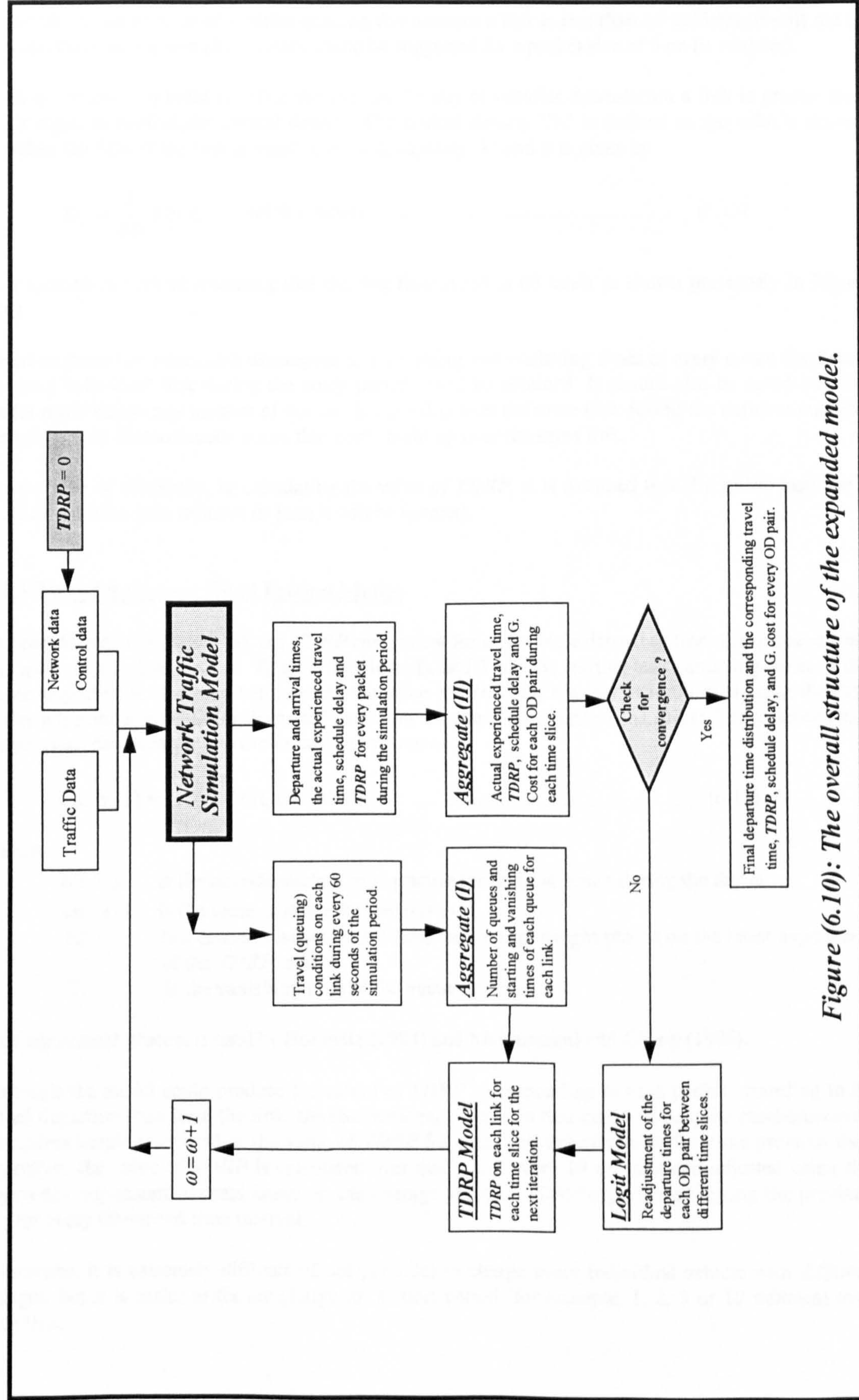


Figure (6.10): The overall structure of the expanded model.

1. The minimum queue length is 10 vehicles (if the packet size is equal to one vehicle). In other words, if the number of vehicles queuing downstream a link is less than 10 vehicles, it will not be considered as a queue (50 vehicles would be suggested for a packet size of 5 or 10 vehicles).
2. A queue starts to build up when the average density of vehicles downstream a link is greater than or equal to twofold the critical density. The critical density ' D_c ' is defined as the vehicle density when the flow of the link is equal to the link capacity ' λ ' and it is given by

$$D_c = \frac{\lambda}{60} \cdot 1000 \quad (\text{vehicles/metre}) \dots\dots\dots (6.12)$$

This relation is derived assuming that the free flow speed is 60 km/h as shown previously in Figure (6.6)

Based on these two reasonable assumptions, the starting and vanishing times of every queue developed on every individual link during the study period could be obtained. It should also be noted that the model could define any number of queues that develop over the same link during the entire simulation period, i.e. any discontinuous queue that could build up over the same link.

For the sake of simplicity, in calculating the value of *TDRP*, it is assumed that if a queue lasts for a short time period (two minutes or less) it will be ignored.

6.4.3 Time-Dependent Road Pricing Model

As been derived in Chapter (3), the *TDRP* component for any vehicle departing from the queue at time t , is given by $(T-t)$ as long as $T_0 \leq t \leq T$, where T_0 and T are the starting and vanishing times of the queue respectively. However, before using the value of *TDRP* for any day, it is adjusted using the day-to-day adjustment process (as discussed earlier in Chapter (4)) to reflect the drivers' experience from the previous day according to the following expression:

$$\tau_i(t, \omega+1) = (1-R_2) \cdot \tau_i(t, \omega) + R_2 \cdot (T_i - t) \dots\dots\dots (6.13)$$

Where:

- $\tau_i(t, \omega)$ is the time-dependent road pricing for link i at time t during the day ω .
- $\tau_i(t, \omega+1)$ is the same as above for day $(\omega+1)$.
- R_2 is a constant non-negative (less than unity) weight placed on the latest experience of the *TDRP*. and
- T_i is the vanishing time of the queue on link i .

This adjustment process is used by Horowitz (1984) and Mahmassani and Chang (1986).

Although the model could produce the values of *TDRP* corresponding to each packet according to its actual departure time from the link, the readjustment process in this case will be quite cumbersome as it requires keeping a record of the value of *TDRP* for every individual vehicle from the previous day. Therefore, the value of *TDRP* is calculated over every link every 10 seconds and adjusted using the day-to-day adjustment process based on the average value of *TDRP* experienced during the previous day for every 60-second time interval.

In practice, it is extremely difficult (if not possible) to charge every individual vehicle with different charges, but it is easier to fix the charge for a short period (for example, 1, 2, 5 or 10 minutes) over each link.

6.4.4 Generalized Cost Function

The generalized cost (or travel utility) used for deciding the route choice and the departure time choice (as explained in the next sub-section) is a function of the following three key variables:

1. Total travel time $r(t)$

It is defined by the time from departing from the origin till arriving at destination, including all kinds of delay. It is denoted by $r(t)$, where t is the departure time from the origin. This cost, of course, depends on the chosen route and the delay experienced on every individual link along that route.

2. Schedule delay $s(t)$ & $p(t)$

Scheduled delay is defined as the absolute value of the difference between the desired and the actual arrival times. This delay is modelled with work start time flexibility, i.e. by considering the time interval $[B-\Delta, B+\Delta]$ is the desired time period for arriving at destinations, where B denotes the centre of the period and Δ is a measure of work start time flexibility ($\Delta \geq 0$). The early and late schedule delay functions are denoted by $s(t)$ and $p(t)$ respectively and they are given by

$$s(t) = \begin{cases} (B - \Delta) - t_a(t) & \text{---} t_a(t) < (B - \Delta) \\ 0 & \text{--- otherwise} \end{cases} \quad (6.14)$$

$$p(t) = \begin{cases} t_a(t) - (B + \Delta) & \text{---} t_a(t) > (B + \Delta) \\ 0 & \text{--- otherwise} \end{cases} \quad (6.15)$$

Where $t_a(t)$ is the time at which the vehicle departing from origin at t reaches its destination.

These two function are explained earlier in more detail in chapter (4).

3. Time-Dependent Road Pricing Function $\tau_r(t, \omega)$

The value of *TDRP* considered in the generalized cost function is the sum of the charges incurred by the vehicle over all links along the route chosen between its origin and destination. It is denoted by $\tau_r(t)$, where t is the departure time from the origin.

The generalized cost function $GC(t, \omega)$ for a vehicle departing from origin at t on day ω is then given by:

$$GC(t, \omega) = \alpha.r(t, \omega) + \beta.s(t, \omega) + \beta\gamma.p(t, \omega) + \alpha.\tau_r(t, \omega) \quad (6.16)$$

Where α , β , $\beta\gamma$ are the shadow values of travel time, early schedule delay and late schedule delay respectively.

The generalized cost is calculated independently for every individual packet during the simulation period. At the end of each iteration (day), the generalized cost is aggregated for each OD pair for different time intervals (using Aggregate II), i.e., the average generalized cost corresponding to every time interval is calculated for every OD pair. These values could be represented in a three-dimension array; the first dimension represents the time interval, and the second and the third dimensions represent the origin and destination respectively.

6.4.5 Departure time choice model

The departure time choice problem under given transport system attributes has been addressed by a number of researchers, including Cosslett (1977), Small (1978), Abkowitz (1980, 1981) and Hendrickson and Plank (1984). These essentially viewed the problem as one of choosing among a finite set of alternative time "slices" (resulting from a discretization of the continuous time range of interest). Each time slice is characterized by a different generalized cost (or utility level) to a trip-maker. This approach is discussed and used for simulating a simple traffic bottleneck in the previous chapter. It is also used in this model in conjunction with the day-to-day adjustment process used by drivers to revise their travel behaviours from one day to another. The mathematical model is given by a "Logit" model (de Palma et al. (1983) and Ben-Akiva et al. (1984)) as expressed by:

$$R_{o,d}(t, \omega) = R1 \frac{Q_{o,d}}{E_{o,d}(\omega)} \exp\left[\frac{1}{\eta} GC_{o,d}(t, \omega)\right] + (1 - R1) \cdot R_{o,d}(t, \omega - 1) \dots\dots\dots (6.17)$$

Where:

$R_{o,d}(t, \omega)$ is the departure rate from origin o to destination d during the time interval $[t, t+\delta]$ on day ω .

$R_{o,d}(t, \omega+1)$ is the same as above for day $(\omega+1)$.

$Q_{o,d}$ is the total number of users departing from origin o to destination d during the study period.

$R1$ is the fraction of users who review their departure time choices every day and it is assumed to be the same for all OD movements.

$GC_{o,d}(t, \omega)$ is the generalized cost for travellers between o and d during the time interval $[t, t+\delta]$ on day ω .

η is a scale parameter reflecting the variability of preferences among individuals.

$E_{o,d}(\omega)$ is the denominator of the logit model for all users travelling between o and d on day ω , and is approximated by the following discrete form:

$$E_{o,d}(\omega) = \frac{1}{2} \left[\exp\left(\frac{1}{\eta} GC_{o,d}(0, \omega)\right) + \exp\left(\frac{1}{\eta} GC_{o,d}(K, \omega)\right) \right] + \sum_{t=1}^{K-1} \exp\left(\frac{1}{\eta} GC_{o,d}(t, \omega)\right) \dots\dots\dots (6.18)$$

where K is the total number of intervals ($K=(T-T_0)/\delta$) and δ is the time increment (or time interval).

At the end of each iteration and as long as the condition for convergence is not satisfied, the above model is independently applied for every individual OD pair according to the average generalized costs experienced during the different time intervals in the previous day.

6.4.6 Checking for convergence

The simulation process is terminated when either the last day of the study is reached or convergence occurs. Convergence is assumed to occur when the weighted coefficient of variation of the generalized cost (weighted over all the OD movements throughout the network) is less than 5 per cent. The weighted coefficient of the generalized cost, cv_w , is calculated as follows:

$$\begin{aligned} cv_w &= \frac{\sum_d \sum_o SD(o, d) \cdot q(o, d) / \sum_d \sum_o q(o, d)}{\sum_d \sum_o GC(o, d) \cdot q(o, d) / \sum_d \sum_o q(o, d)} \\ &= \frac{\sum_d \sum_o SD(o, d) \cdot q(o, d)}{\sum_d \sum_o GC(o, d) \cdot q(o, d)} \dots\dots\dots (6.19) \end{aligned}$$

where $\overline{GC}(o, d)$, $SD(o, d)$ and $q(o, d)$ are the average generalised cost, the standard deviation and the flow demand respectively for all movements between the origin o and destination d . $\overline{GC}(o, d)$ and $SD(o, d)$ are given by

$$\overline{GC}(o, d) = \frac{\sum_t GC(t, o, d) \cdot q(t, o, d)}{\sum_t q(t, o, d)} \quad \dots\dots\dots (6.20)$$

and

$$SD(o, d) = \sqrt{\frac{\sum_t (\overline{GC}(o, d) - GC(t, o, d))^2 \cdot q(t, o, d)}{\sum_t q(t, o, d)}} \quad \dots\dots\dots (6.21)$$

Where $GC(t, o, d)$ and $q(t, o, d)$ are the generalized cost and the flow demand respectively for all movements between the origin o and the destination d during the time interval t .

6.4.7 Final output

In addition to the output of the original model (see section (6.3.7)) the expanded model produces the final departure time distribution for every O-D pair and the corresponding total travel time, schedule delay, *TDRP* and the generalized cost over every time interval. It also gives the number of queues as well as their starting and vanishing times for every individual link of the traffic network during the study period.

The simulation model is coded in FORTRAN 77 and implemented on an ALFA-network/Cranfield University Computer Centre and a Supper Computer/University of London Computer Network.

6.5 Interaction between nodes and its effect on the value of *TDRP*

6.5.1 Interaction between nodes (case with no merging/no diverging flow)

In the case of a simple junction (bottleneck) along a route, as discussed earlier, the delay imposed by the arriving vehicle on others and consequently the *TDRP* are dependent on the ratio of demand over capacity during the study period. On the other hand, if the number of nodes (junctions) along the route chosen by vehicles is two or more, the output flow pattern of each node is considered as an input for the following node as long as the traffic flow does not change along the route (i.e. there is no merging or diverging at nodes). Thus, the queuing delay at any node will be dependent on the queuing delay experienced at the previous node(s). To investigate the relation between the queuing delay experienced at different nodes along the same route, the simple example shown in Figures (6.11) to (6.14) is considered.

Figure (6.11) shows the two cases being examined under the same flow pattern; in case (1) only node "A" is considered, while in case (2) nodes "A" and "B" are considered such that the capacity of node "B" is greater than the capacity of node "A", i.e. $\lambda_B > \lambda_A$, (if $\lambda_B \leq \lambda_A$, no delay is experienced at A by any vehicle since node B precedes node A and the flow does not change at B). Figure (6.12) illustrates the flow pattern at any time along the route $\mu(t)$, the capacities of nodes A and B, and the output flow pattern from node B. The latter is considered as the input flow pattern for node A (in case (2)). The queue distributions for the two cases at A and B are illustrated in Figure (6.13). The total delay "D" during any time period is given by the area under the queue length curve during this time period. Thus, the total delay during the time period $[t_1, t_2]$ is given by

$$D_A = \int_{t_1}^{t_2} (\mu(t) - \lambda_A) \cdot dt = \int_{t_1}^{t_2} \mu(t) \cdot dt - (t_2 - t_1) \cdot \lambda_A \quad \dots\dots\dots (6.22)$$

$$D_B = \int_{t_1}^{t_2} (\mu(t) - \lambda_B) \cdot dt = \int_{t_1}^{t_2} \mu(t) \cdot dt - (t_2 - t_1) \cdot \lambda_B \quad \dots\dots\dots (6.23)$$

$$D_{A/B} = \int_{t_1}^{t_2} (\lambda_B - \lambda_A) \cdot dt = (\lambda_B - \lambda_A) \cdot (t_2 - t_1) \quad \dots\dots\dots (6.24)$$

From Equations (6.22), (6.23) and (6.24), it could be concluded that the total delay at A for case (1) is equal to the sum of the total delay at A and B for case (2) during the time period $[t_1, t_2]$, i.e.

$$D_A = D_{A/B} + D_B \quad \dots\dots\dots (6.25)$$

Since the total delay during the time periods $[T_0, t_1]$ and $[t_2, T]$ is the same in the two cases (see Figure (6.13)), then the total delay during the whole period $[T_0, T]$ is also the same for the two cases. Hence, vehicles delayed at node B will experience less delay at node A than the delay experienced in case (1) such that the total delay experienced by every individual vehicle is the same in the two cases. Accordingly, every individual vehicle will have the same departure time from node A in the two cases (see Figures (6.12) and (6.13)). This implies that adding node B (with a capacity greater than the capacity of the existing node A) to the system does not increase the overall delay, and has no influence on the starting and vanishing times of the queue at A. Thus, node B would only work as a traffic regulator for the queue length (i.e. the number of vehicles waiting in the queue) at A. Since the total delay does not change by adding node B to the system, then the delay imposed by each vehicle on one another should not change either. Therefore, imposing *TDRP* (based on the starting and vanishing times of the queue) independently at every individual node and ignoring the interaction with other nodes will lead to an overcharging for all vehicles delayed at node B. This is because in case (2) vehicles are charged the same charge at node A as in case (1) in addition to the charge at node B.

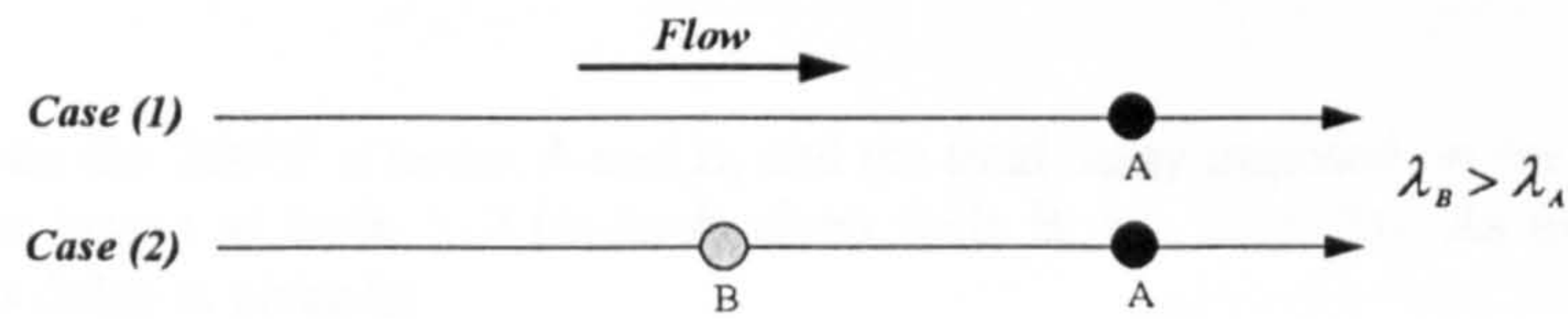


Figure (6.11): The two cases under consideration.

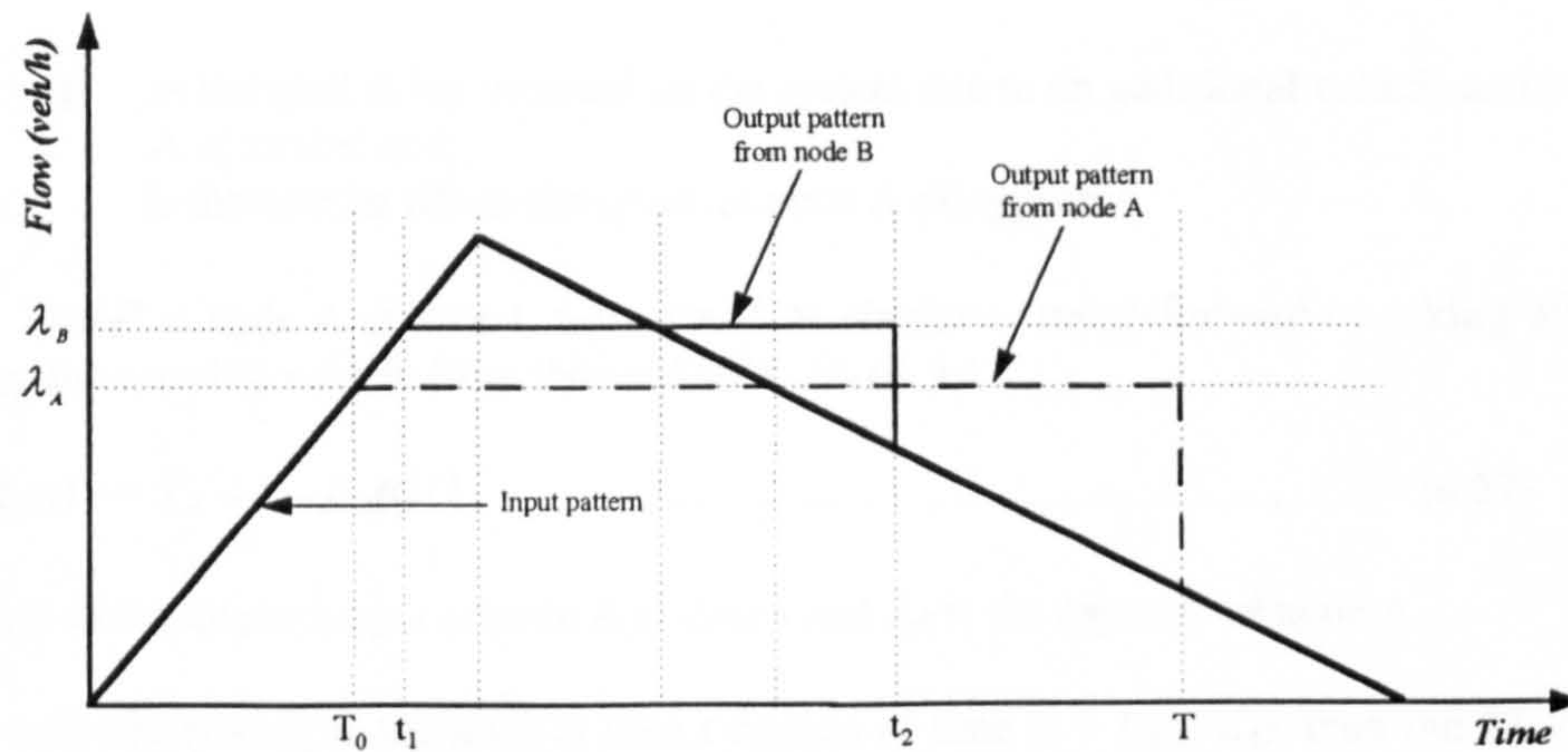


Figure (6.12): The flow patterns and the capacities at nodes A and B for the two cases being examined.

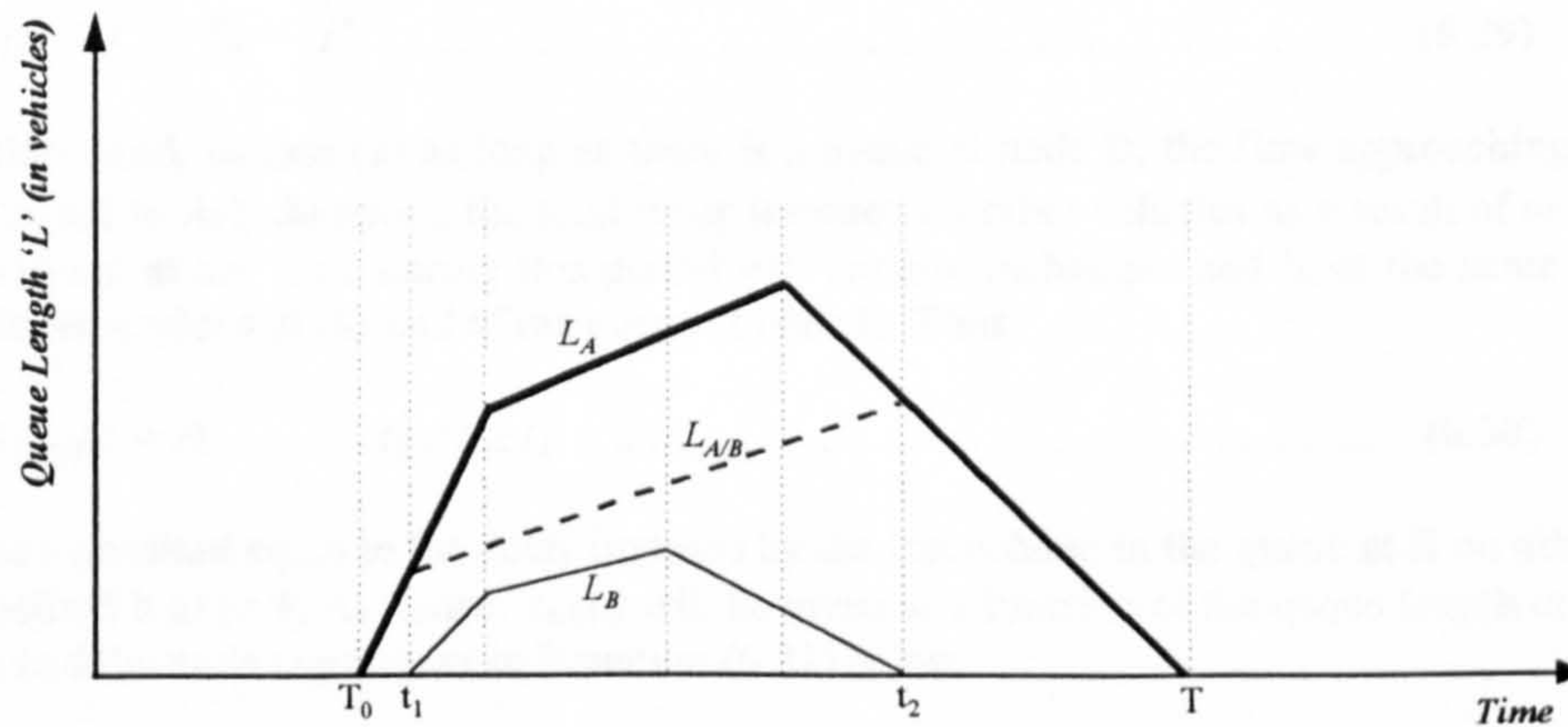


Figure (6.13): Queue lengths at nodes A, B and A/B.

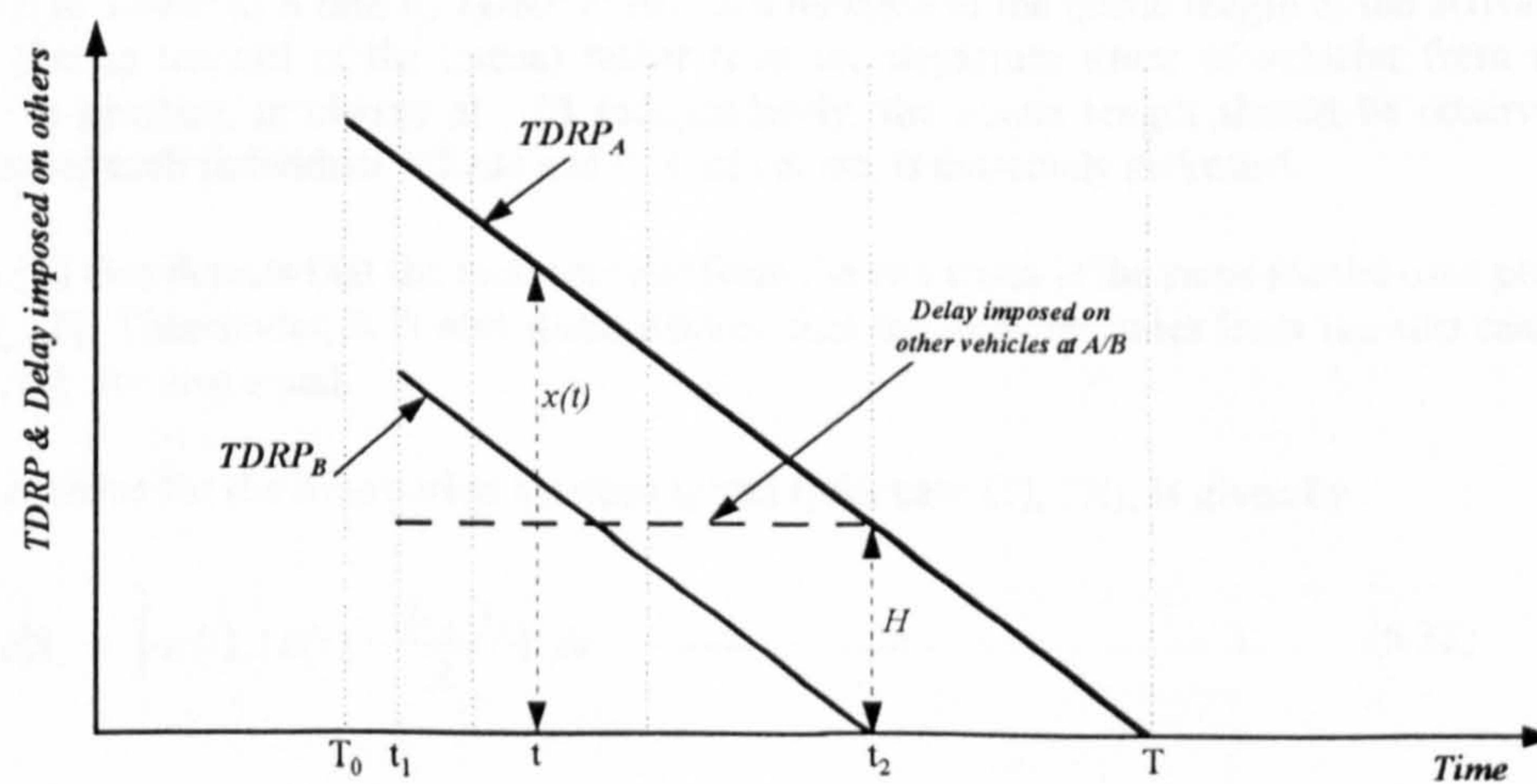


Figure (6.14): The TDRP at nodes A and B and the delay imposed on other vehicles at A/B.

Figure (6.14) shows the *TDRP* at nodes A and B, and the total delay imposed on the system due to any vehicle arriving at time t at node A/B (node A given node B, i.e. case (2)). As explained earlier in chapter three, this delay is given by

$$D_A(t) = T_A - t \quad \dots\dots\dots (6.26)$$

Where

$D_A(t)$ is the total delay imposed on the system due to an additional vehicle arriving at node A at time t ; and
 T_A is the time at which the queue at node A disappeared.

Thus, the *TDRP* at node A at time t , $\tau_A(t)$, could be obtained straightforward by taking off the delay incurred by the arriving vehicle from the total delay $D_A(t)$, i.e.

$$\tau_A(t) = T_A - t - L_A(t)/\lambda_A \quad \dots\dots\dots (6.27)$$

Where $L_A(t)$ is the queue length at node A at time t and λ_A is the capacity of node A.

Since the vehicle arriving at the node at time t departs at time $(t + L_A(t)/\lambda_A)$, then the *TDRP* function could be given as a function of the departure time from the node t' .

$$\tau_A(t') = T_A - t' \quad \dots\dots\dots (6.28)$$

$$\tau_B(t') = T_B - t' \quad \dots\dots\dots (6.29)$$

On the other hand, in case (2) as long as there is a queue at node B, the flow approaching node A is constant (equal to λ_B), therefore, the total delay imposed on other vehicles as a result of an additional vehicle arriving at any time during this period will remain unchanged and have the same value as if the vehicle were added at the end of the queue at node B. Thus

$$D_{AB}(t) = H \quad t_2 \geq t \geq t_1 \quad \dots\dots\dots (6.30)$$

where H is a constant equal to the delay imposed by the last vehicle in the queue at B on other vehicles arriving behind it at node A. Hence, $\tau_{AB}(t)$ will be given as a function of the queue length at the arrival time at A and the node capacity as in Equation (6.31) below

$$\tau_{AB}(t) = H - L_A(t)/\lambda_A \quad \dots\dots\dots (6.31)$$

As opposed to *TDRP* at A and B, *TDRP* at A/B is a function of the queue length at the arrival times to the node (i.e. to the tail of the queue) rather than the departure times of vehicles from the node. However, in practice, to charge at A/B independently, the queue length should be observed at the arrival time of each individual vehicle and this, of course, is extremely awkward.

Figure (6.14) also depicts that the total revenue from the two cases is the same for the time periods $[T_0, t_1]$ and $[t_2, T]$. Thereunder, it is also demonstrated that the total revenues from the two cases for the period $[t_1, t_2]$ are also equal.

The total revenue for the time period between t_1 and t_2 for case (1), TR_1 , is given by

$$TR_1 = \int_{t_1}^{t_2} n(t) \cdot \left(x(t) - \frac{L_A(t)}{\lambda_A} \right) dt \quad \dots\dots\dots (6.32)$$

Where $L_A(t)$ is the queue length at node A at time t and $x(t)$ is the total delay imposed by the vehicle arriving at node A at time t on other vehicles behind (for case (1)). $n(t)$ is the number of vehicles arriving at node A at time t .

Similarly, the total revenue for the time period between t_1 and t_2 for case (2), TR_2 , is the sum of the total revenue at node B and A/B and is given by

$$TR_2 = \int_{t_1}^{t_2} n(t) \cdot (x(t) - H - \frac{L_B(t)}{\lambda_B}) \cdot dt + \int_{t_1}^{t_2} n'(t) \cdot (H - \frac{L_{A/B}(t)}{\lambda_A}) \cdot dt \quad \dots\dots (6.33)$$

Where

- $n(t)$ & $n'(t)$ are the number of vehicles arriving at time t at nodes B and A respectively.
- $L_B(t)$ & $L_{A/B}(t)$ are the queue lengths at time t for nodes B and A respectively.
- λ_B & λ_A are the capacities of nodes B and A respectively.
- $x(t)$ & H as defined above.

Since the total number of vehicles arriving at nodes B and A between times t_1 and t_2 is the same (see Figure (6.12)), then

$$\int_{t_1}^{t_2} n(t) \cdot dt = \int_{t_1}^{t_2} n'(t) \cdot dt \quad \dots\dots\dots (6.34)$$

Also, from Equation (6.25), the total delay at A for case (1) is equal to the sum of the total delay at B and A/B for case (2), then

$$\int_{t_1}^{t_2} n(t) \cdot \frac{L_A(t)}{\lambda_A} \cdot dt = \int_{t_1}^{t_2} (n(t) \cdot \frac{L_B(t)}{\lambda_B} + n'(t) \cdot \frac{L_{A/B}(t)}{\lambda_A}) \cdot dt \quad \dots\dots\dots (6.35)$$

Substituting by Equations (6.34) and (6.35) into Equation (6.33) and comparing with Equation (6.32), demonstrates that the total revenues from the two cases between t_1 and t_2 are equal. Thus, the total revenues from the two cases between T_0 and T are also equal.

The main conclusion that emerges from this simple example, therefore, is that the overall delay and the total revenue levied along a route are dependent on the capacity of the critical node (the node with the lowest capacity) as long as the traffic flow along the route does not change from one node to another during the study period. In effect, adding one or more nodes (with higher capacity than the critical node) along that route will have no influence on the overall delay, the starting and vanishing times of the queue at the critical node or the total revenue from *TDRP*. This is because the delay incurred by each individual vehicle at every node is dependent on the delay incurred at other nodes such that the total delay incurred by each individual vehicle over all nodes remains unchanged. Of course, this delay will increase if the added node has a lower capacity than the existing node.

Thus, instead of levying *TDRP* at every node along a single route, the charges could be levied only at the critical node on the basis of starting and vanishing times of the queue at this node as if it were not influenced by any other nodes. The total revenue, in this case will be the same as that from charging vehicles at every individual node taking into consideration the interactions between these nodes. In fact, this is quite useful in practice for two reasons: first, minimizing the total costs of installing the road side equipment for road pricing; second, the *TDRP* is a function of the departure time from the critical node, while for other nodes it is a function of the queue length. In practice, it is much easier to observe the departure time for every individual vehicle than observing the queue length at the node when every vehicle arrives and joins the queue.

On the other hand, if every node is charged independently of other nodes along the route (i.e. ignoring the node interaction) and the charge is based on the queuing duration at each node, this will definitely lead to an overcharging for some vehicles along the route. This explain why the phenomenon of node interaction is of a great importance in charging vehicles throughout a traffic network. Perhaps the most striking merit of this phenomenon is: does charging only the critical node along the route incorporate all the charges due at all other nodes under all circumstances? and if so, how the critical node along every individual route should be defined? and upon what criterion should it be decided?

To properly answer these questions, the next section considers a more practical case; where the traffic flow along a single route changes from one node to another as a result of the merging and diverging flows at nodes. The overall delay of the whole system, in this case, will be dependent on the ratio of demand over capacity for every individual node during the study period.

6.5.2 Introducing merging and diverging flows at nodes

For a single route without merging or diverging nodes, there is no difficulty in determining the critical node since it is the node with the lowest capacity along that route. In other words, the critical node is the one with the longest queue duration. Unfortunately, this criterion can not be applied straightforward for a single route in practice, because of the merging and diverging flows at nodes and the complications of the flow movements and exit capacities allocated to different traffic movements (i.e. straight, left or right movements) at each node.

In the example shown in Figure (6.15) below, if the capacity of node A is higher than the capacity of node B, no queue is expected to build up at A as long as the sum of the flow proceeding from B to A and the merging flows at A is less than the capacity of node A (it should also be noted that the flow proceeding from B to A is affected by the diverging flow at B). On the contrary, if this sum exceeds the capacity of node A, the queue starts to build up and the queue duration depends on the sum of all the traffic flows approaching node A. If the merging flow at A is too high and extends for a long period, the queue duration at A could exceed that at B, although the capacity at A is higher than that at B. Likewise, if the capacity at node G is less than that at node A and the diverted flow at A is too high, node A could be more critical than node G as well. Hence, the critical node along a route can not only be defined according to its capacity compared with others, but it also depends on the merging flows at that node and the diverting flows from the previous one. Unlike node capacity, the merging and/or diverging flows could vary with time, therefore the critical node along the same route could also change from one node to another during the study period according to the ratio of the demand over capacity.

Moreover, for a single route without merging or diverging flows, the queue duration at any node could be considered as a sub-set of the queue duration of the critical node. On the contrary, due to merging and diverging flows, the queue duration at different nodes could overlap one another. For example, if $\lambda_A < \lambda_B$, the queue at A could start after the queue at B and also disappear after it, as shown in Figure (6.15) below.

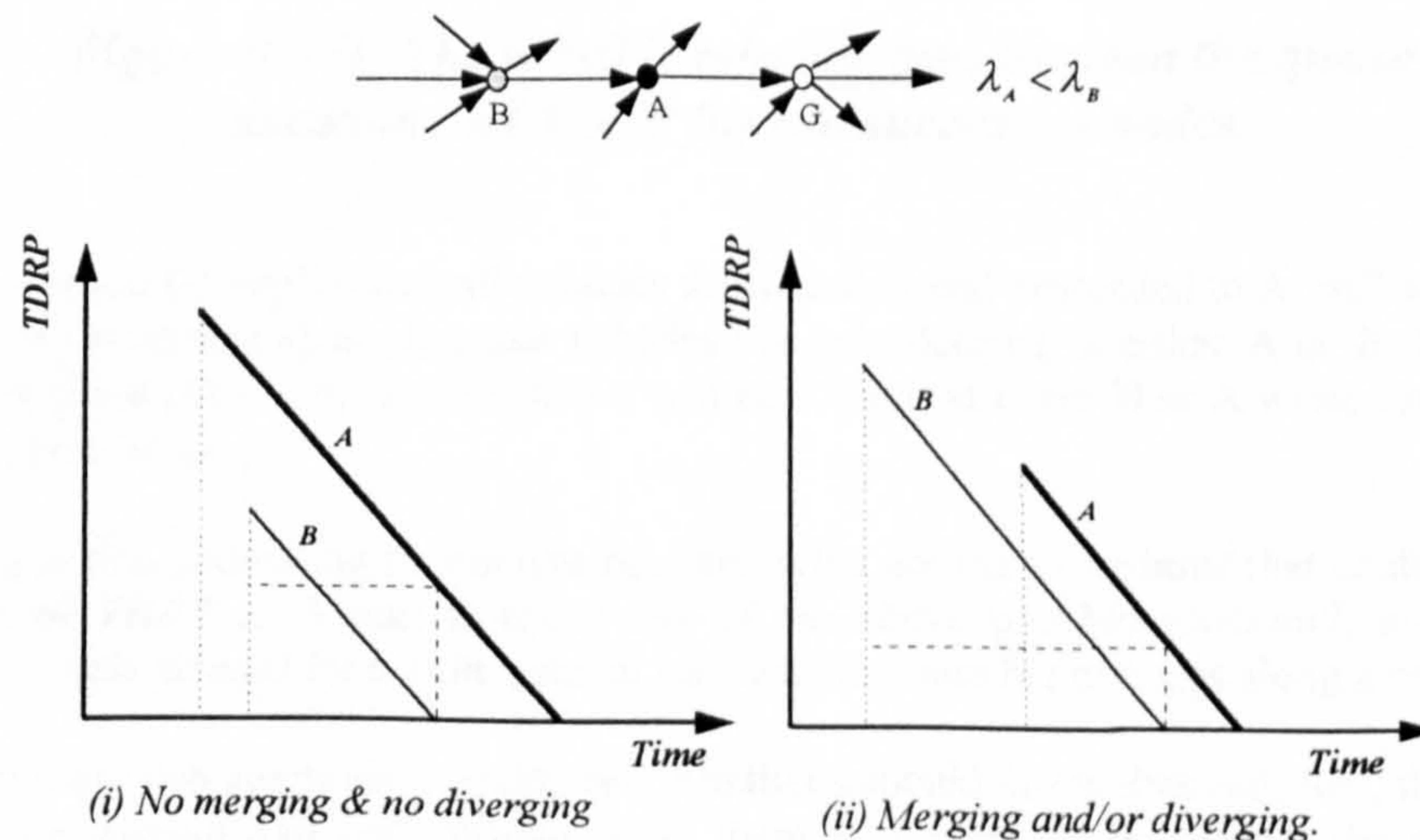


Figure (6.15): The effect of merging and diverging flows on the value of TDRP at nodes A and B.

However, unlike the case with no merging/no diverging, *TDRP* can not be charged all the time at node A alone or node B alone, but the two nodes and the interaction between them should be considered. Hence, some vehicle will be charged at only A or B and some others will be charged at both of them. Therefore, defining the critical node will no longer be useful as a basis for charging *TDRP* and the charge should be adapted at each individual node according to the interaction with the preceding node.

The procedures used to adapt the charges at each individual node along a single route with two nodes or more are considered in the following section.

6.5.3 Procedures for adapting the charges at each individual node

To adapt the charges at every node along a single route, consider first the case with only two nodes with merging and diverging flows. The queue duration for these two nodes could be presented on the same axis ignoring the time shift (i.e. travel time) between them; i.e. the departure time of any vehicle from the first node is considered to be the arrival time to the next one. This will depict the share of each vehicle in either or both of the two queues along the route. Typically, the relation between the queue duration for any two (successive) nodes along the route, will take one of the six hypotheses shown in Figure (6.16) below.

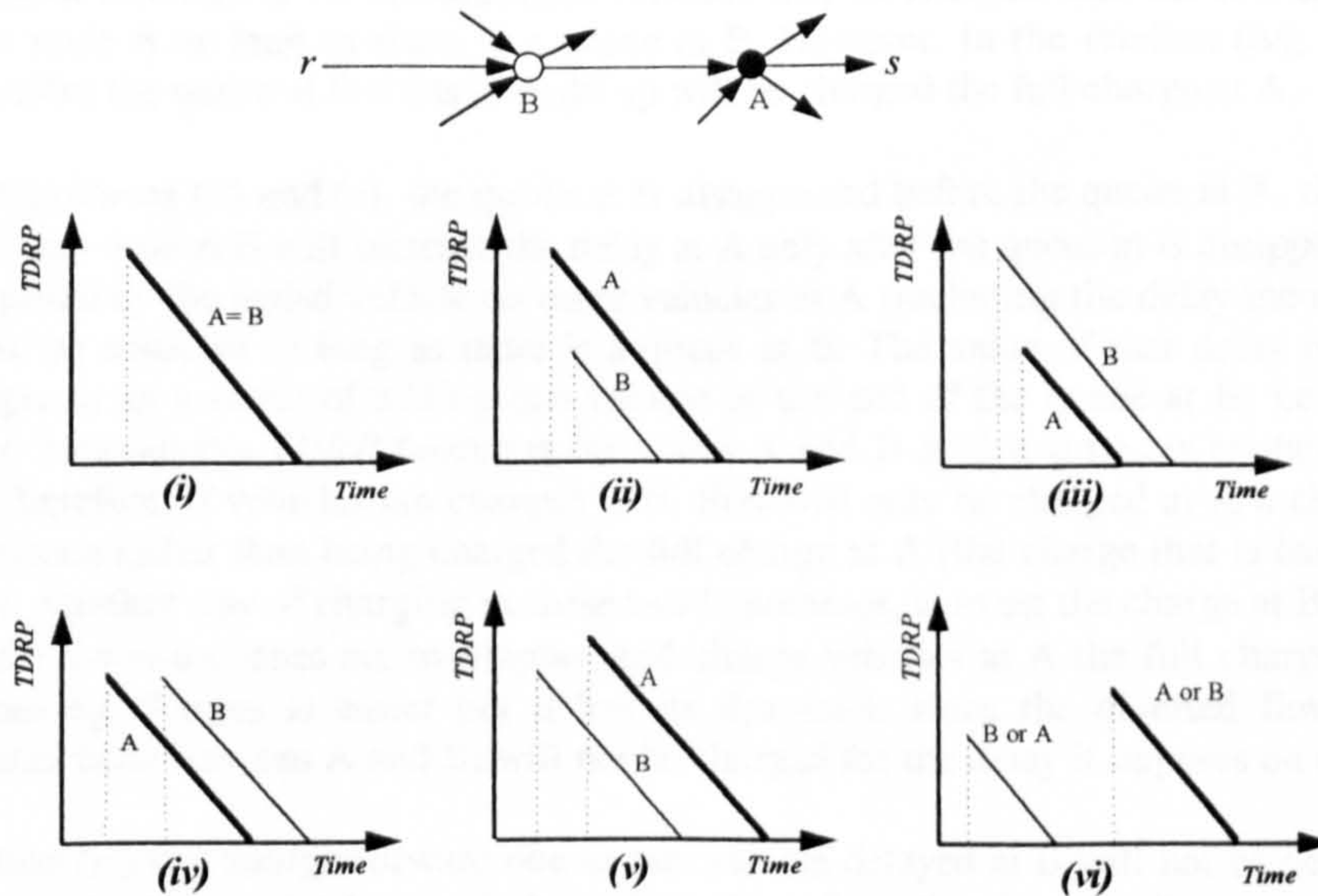


Figure (6.16): The possible relationships between the queue duration and *TDRP* for two successive nodes.

The first relation (i) implies that all vehicles delayed at B and proceeded to A, will also be delayed at A, while the relation (vi) implies that vehicles are only delayed at either A or B. In the other four hypotheses (from (ii) to (v)), some vehicles will be delayed at either B or A while some others will be delayed at both of them.

The two question addressing themselves now are: what are the procedures that could be used to give the value of *TDRP* at A and B under any of the above possible relations?, and whether these procedures could be used for a more general case with any number of nodes along a certain route?

To answer these two questions, two different situations should be emphasized; first, the diverging flow at B is time varying and not constant while there is a queue at B; second, the diverging flow is constant and equal to a fixed proportion of the flow departing from the queue at B. In the first situation, as a result of the diverging flow at node B, the flow proceeding from B to A, while there is a queue at B, will not be constant (unlike the case with no merging/no diverging flows). Therefore, the

queue at node A would behave independently from that at node B. Hence, vehicles would be charged also independently at both A and B in all different hypotheses shown in Figure (6.16). In fact, interaction between nodes, in this case, is not completely independent but it depends on the proportion of flow proceeding from one node to another. However, this is very difficult to investigate for a general case and hence it is worth expanding further.

On the contrary, if the proportion of the main stream flow diverted at B remains constant during the queue at B, then the main stream flow proceeding from B to A will also be constant. Hence, the queue at node A will be dependent on the queue at B and *TDRP* at A should be adapted accordingly. To adapt the *TDRP* at node A according to the interaction with node B, consider the case of two nodes along the route, *rs*, with each of the six hypotheses shown in Figure (6.16). For the relation (i), the two queues coincide with one another. The queue at A is mainly developed as a result of the interaction between the flow proceeding from B, the merging flow at A and the node capacity at A. In this case, the proceeding flow from B to A is fixed (equal to a certain proportion of the capacity at B) because the diverging flow at B is assumed to be constant. Thus, adding one vehicle at any time to the main stream at B will have no influence on the queue duration at A until the queue at B disappeared. Since the two queues disappeared at the same time, then every vehicle proceeding from the queue at B to node A will impose zero delay on all other vehicles behind at node A. Thus, while vehicles are charged the full charge at B, the charge at A is set to zero. Similarly, in the hypotheses (iii) and (iv) because the queue at B disappeared after the queue at A, then any added vehicle at any time at B will have no influence on the queue duration at A. Accordingly, vehicles will be charged also the full charge at B and no charge at node A as long as there is a queue at B. However, in the relation (iv), vehicles departing from A before the queue at B starts to build up will be charged the full charge at A.

For the hypotheses (ii) and (v), the queue at B disappeared before the queue at A, therefore any added vehicle at any time at B will increase the delay at A only after the queue at B disappeared, and the total delay imposed by the added vehicle on other vehicles at A (including the delay incurred by the vehicle itself) will be constant as long as there is a queue at B. The value of that delay is equivalent to the delay imposed as a result of adding one vehicle at the end of the queue at B; i.e. it is equal to the difference between the *TDRP* functions for nodes A and B at that time (as explained in some detail below). Therefore, if vehicles are charged at B, they will only be charged at A a charge equivalent to that difference rather than being charged the full charge at A (the charge that is based on the queuing duration). Another way of charging in these two hypotheses, is to set the charge at B to zero during the time where the two queues are overlapped and charge vehicles at A the full charge throughout. This way of setting charges is easier but it has its drawback since the diverted flow at B (or at any uncongested node between A and B) will not be charged for the delay it imposes on others at node B.

The relation (vi) is a straightforward one as any vehicle delayed at B will not be delayed at A because the two queues are completely time-independent. Therefore, the charge will be independently levied at each node.

Adjusting the charges for the two hypotheses (ii) and (v) is the most important one, therefore, it is explained in some detail below.

Adjusting the charges for the hypotheses (ii) and (v)

As illustrated in Figure (6.17), any vehicle departing from the queue at node B at time "O" imposes a delay on other vehicles behind equal to "OS". Ignoring the time offset between nodes B and A, then this vehicle is assumed to arrive at node A at the same time "O" and imposes a fixed amount of delay on others including itself equal to "H" as a result of considering the interaction between the two nodes. Assuming that the queue length at A at time "O" is *m* vehicles and the capacity of node A is λ_A vehicles per hour, then this vehicle will be delayed at A by m/λ_A and departs at time "P". Thus, the charge due on this vehicle at A will be equal to $(H - m/\lambda_A)$. This value is equivalent to the difference between "PY" and "OS" which are the charges based on the departure times at "P" and "O" respectively (see Figure (6.17)). Accordingly, observing the queue duration and the departure times at nodes A and B will be sufficient to calculate the value of *TDRP* without observing the queue length at

any time. However, the charge at node A is dependent on the charge paid at node B for all vehicles delayed at B, therefore, every vehicle should be identified at the charging point by a unique record with the charge paid at the preceding node. This implies that the charge to be paid at any node depends on the route chosen by the vehicle before approaching this node. In other words, vehicles arriving at node A from different routes, as shown in Figure (6.18), will have different values of charges even if they depart from that node at the same time. Thus, the route chosen by the vehicles will influence the value of charge paid at the same node at the same time of the day.

Hence, unlike the case of a single node along the route, the charge here can not be simplified to be fixed during a short time interval for all vehicles departing from the node during this time interval. This is because the charge does not only depend on the departure time from the node but also depends on the route chosen by the vehicle and the charges paid at the preceding node along that route.

In conclusion, as a result of considering the interaction between nodes, the charges at each node are demonstrated to be dependent on the route chosen by the vehicles. Therefore, the charges should be time varying on a vehicle-basis rather than a time interval-basis.

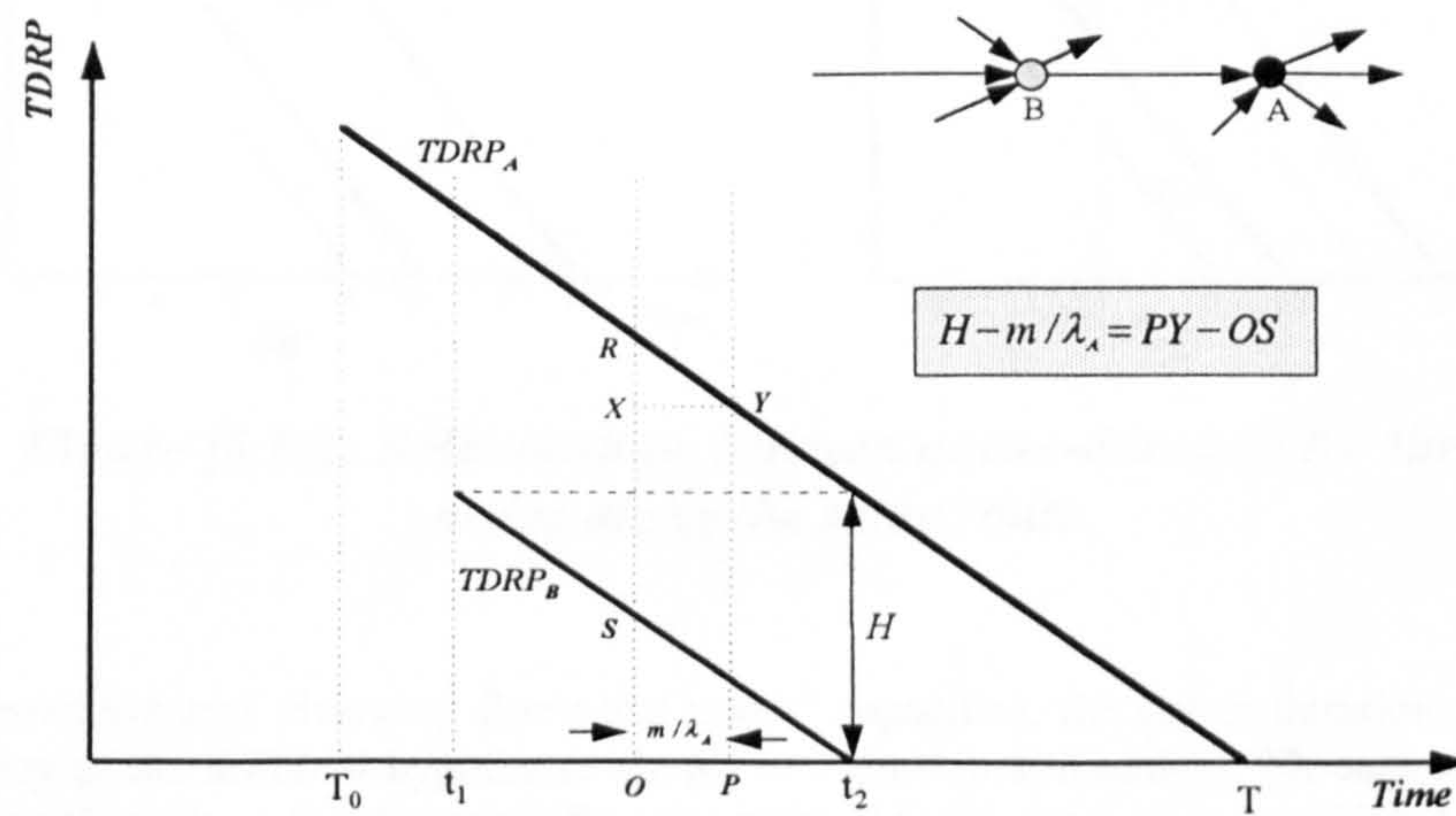


Figure (6.17): Adjusting TDRP at node A according to the node-interaction between nodes A and B.

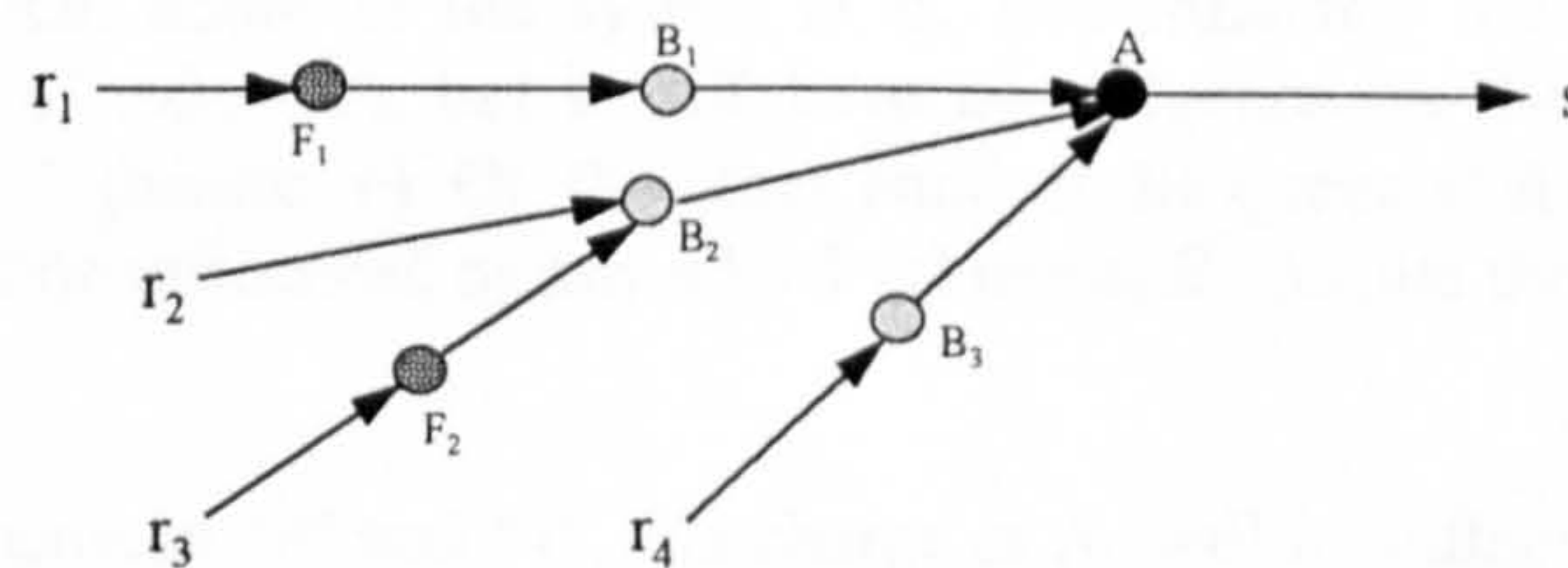


Figure (6.18): Different routes that could lead to node A.

The case with more than two nodes along the same route

The general case representing a real life situation is to have more than two nodes along the same route and most of them (or all) are congested. In this case, the relation between any two successive queuing duration (not necessarily occurring at two successive nodes) could be represented by any of the six hypotheses discussed earlier.

One might ask: is the node interaction restricted to only every two successive nodes or it extends to any pair of nodes (not necessarily successive) along the route chosen by the vehicles from origin to destination? To properly answer this question, consider these two illustrative examples shown in Figure (6.19) below.

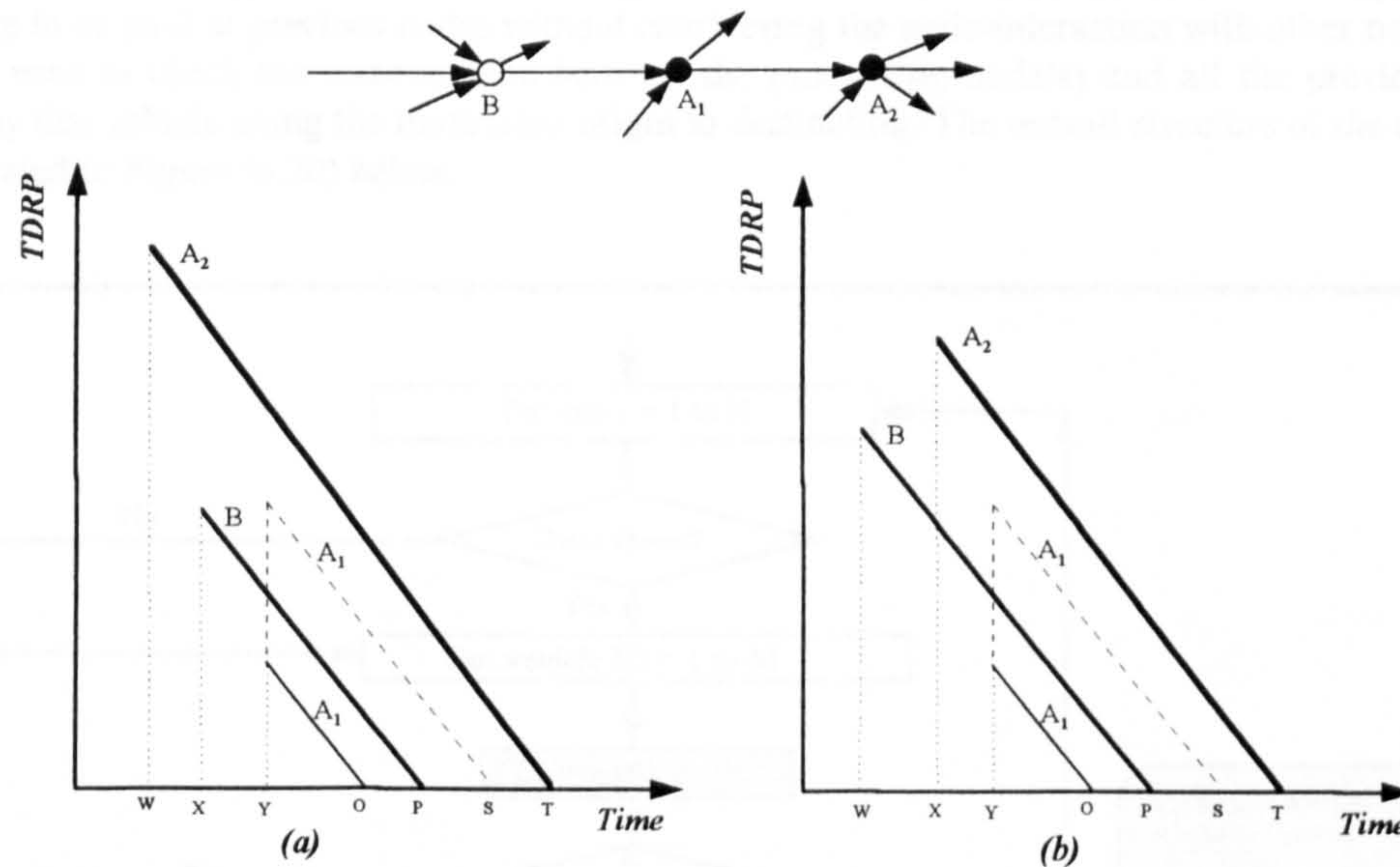


Figure (6.19): Relationships between queue-duration for three nodes along the same route.

Due to the merging and diverging flows and nodes' capacities, the queue duration at B, A₁ and A₂ could take any of the different hypotheses shown in Figure (6.19a and b). Throughout the discussion below the situation where the diverging flow is assumed to be constant is considered, since otherwise the queues at different nodes are independent.

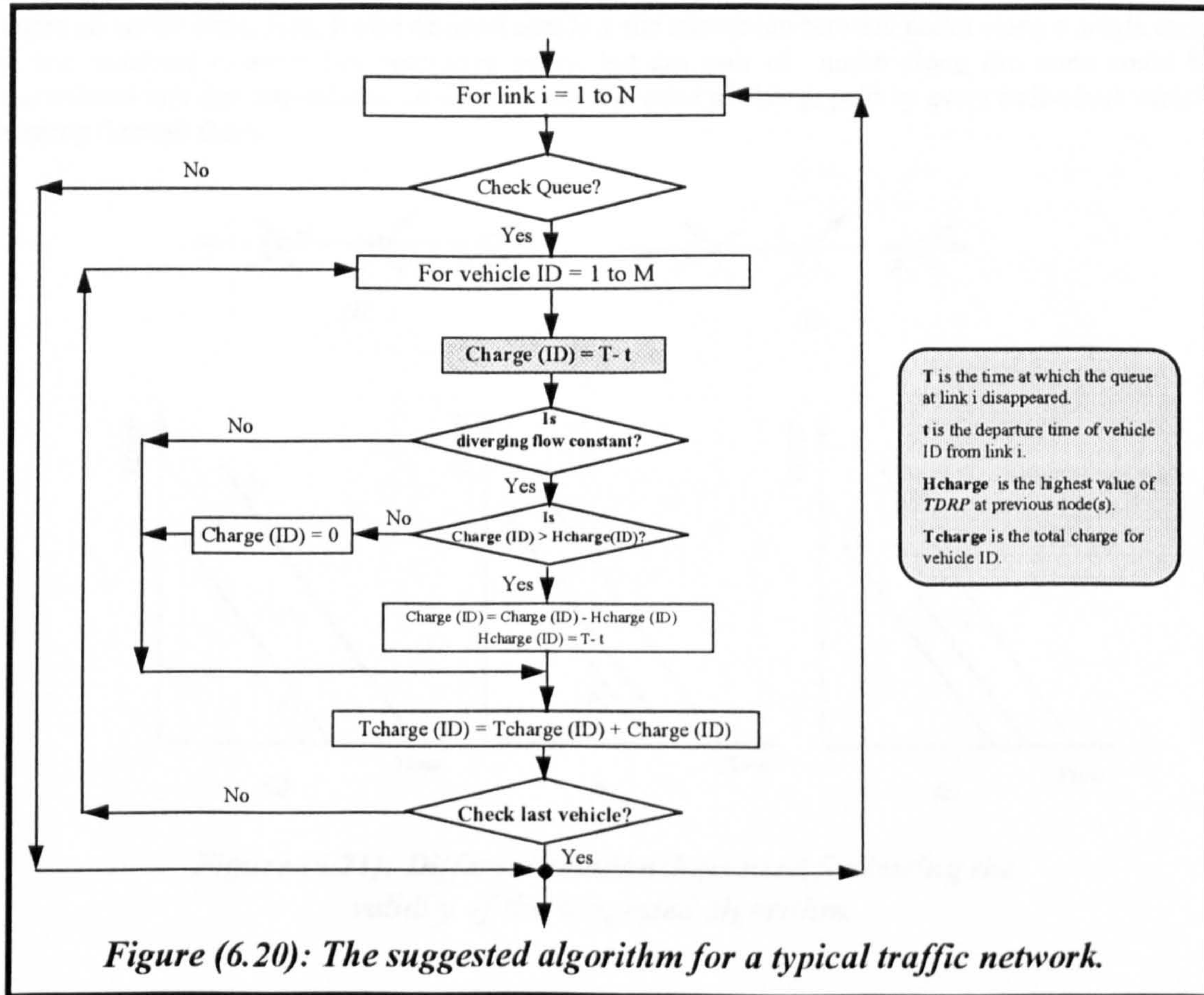
If the three queue duration at B, A₁ and A₂ are represented on the same horizontal axis ignoring the time offset between the three nodes (travel time between each two successive nodes), then the relation between the queue duration for the three nodes will be as shown in Figure (6.19a and b). As explained earlier, any vehicle added to the system at B, will influence the charge at A₁ if the queue there disappeared at "s" (after P), but it will have no influence on the charge at A₁ if the queue there disappeared at "O" (before P). On the other hand, if the queue at A₂ disappeared at "T" (after P), the charge at A₂ will be influenced by any added vehicle at B, despite the queue at A₁ which disappeared at "O".

Also if "s" lies between "P" and "T", the charge at A₂ will be influenced by that at A₁ between "Y" and "s", while the charge at A₂ between "X" and "Y" will be influenced by the charge at "B". This is because if there were no queue at A₁, the delay at A₂ will be influenced by that at B and once the queue starts to develop at A₁, the delay at A₂ will be influenced by that delay at A₁ as long as "s" lies between "P" and "T". In this case, A₂ will be more critical at that time than A₁, thus the total delay at A₁ and A₂ will be the same as the total delay at A₂ as if A₁ were not there; i.e. A₁ works only as a traffic regulator for the delay experienced by the vehicles proceeding from A₁ to A₂. On the other hand, if "s" lies after "T", A₁ will be more critical at that time than A₂ and the delay at A₂ will not be influenced by that at A₁.

These two simple examples demonstrate that the charge at any node is not only influenced by the queue duration at the node just before it, but it could also be influenced by the queue duration at any preceding node along the route chosen by the vehicles.

6.5.4 Suggested algorithm for a typical network

This section suggests an algorithm for a typical traffic network to define the *TDRP* for every individual vehicle at every node, taking into account the interaction between nodes. This algorithm is based on the information from network observation or from simulation results (i.e. queue duration at each node and the starting and vanishing times of each queue). It is assumed that while every vehicle is proceeding from origin to destination throughout the network, it keeps a record of the highest value of charge to be paid at previous nodes without considering the node-interaction with other nodes. This value is used to check the inter-relation between the proceeding node(s) and all the previous nodes passed by this vehicle along the route from origin to destination. The overall structure of the algorithm is illustrated in Figure (6.20) below.



Six different cases are used to check the validity of this algorithm and to demonstrate that it could be used to define the value of *TDRP* under any number of nodes. These six different cases are shown in Figure (6.21), which exhibits three different relations between queue duration at nodes B, A_1 and A_2 . These three relations are examined twice: first, node B is proceeded by node A_1 and then A_2 ; second, node A_1 is proceeded by node A_2 and then node B.

Consider first case (1) where B is proceeded by A_1 and then A_2 . For Figure (6.21a), the algorithm gives a full charge at B, and the charge at A_1 is equal to the difference between the full charges at nodes A_1 and B. Whether the queue at A_2 disappeared before that at node B or between that at nodes B and A_1 , no charges is levied at A_2 , i.e. the charge given by the algorithm at A_2 is equal to zero. For Figure (6.21b), a full charge is levied at node B. The charge at A_1 is set to zero, while the charge at A_2 is equal to the difference between the full charges at A_2 and B. For Figure (6.21c), a full charge is also levied at node B, while the charge at A_1 is equal to the difference between the full charges at nodes A_1 and B. The charge at A_2 for all vehicles departing from A_2 before the queue at A_1 starts to build up, is

equal to the difference between the full charge at A_2 and the charge already paid at B. Once the queue starts to build up at A_1 , the charge at A_2 will be equal to the difference between the full charges at A_2 and A_1 , i.e. the difference between the full charge at A_2 and the sum of the charges already paid by every individual vehicle at nodes B and A_1 (which is equal to the full charge at A_1).

For case (2) where A_1 is preceded by A_2 and then B, the three situations shown in Figure (6.21) are also considered. For Figure (6.21a), a full charge is levied at node A_1 , zero charge at node A_2 , and a full charge at node B only before the queue at A_1 starts to build up and zero charge afterwards. For Figures (6.21b) and (6.21c), a full charge is also levied at A_1 , while the charge at A_2 is equal to the difference between the full charges at nodes A_2 and A_1 . At node B, vehicles are charged a full charge only before the queue at A_2 starts to build up and zero charge afterwards.

This interpretation demonstrates that the algorithm is valid for dealing with any number of nodes under all traffic conditions. It also demonstrates that the interaction between nodes along a single route is not restricted to every two successive nodes, but any pair of nodes along the route could be interrelated and this interrelation could influence the value of charge paid by every individual vehicle passing through them.

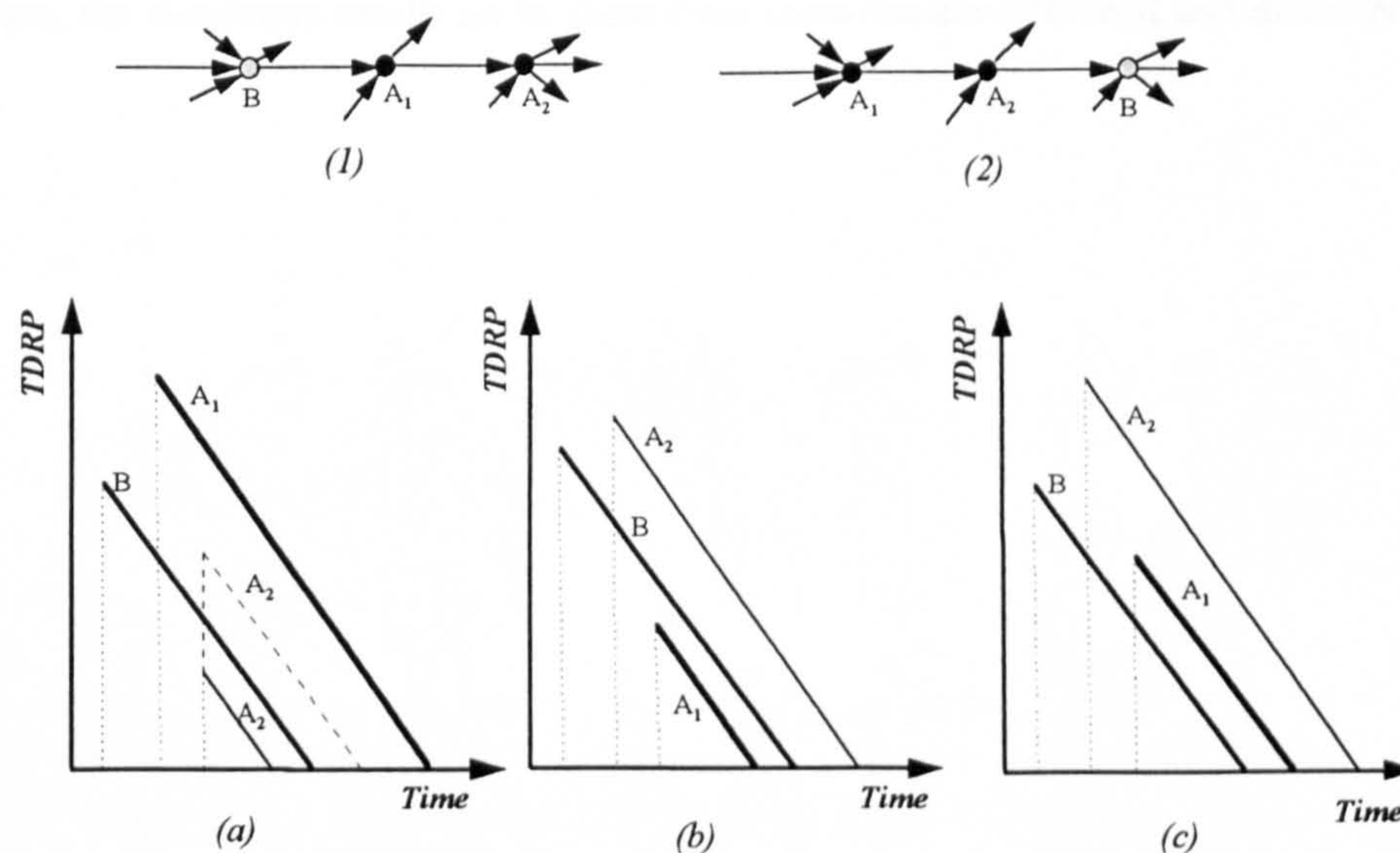


Figure (6.21): Different relationships used for testing the validity of the suggested algorithm.

6.5.5 Summary and implications for practical applications

The interaction between nodes is of greatest importance when the nodes concerned have no merging/no diverging flows or with fixed diverging flows throughout the study period. However, as these conditions change, the importance of interaction between nodes declines and it becomes very difficult to investigate. In this case, the interaction between nodes (if any) will depend on the proportion of the flow proceeding from one node to another. The general algorithm suggested, in this chapter, for modelling the interaction between nodes under all circumstances did not consider the interaction between nodes when the diverging flow is not constant. Therefore, the issue of node interaction has not been completely resolved yet and more research is still needed in this area.

It is also demonstrated, in this chapter, that interaction between nodes along the route is not restricted to every two successive nodes but any pair of nodes along the route could be interrelated and this interrelation could influence the value of charge paid by every individual vehicle passing through them. Therefore, the charges paid at a certain node will be dependent on the route chosen by the vehicles before approaching this node. As a consequence, it is necessary to define the route chosen by

every individual vehicle from origin to destination. This can be done by giving each vehicle a unique identification number "ID" and being traced throughout the network. Although this is quite possible to be done in practice using the available technology of Electronic Road Pricing *ERP*, it is computationally laborious and this will increase the total cost of running the system.

To apply the system of *TDRP* in practice with less costs and less complications, particularly when there is no merging/no diverging flows or when the diverging flow at each node remains constant, only the most critical nodes (nodes with very long queues) could be identified for charging. For a big network, the delays at the most critical nodes are less likely to be inter-related (or the most critical nodes to be charged could be chosen such that this condition is satisfied). Thus, the main principle of *TDRP* could be applied for a limited number of non-correlated (or almost non-correlated) nodes throughout the network. However, charging some nodes may overestimate the strict optimum charges of *TDRP* if the nodes chosen for charging are correlated with each other. It also may fall short of achieving the optimal benefits from the system.

Since the interaction between nodes is difficult to investigate in reality for a typical network under general traffic conditions, three different scenarios for *TDRP* are suggested in the next chapter: charging all nodes, charging only the entrances to the network, and charging only the most congested nodes. Then, the simulation results under these three scenarios are examined and discussed in some detail.

CHAPTER SEVEN

7. Evaluating the different effects of *TDRP* on a traffic network

7.1 Introduction

7.2 Description of the case study used for evaluation

7.3 The base case simulation results

7.4 The effects of congestion on the base case results

7.5 The different effects of *TDRP* on the travel conditions

7.5.1 The different effects of *TDRP* under scenario (I).

7.5.2 Comparison between the different *TDRP* scenarios.

7.6 Summary

7.1 Introduction

Having discussed the different effects of *TDRP* on a single traffic bottleneck along a route and developed the network simulation model in previous chapters, the aim of this chapter is to exhibit the potential benefits of applying *TDRP* for a traffic network in urban area. To avoid the hardship and the complexity of the real life data, a typical network and a typical flow demand for an urban area are therefore chosen for conducting the simulation. The different characteristics of the chosen network is investigated under different demand levels (i.e. different level of congestion). Then, the potential benefits of different scenarios of *TDRP* are investigated under different traffic conditions.

7.2 Description of the case study used for evaluation

The traffic network as well as the traffic demand and the traffic parameters used for conducting the simulation throughout this chapter are discussed each in turn below.

a. *Traffic Network*

The network used in this study has a similar configuration to the test network designed to demonstrate the use of the facilities in CONTRAM (see Leonard and Gower, 1982). This particular network is chosen as it contains many of the network conditions found in urban areas. The network represents a small town with a main through route which passes through the centre of the shopping area as depicted in Figure (7.1) below. There are two alternative routes around the town: a fast, but longer, ring road and a back route, through a residential area, on which there are two give-way junctions. The network contains a roundabout, four signal-controlled junctions— one of which is a pedestrian signal, and a car park.

However, the original configuration of CONTRAM test network is modified to suit the limitations of the simulation model and to demonstrate the different effects of *TDRP*. The modified network is depicted in Figure (7.2). Because the simulation model does not deal with signal-controlled junctions, all junctions are changed to priority junctions (give-way junctions or roundabout), while the pedestrian signal is removed for simplicity. To give vehicles more alternatives for route choices, when *TDRP* is introduced, the ring road is extended to go around the residential area.

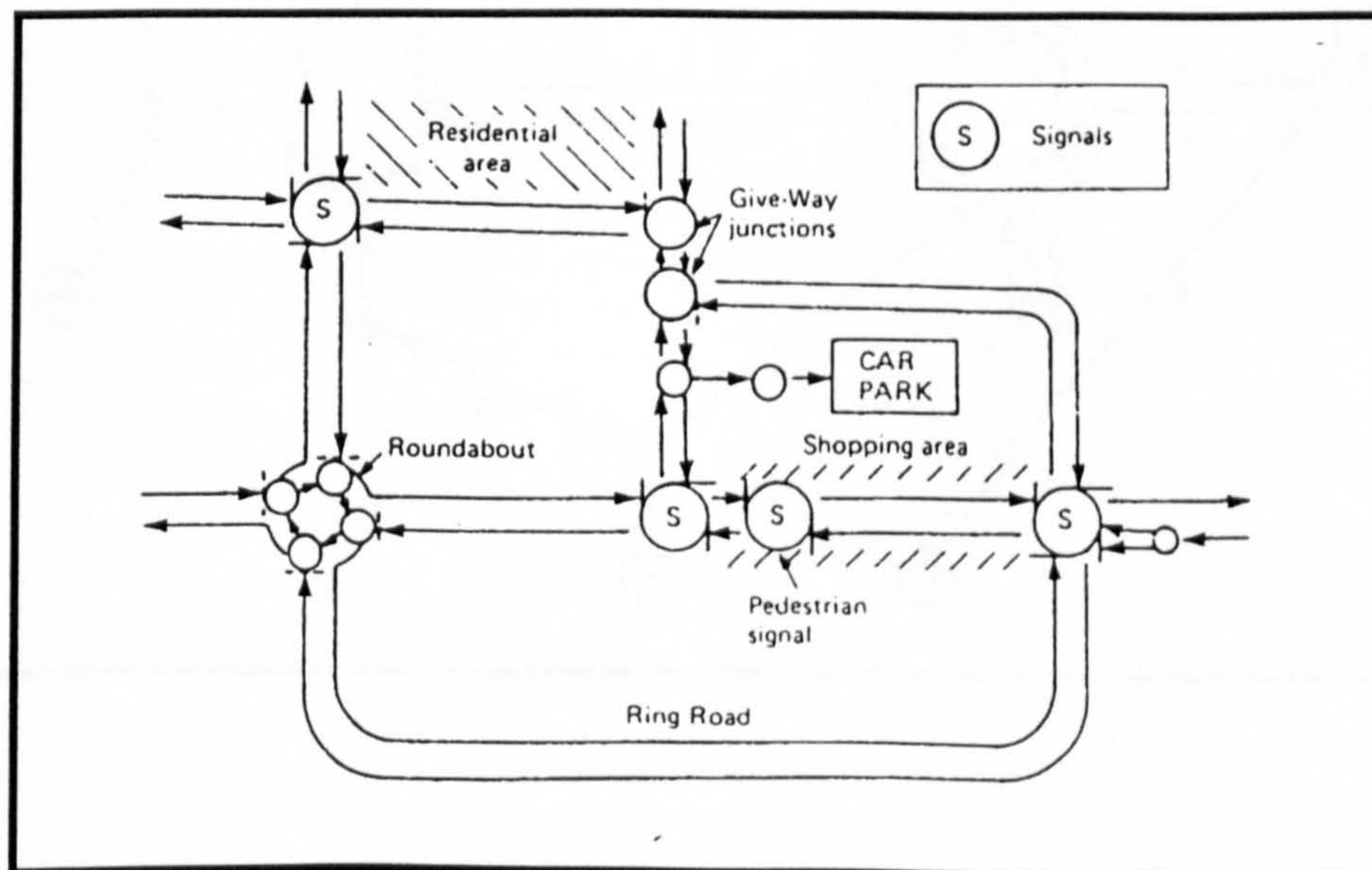


Figure (7.1): CONTRAM test network.

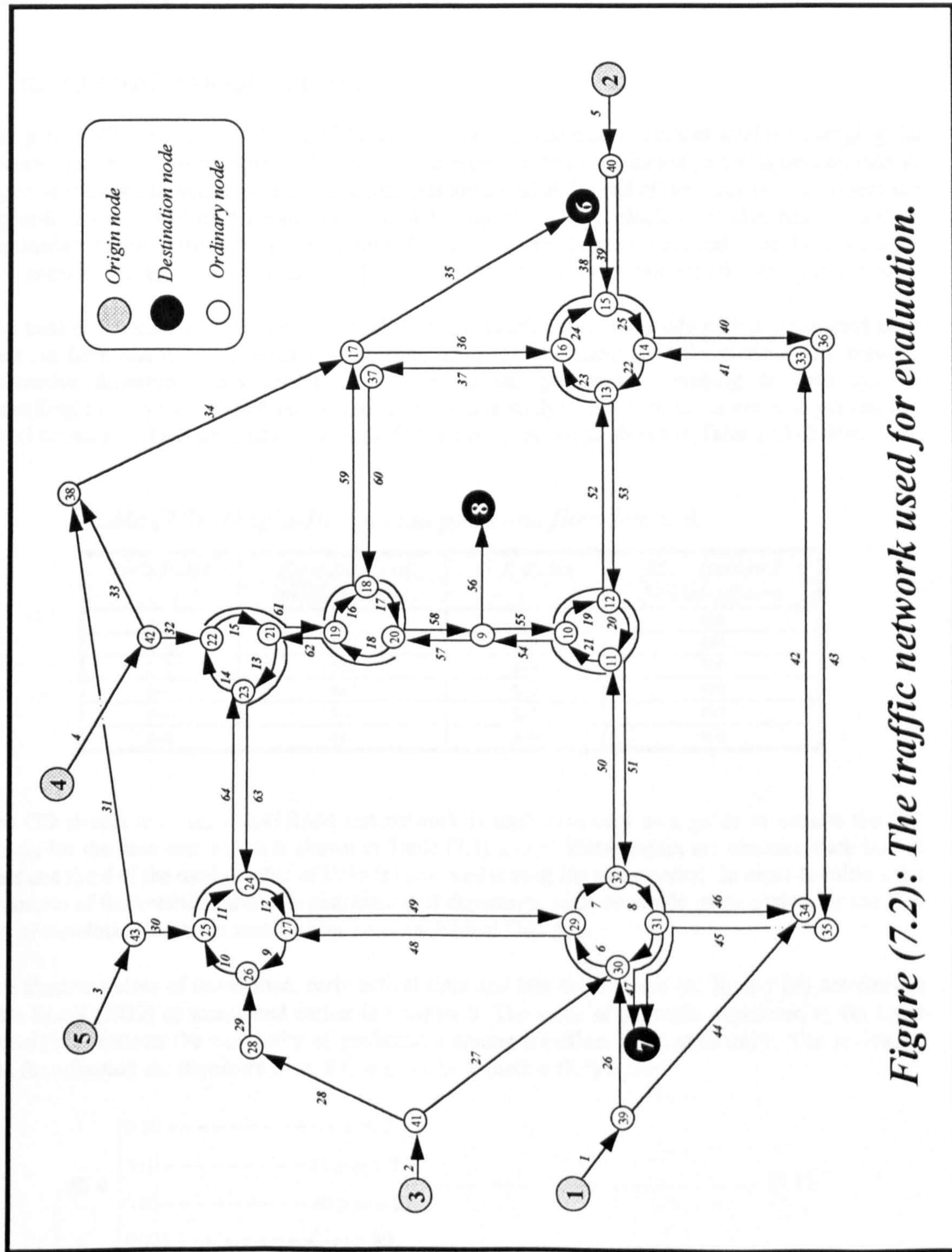


Figure (7.2): The traffic network used for evaluation.

The network has five origins and three destinations. The total number of nodes (including origin and destination nodes) is 43 and the total number of links and turns is 64. The lengths and capacities of almost all the links are adapted to go on line with the different demand levels used in the study. A full detail of the network data is given in Appendix (B).

b. Traffic Demand and traffic parameters

The period of the day analysed is [05:00 a.m.–10:00 a.m.]. The time increment used for changing the demand and the departure time is 10 minutes. Therefore, the total simulation period is divided into 30 equal time intervals, and three more time intervals are added at the end of the study period to clear the network at the end of each iteration (i.e. to make sure that every vehicle or packet has reached its destination before starting the next iteration). The desired arrival time is assumed to be the same at all destinations. The base case desired arrival time is 09:00 a.m. with no work time flexibility (i.e. $\Delta=0$).

The total demand for each O-D pair throughout the network during the study period is assumed to be constant from one day to another (i.e. inelastic demand), and users have the choices only between alternative departure times and/or routes (i.e. neither generation, changing destinations, nor cancelling journeys are considered in the scope of this study). Network users are also assumed to travel between 12 O-D pairs (the other three O-D pairs are zeros) as shown in Table (7.1) below.

Table (7.1): Origin-Destination pairs and flow demand.

<i>O-D Pairs</i>	<i>Flow Demand (vehicles/hour)</i>	<i>O-D Pairs</i>	<i>Flow Demand (vehicles/hour)</i>
1-6	720	4-6	480
1-8	480	4-7	480
2-7	540	4-8	360
2-8	480	5-6	480
3-6	540	5-7	480
3-8	480	5-8	360

The OD matrix used for CONTRAM test network is used here only as a guide to assume the OD matrix for the base case which is shown in Table (7.1) above. These values are assumed such that at least one third of the total number of links is congested during the study period. In order to initiate the dynamics of the system, a uniform distribution of departures over the whole study period for the first day of simulation has been specified for every individual O-D pair.

The shadow values of travel time, early arrival time and late arrival time (α , β , and $\beta\gamma$) are derived from Small (1982) as mentioned earlier in Chapter 5. The value of the scale parameter, η , (in Logit Model) that reflects the variability of preferences among travellers is equal to unity. The reviewing rate for adjusting the departure time, R_1 , is given by Equation (7.1) below.

$$R_1 = \begin{cases} 0.20 & \text{----- } \omega \leq 7 \\ 0.10 & \text{----- } 21 \geq \omega > 7 \\ 0.05 & \text{----- } 40 \geq \omega > 21 \\ 0.025 & \text{----- } \omega > 40 \end{cases} \quad \dots\dots\dots (7.1)$$

These different values of R_1 are chosen to help the system to converge to a stationary state and to reduce the number of iterations needed for convergence. However, using differential values for R_1 has a practical implication because at the early stages of implementing any new policy, the percentage of travellers reviewing their trip decisions is expected to be high and decreases eventually as the system approaches its final state.

The simulation terminates if the weighted coefficient of variation of the generalized cost for all OD pairs is less than 5%, or after 60 iterations (whatever reaches first).

These initial conditions are used in most of the simulation conducted herein. Other specifics that have been varied across simulation are discussed later (in the following chapter) in conjunction with the experimental factor under investigation.

7.3 The base case simulation results

This section represents the results of the base case which plays the role of a reference point for the simulation experiments. The base case simulation experiments begin with no-congestion pseudo-equilibrium departure rate distribution for every individual OD pair. The simulation experiment is terminated after 60 iterations. As depicted in Figure (7.3), the fluctuation of the weighted coefficient of variation WCV almost settles down (with some oscillations) after 45 iterations. The average value of the WCV for the last ten iterations is equal to 0.16. Figure (7.3) also exhibits that the average speed reaches its final state (of equilibrium) with some insignificant oscillations after only 15 iterations. This is because the route choice and consequently travel time and speed are expected to settle down before the departure time choice. Since the departure time choice significantly affects the schedule delay costs and consequently the generalized costs, therefore the WCV could take much longer to settle down than the average speed.

The base case departure rate distribution, travel time costs, schedule delay costs and generalized costs for all movements are depicted in Figure (7.4). These values are the average of the last five days of simulation after they have been weighted by the number of vehicles departing during each time interval throughout the study period. The GC distribution is not consistent with Wardrop's first principle adjusted for the time dependent trip characteristics, since in contrast to Wardrop's deterministic approach, a stochastic one is adopted in this study. In other words, because of the stochastic element in the "Logit" model (used for the departure time choice) and the interaction between different OD pairs, the resulting GC distribution during the non-zero departure period is not linearly horizontal. (It should be noted that the GC as well as the GC components for the periods of zero departure are the average values for the last vehicles departing from these intervals before they are allocated to another time intervals; and of course, they are not weighted since the departure rate is zero). The distribution of the departure rate depicted in Figure (7.4) is not normally distributed over the period of non-zero departure, but its peak is skewed towards early departures. The reason for that is because of the higher disutility of late schedule delay (over two times the disutility of early schedule delay), therefore, travellers try to avoid late arrivals to their destinations by shifting to earlier departures. Figure (7.4) also shows that the maximum travel time cost is not associated with the peak departure rate but it is shifted slightly to the right. This is because travellers departing just after the peak departure rate are expected to experience the maximum travel time and delay. The reason that the schedule delay cost is not zero at the desired arrival time (i.e. at 9:00 a.m.) is because, as mentioned earlier, the values presented in this figure are the average costs for all movements throughout the network; i.e. while the schedule delay cost for some movements is zero, for others departing during the same time interval it is higher than zero. This is because the schedule delay cost for travellers departing during the same time interval depends on the distance between the origin and the destination of every individuals.

The departure pattern and the GC distribution for individual OD pairs as well as for all movements from individual origins are depicted in Figures (7.5) and (7.6). Figure (7.5) shows the WGC versus the departure rate for the three OD pairs: 2-7, 3-6 and 4-8, while Figure (7.6) shows the same results for all movements from origins: 1, 2 and 5. For all movements, the WGC distribution seems to be flat during the period of the peak departure rate and then starts to increase afterwards. The resulting departures from all origin take place between 6:25 a.m. and 9:15 a.m. The timing of the peak departure rate for individual OD pairs depends on the distance between origin and destination. As this distance increases, the peak departure rate takes place much earlier and vice versa (as depicted in Figure (7.5) for O3-D6 and O4-D8).

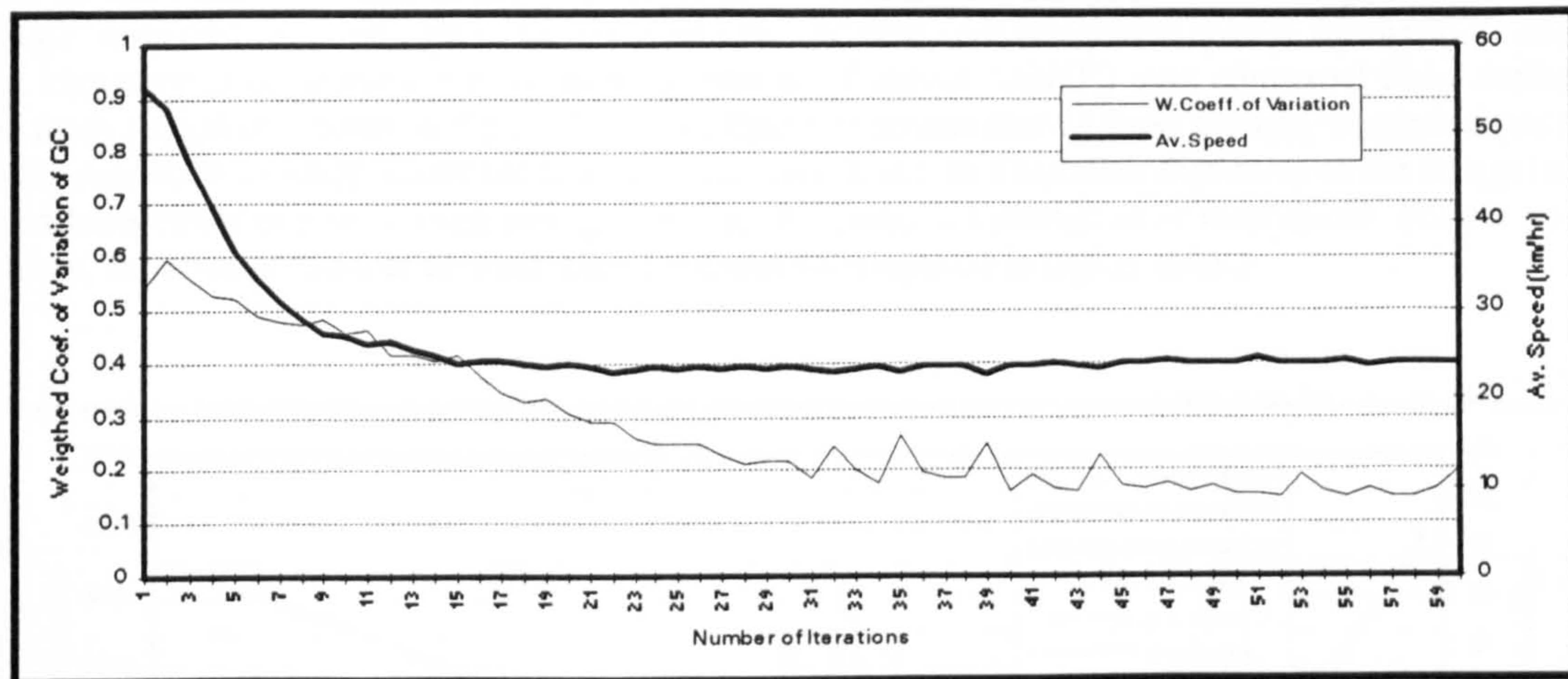


Figure (7.3): The Fluctuation of the Average Speed and the Weighted Coefficient of Variation of G. Cost under the Base Demand.

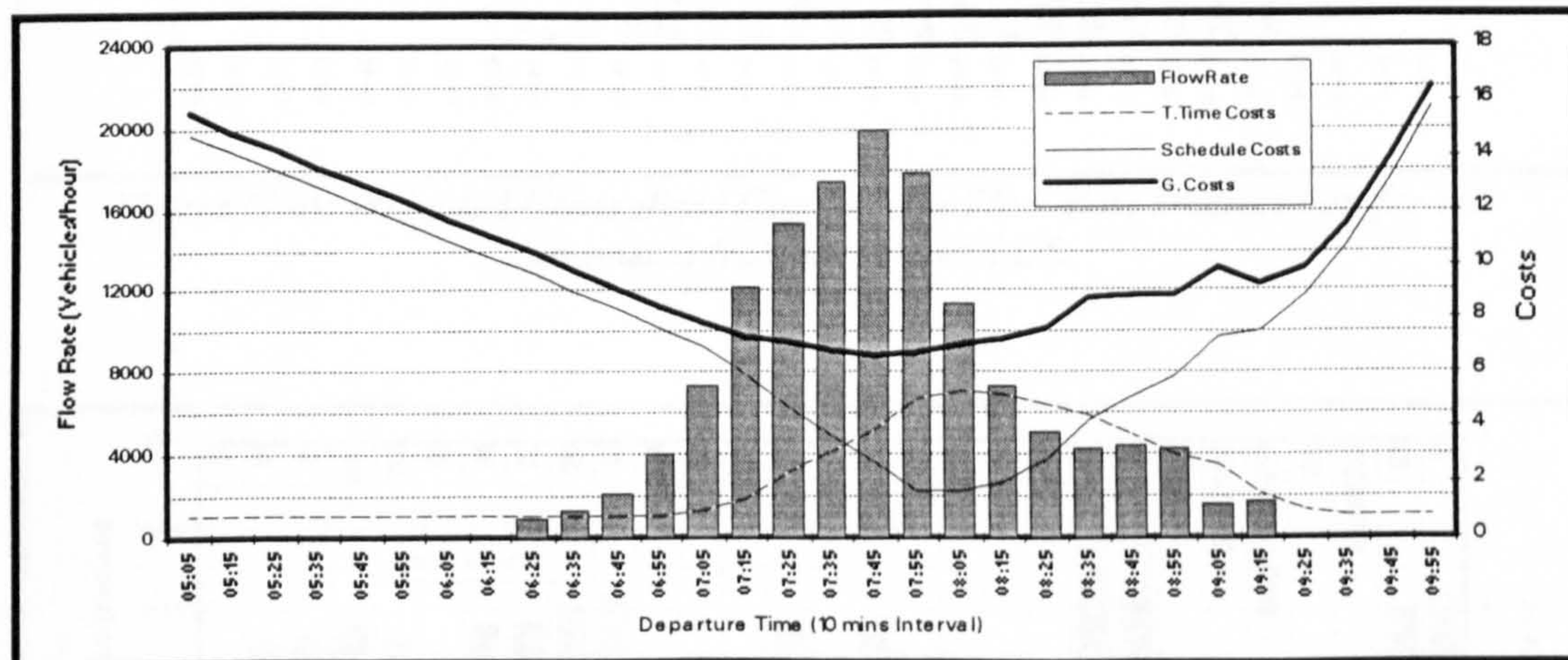


Figure (7.4): Weighted G. Cost and W. G. Cost Components versus Flow Rate Distribution for all movements under the base demand.

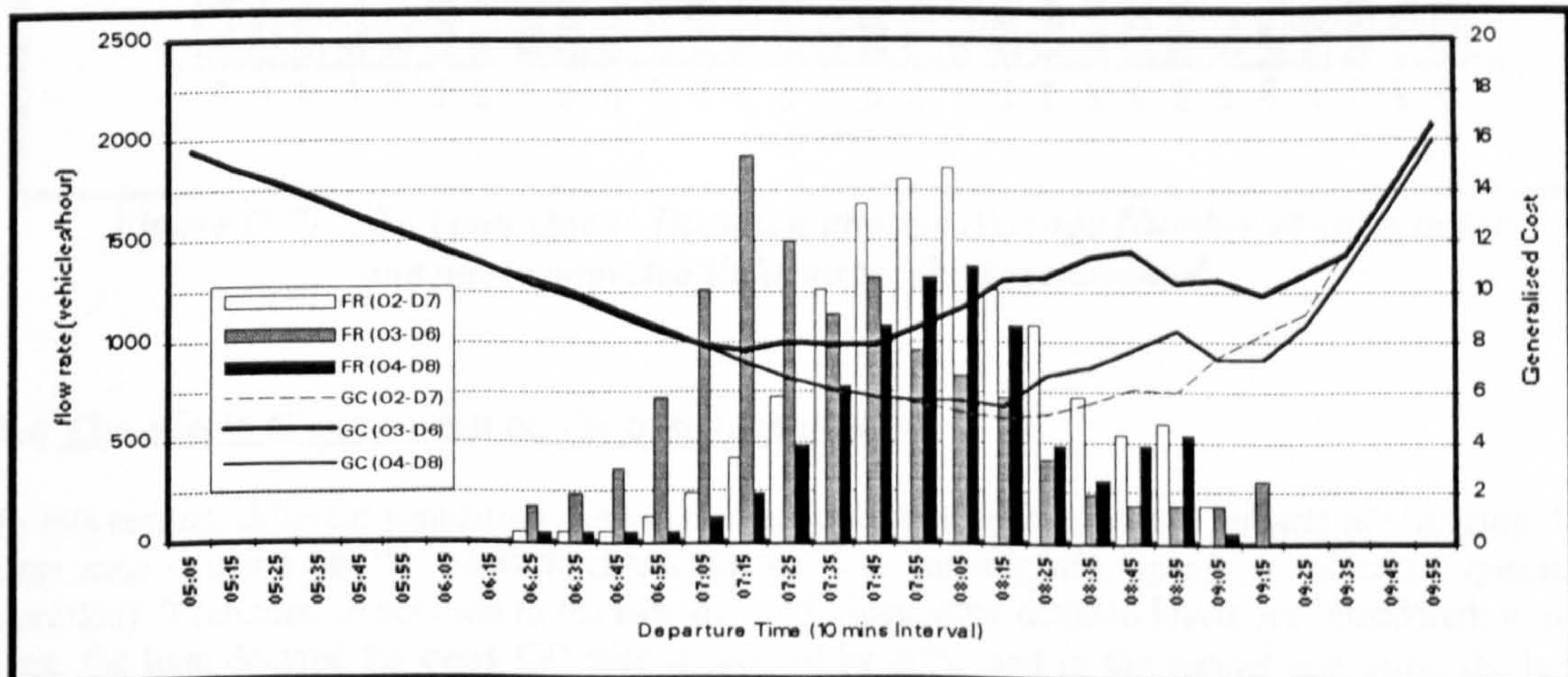


Figure (7.5): Average Generalized Cost versus Flow Rate Distribution for O2-D7, O3-D6 and O4-D8.

One of the most important measure of congestion, used in this study, is the total queue duration "TQD" over each link throughout the study period. The base case simulation results for the TQD and the frequency of occurrence (i.e. the average number of queues "ANQ") over congested links during the study period are shown in Figure (7.7). The figure illustrates that 24 links are very congested (with TQD over 4000 seconds) under the base demand level. It is also illustrated that most of the congested links have either one or at most two queues (i.e. the queue is interrupted at most once). For shorter queues, the average number of occurrence (or queue interruption) is higher than 2.

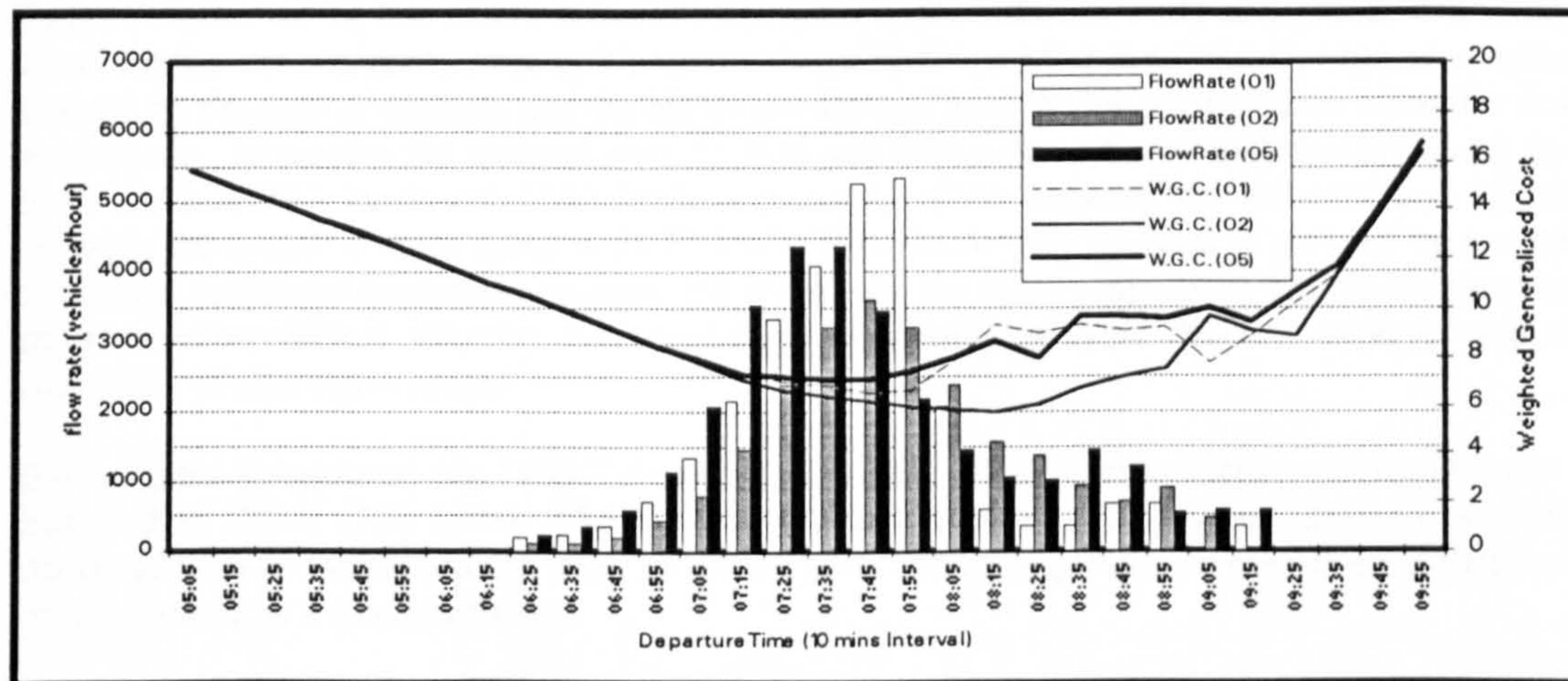


Figure (7.6): Weighted Generalized Costs versus Flow Rate Distribution for all movements from O1, O2 and O5.

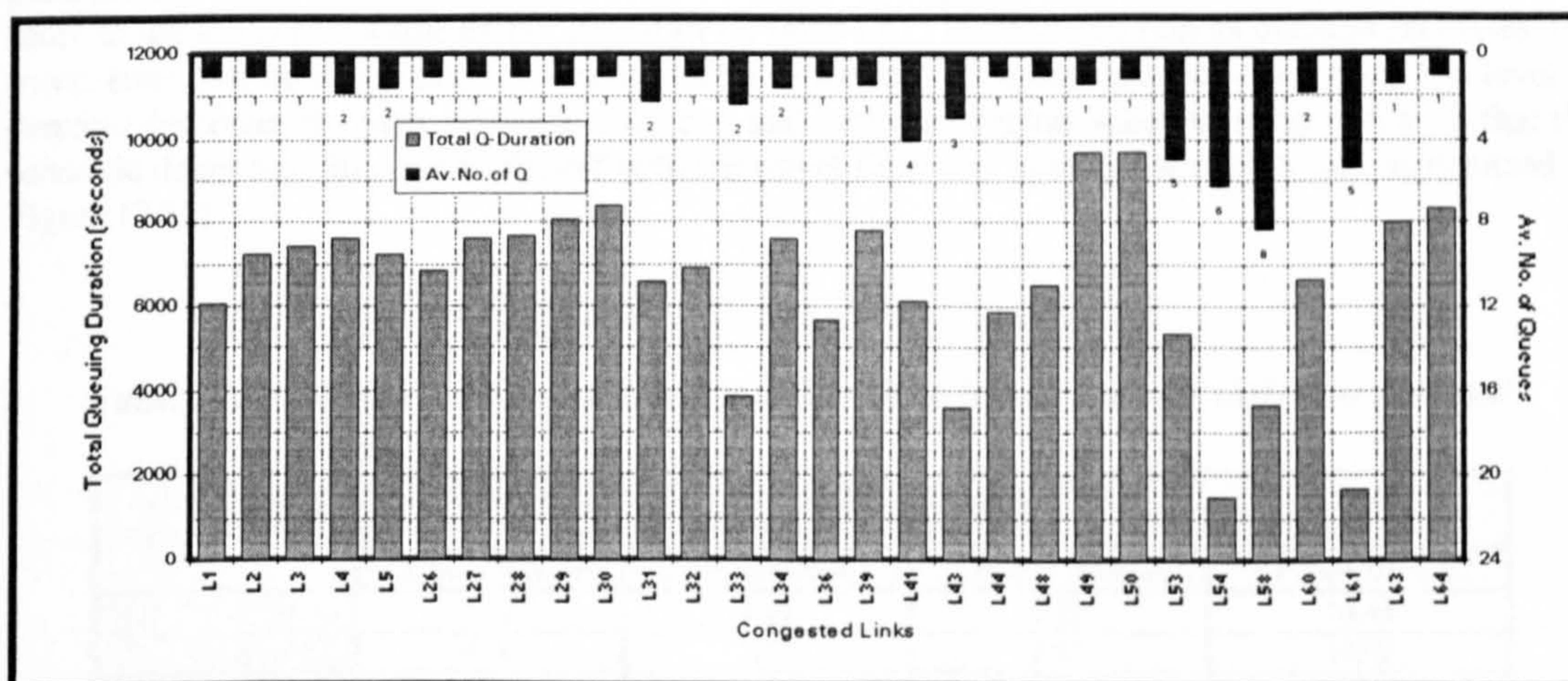


Figure (7.7): The Total Queue Duration and the Average Number of Queues for the most congested links under the base demand.

7.4 The effects of congestion on the base case results

In this section, different simulation experiments are conducted to explore the impacts of changing the base case demand on the network characteristics (i.e. convergence, speed, travel costs, queuing duration). Therefore, in addition to the base demand, three other demand levels are considered: in the first, the base demand for every OD pair is reduced by 20%, and in the second and third, the base demand for every OD pair is increased by 20% and 40% respectively.

The fluctuation of the WCV of the generalized cost under different levels of demand is presented in Figure (7.8). The Figure shows that under different levels of demand, the system converges to a stationary state with some oscillations, and the WCV for 0.8 base demand (i.e. 0.8 BD) is slightly higher than the others at equilibrium. This could be because as the demand level decreases, the travel time cost also decreases and it would become difficult for the schedule delay cost function and the travel time cost function to trade-off with a small coefficient of variation at equilibrium. In Figure (7.8), although the 1.4 BD shows the lowest WCV, the difference between its value and that for the BD and 1.2 BD is not significant.

Figure (7.9) shows the fluctuation of the average speed under different levels of demand. The average speed reaches its stationary state (with very little oscillations) after a certain number of iterations depends on the level of demand, and the higher the demand level the longer it takes to reach this state. Furthermore, increasing the demand level by 20% and 40% reduces the overall average speed from 24.02 km/h to 20.8 km/h and 17.82 km/h respectively, while reducing the demand level by 20% increases the overall average speed to 28.62 km/h. Thus, increasing the level of demand by a certain percentage might not necessarily decrease the speed with the same proportion. This is because there are some other factors affecting the speed-flow relationship such as the link capacity and the availability of alternative routes.

The weighted generalized cost 'WGC' for every individual OD pair under different demand levels is presented in Table (7.2) below. The table shows that increasing the overall demand is always associated with an increase in the WGC for all OD pairs and consequently the overall generalized cost for the network as a whole increases.

The WGC and the WGC components versus the departure rate distribution under different demand levels are depicted in Figures (7.10), (7.11) and (7.12). As discussed earlier, the WGC has a similar pattern (U-shaped pattern) and as the demand level increases the WGC increases and its pattern becomes more flatter. Figure (7.10) shows that increasing the level of demand does not necessarily increase the departure rate in all time intervals, but the departure rate distribution would be shifted more to the left (i.e. towards earlier departure). Figure (7.11) also shows that as demand increases, the travel time cost increases and its peak is shifted to the left. As a consequence, increasing the level of demand decreases the early schedule delay cost and increases the late schedule delay cost such that the schedule delay cost function trades-off with the travel time cost function at equilibrium as depicted in Figure (7.12).

Table (7.2): Weighted Generalized costs for all O-D Pairs under different demand levels.

O-D Pairs	Demand Levels			
	0.8 Base Demand	Base Demand	1.2 Base Demand	1.4 Base Demand
1-6	5.72	7.10	8.37	10.48
1-8	6.40	7.57	8.82	10.92
2-7	4.19	5.80	7.84	9.93
2-8	5.71	7.35	9.35	11.51
3-6	6.70	8.61	10.25	11.52
3-8	6.81	8.55	10.43	11.67
4-6	4.29	6.22	6.97	8.32
4-7	6.13	7.49	8.41	10.11
4-8	4.83	6.34	7.36	8.66
5-6	5.34	7.31	8.39	10.44
5-7	6.33	8.03	9.48	11.66
5-8	6.74	8.36	10.05	12.13
Average GC	5.77	7.39	8.81	10.61

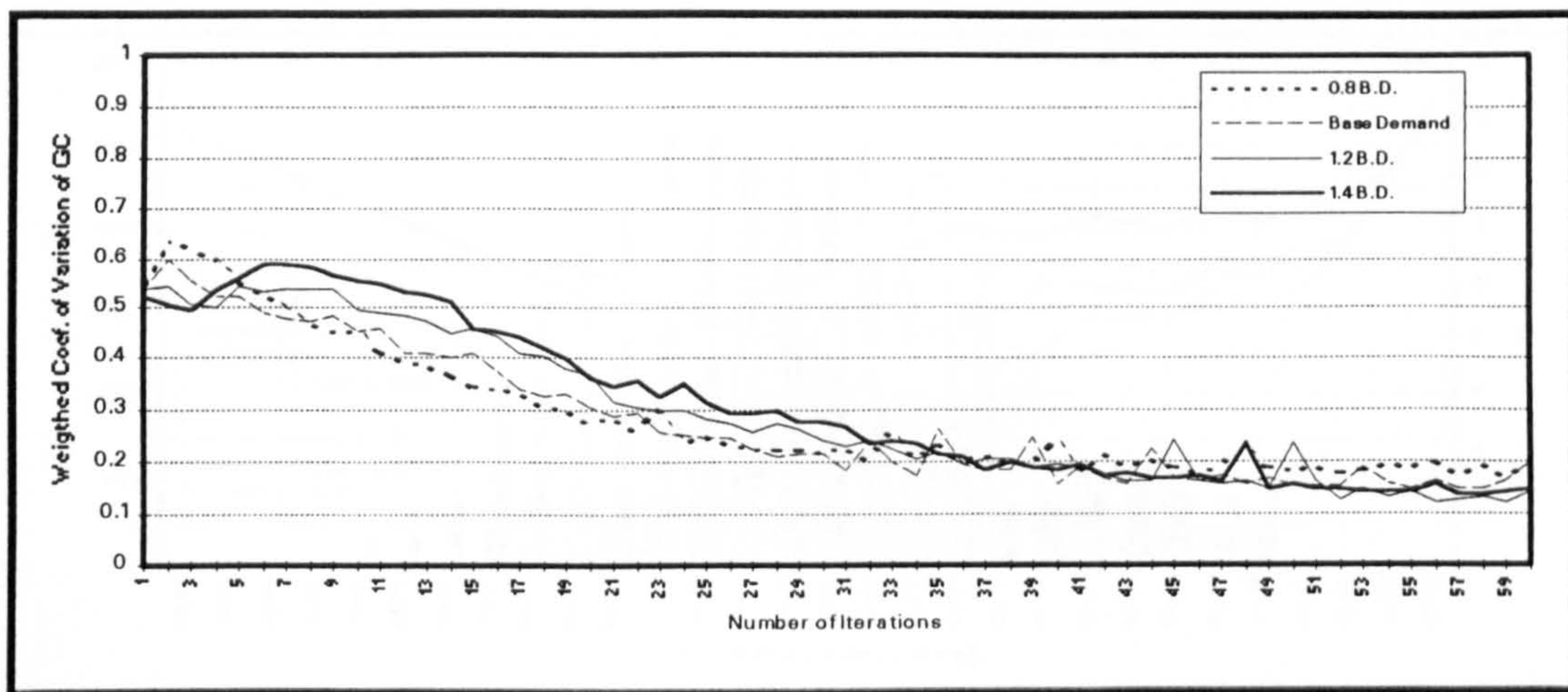


Figure (7.8): Fluctuation of the Weighted Coefficient of Variation of G. Cost under different demand levels.

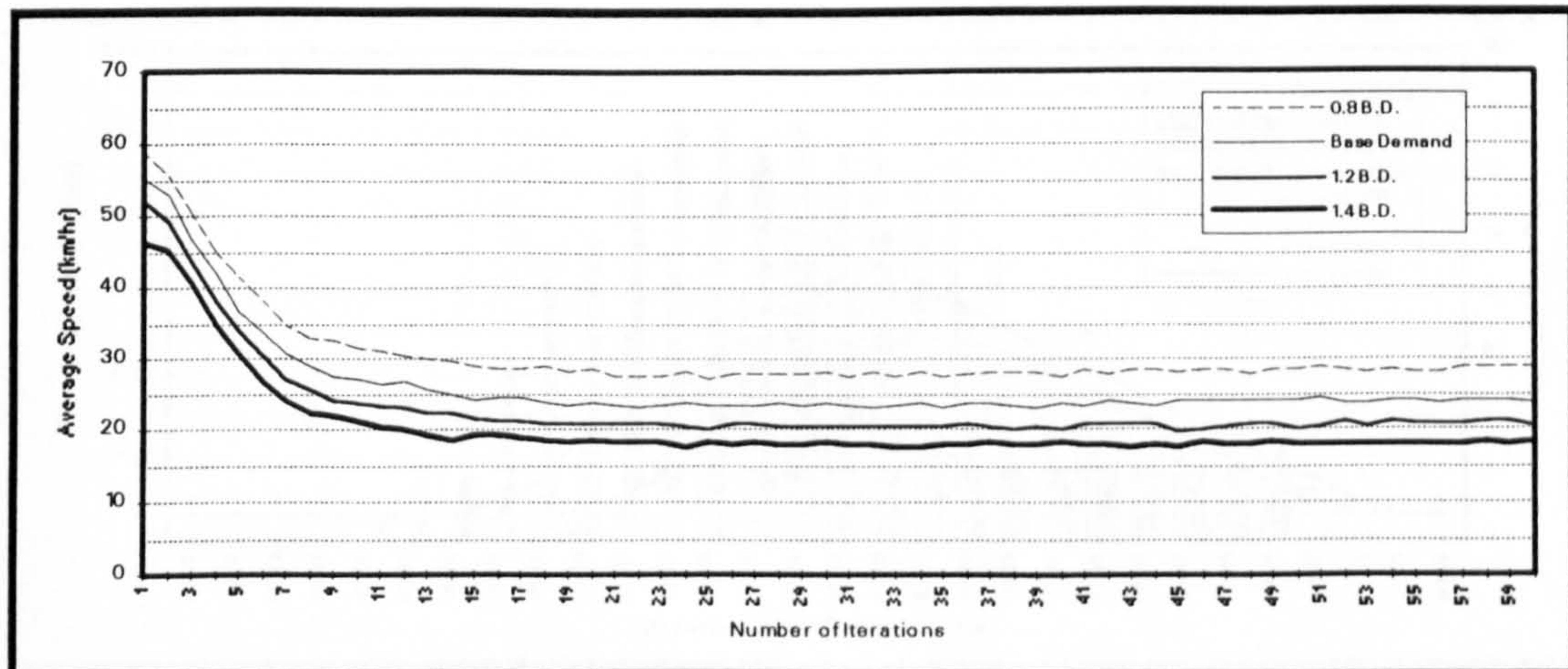


Figure (7.9): Fluctuation of the Average Speed under different demand levels.

Another impact of increasing the level of demand is the increase of the level of congestion over individual links. This is expressed by the total queue duration TQD as well as the average number of queues ANQ over each individual link during the study period. On the one hand, as the level of demand increases, the TQD over congested links increases as presented in Table (7.3). On the other hand, for congested links with a very high TQD, the number of queue occurrence (interruptions) does not increase with increasing the level of demand, while for most of links with a low queue duration, the number of occurrence increases with increasing the level of demand.

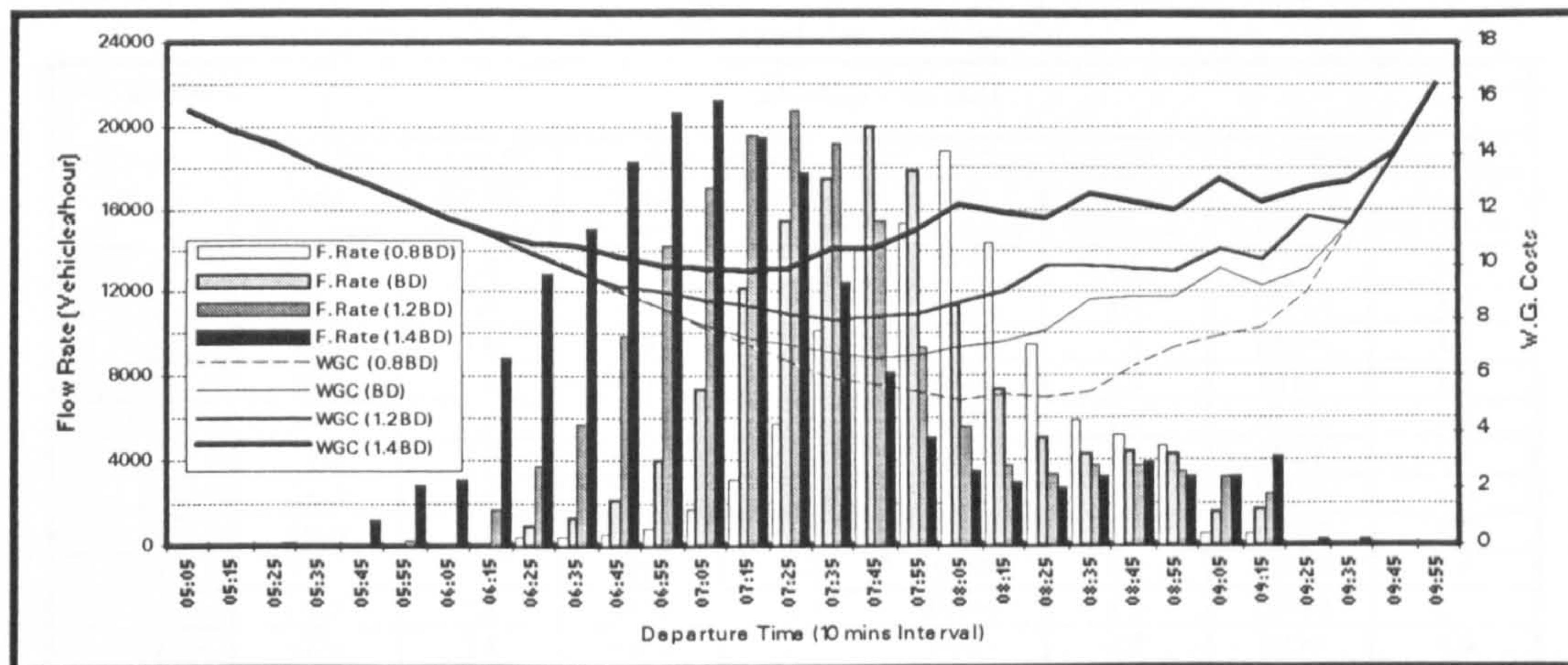


Figure (7.10): Weighted G. Cost versus Flow Rate Distribution for all movements under different demand levels.

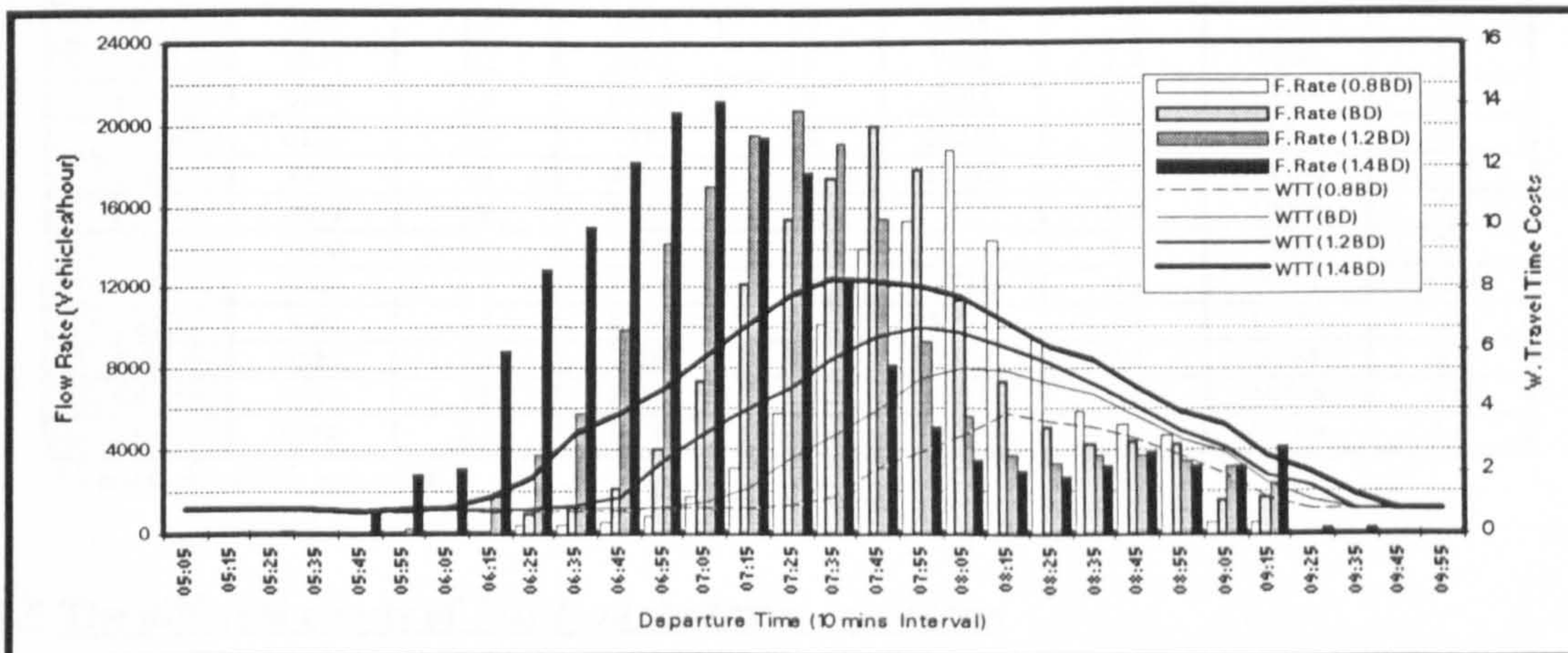


Figure (7.11): Weighted Travel Time Cost versus Flow Rate Distribution for all movements under different demand levels.

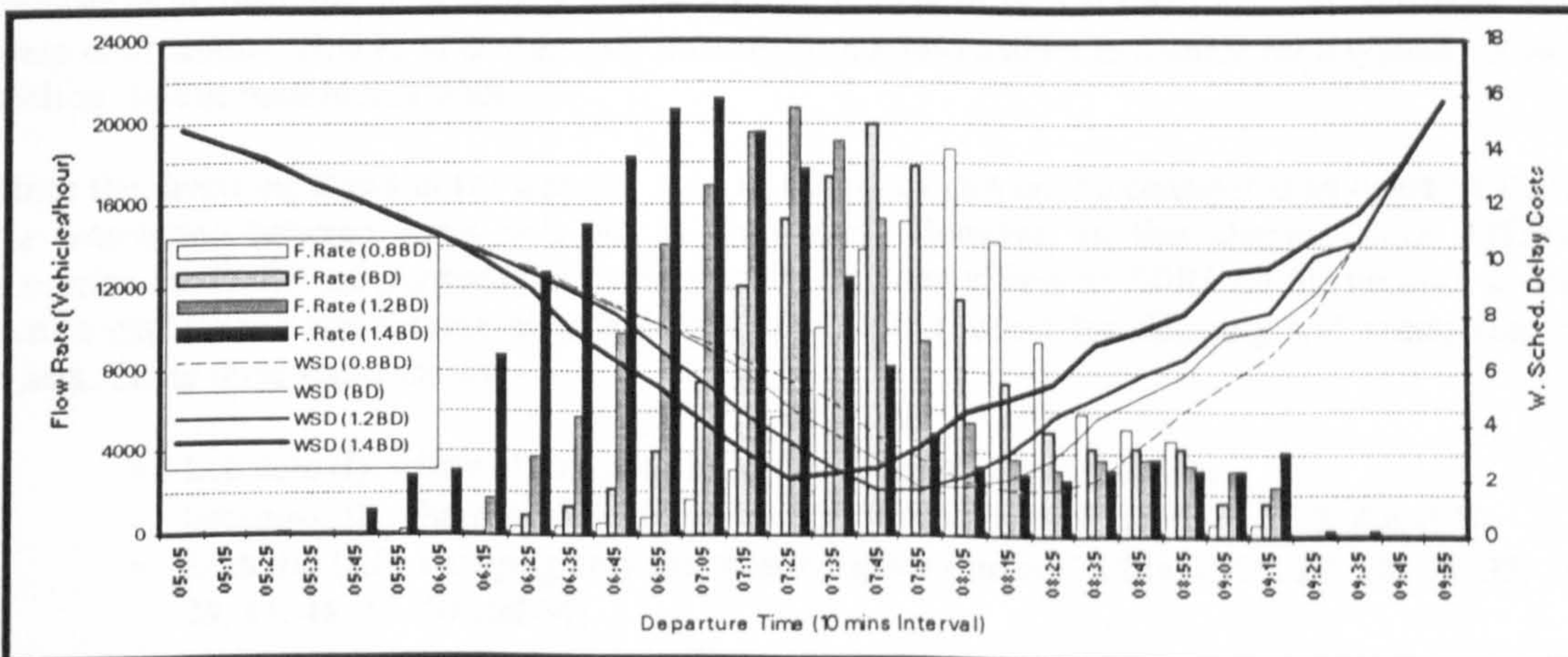


Figure (7.12): Weighted Schedule Delay Cost versus Flow Rate Distribution for all movements under different demand levels.

Table (7.3): Total queue duration and Average number of queues for the most congested links under different demand levels.

Congested Links	Demand Levels							
	0.8 Base Demand		Base Demand		1.2 Base Demand		1.4 Base Demand	
	Total Q-Duration	Av. No. of Queues	Total Q-Duration	Av. No. of Queues	Total Q-Duration	Av. No. of Queues	Total Q-Duration	Av. No. of Queues
L1	4512*	1.0	6072	1.0	8431	1.2	9780	1.0
L2	5592	1.0	7224	1.0	10332	1.2	11592	1.0
L3	5868	1.0	7404	1.0	9228	1.0	10836	1.0
L4	5693	1.6	7614	1.8	8040	1.0	9588	1.0
L5	4416	1.0	7219	1.6	8592	1.0	10668	1.0
L26	5376	1.0	6864	1.0	8460	1.0	10776	1.0
L27	6308	1.4	7632	1.0	9780	1.0	12036	1.0
L28	5699	1.8	7692	1.0	9576	1.0	11592	1.0
L29	5352	1.0	8064	1.4	9547	1.8	12457	1.4
L30	6689	1.2	8436	1.0	10260	1.0	12024	1.0
L31	4739	2.2	6600	2.2	6405	4.8	8750	4.0
L32	5112	1.0	6900	1.0	8232	1.0	10512	1.2
L33	2452	2.0	3828	2.4	2878	2.2	1936	3.2
L34	5776	1.6	7642	1.6	9896	1.8	12096	1.4
L36	2383	1.8	5676	1.0	7337	1.2	9475	1.4
L39	5453	1.6	7846	1.4	9518	1.2	12470	1.6
L41	4742	2.6	6141	4.0	7817	4.2	9838	4.0
L43	2899	1.6	3563	3.0	3858	2.8	3670	2.2
L44	4810	1.2	5892	1.0	7560	1.2	9600	1.0
L48	4908	1.0	6540	1.0	7968	1.0	10356	1.0
L49	6120	1.0	9694	1.4	11933	1.6	11844	1.0
L50	7603	1.2	9691	1.2	12748	2.0	14134	1.2
L53	4632	3.6	5331	5.0	7415	6.8	10490	5.4
L54	1440	5.4	1508	6.2	2123	7.2	1989	7.8
L58	3303	6.8	3659	8.4	4720	9.2	5652	12.6
L60	5281	1.4	6674	1.8	8154	3.2	12153	3.0
L61	2188	3.6	1687	5.4	2529	7.4	3344	8.6
L63	5592	1.0	8014	1.4	9014	1.2	10548	1.0
L64	6360	1.0	8376	1.0	10176	1.0	12168	1.0

* Total Q-duration is in seconds.

7.5 The different effects of *TDRP* on the travel conditions

The optimal *TDRP* (i.e. *TDRP* modified by considering the changes in schedule delays imposed by one another) can not be explored for a network since it requires a pre-defined desired arrival time at every individual node. As a matter of fact, the desired arrival time at any intermediate node depends on the desired arrival time at the end destination. Because any intermediate node could serve a certain number of destinations, it requires a similar number of desired arrival times that are consistent with these destinations. This is, of course, very difficult (if not impossible) to assume for a typical network such as the one used in this study.

Since the diverging flows in the network used in this study can not be considered as constant flows, the interaction between nodes will not be considered. However, in this chapter, three different scenarios for *TDRP* are suggested to investigate the different effects of *TDRP* on the network system and to explore the significance of considering some links (nodes) for charging and exempting the others. These three scenarios are:—

- Scenario (I): where all congested links are charged for congestion.
- Scenario (II): charging only the entrances to the network [5 Links: 1, 2, 3, 4 and 5].
- Scenario (III): charging only the most congested links [12 Links: 26, 27, 28, 30, 32, 34, 39, 41, 44, 49, 50 and 64].

The simulation model is then employed to investigate the different effects of *TDRP* on the network system under these three scenarios. The results of the first scenario are initially discussed, and then compared with the results of the other two scenarios.

Throughout the simulation, the day-to-day adjustment factor for *TDRP*, R_2 , is given by

$$R_2 = \begin{cases} 0.04 & \text{---} \omega \leq 21 \\ 0.02 & \text{---} 40 \geq \omega > 21 \\ 0.01 & \text{---} \omega > 40 \end{cases} \quad (7.2)$$

The charge over every individual link is calculated every 10 seconds, and adjusted by the average charge experienced over the corresponding 60-second time interval during the previous iteration.

7.5.1 The different effects of *TDRP* under Scenario (I)

System Convergence

In considering *TDRP*, the simulation started with the same initial departure pattern considered before introducing *TDRP* (i.e. pseudo-stationary state). Figure (7.13) shows the day-to-day evolution of the WCV of the generalized cost for before and after the provision of *TDRP*. The figure exhibits a better convergence to a steady state (without oscillations) under *TDRP*. It is also exhibited that the system reaches this state of equilibrium almost at the same time as the case before *TDRP*. In addition, the Figure shows that the coefficient of variation at the final state of the system decreases from 0.18 to 0.1 as a result of introducing *TDRP*.

Table (7.4) summarises the different effects of *TDRP* scenario (I) under different demand levels. It is depicted from this table that the convergence is improved under *TDRP* for all the demand levels and as the level of demand increases the value of WCV at equilibrium decreases. This could be because, as the demand level increases, the total travel cost incurred by each individual vehicle increases and becomes more correlated to the travel cost of the system as a whole and less deviated from that experienced by other vehicles. This level of convergence (under *TDRP*) would be difficult to achieve when the system is only under the two travel cost functions (i.e. travel time cost function and schedule delay cost function).

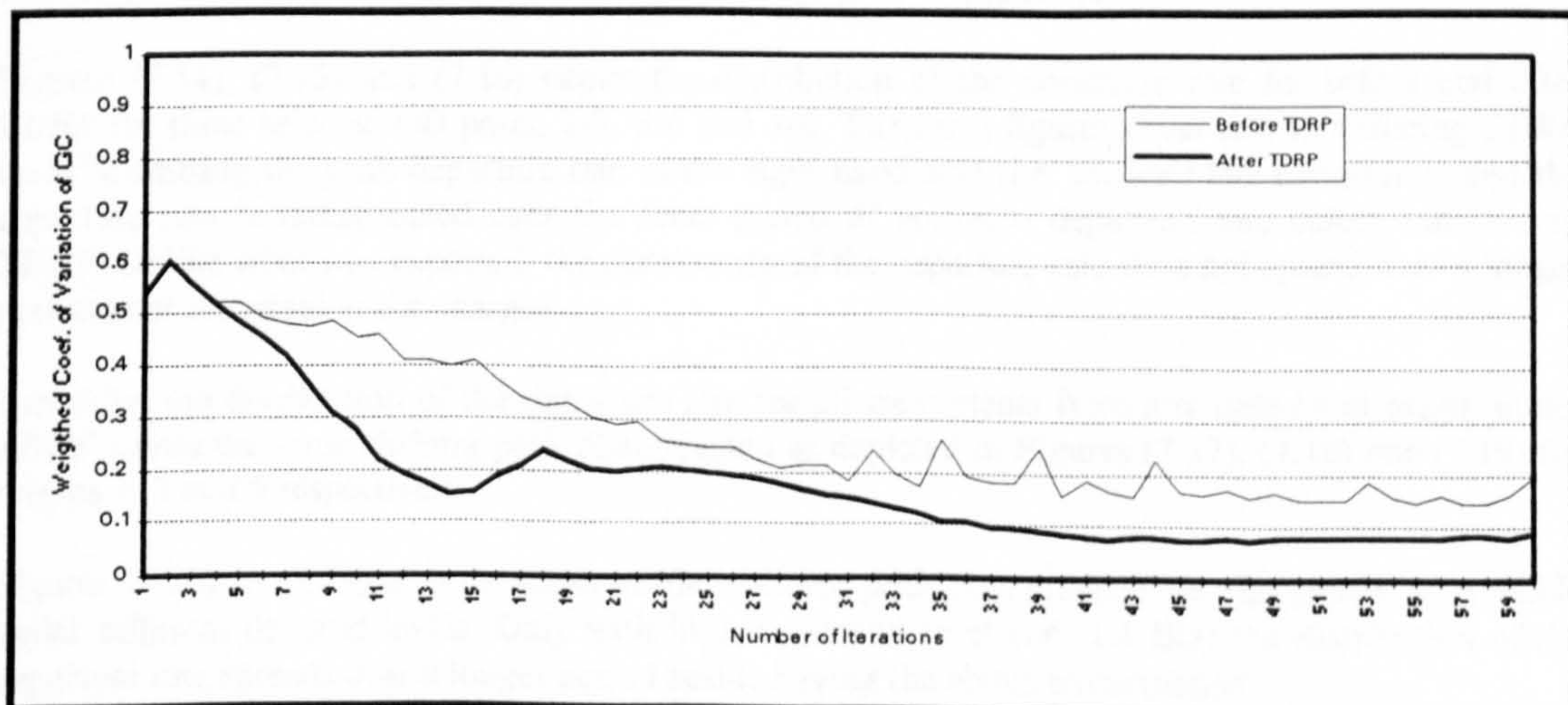


Figure (7.13): The Fluctuation of the Weighted Coefficient of Variation of G. Cost under the Base Demand (before and after *TDRP*).

Table (7.4): The effects of congestion on the base case results*.

Demand Level	Coeff. of Var.		Speed (km/h)			T. Time C.		S. Delay C.		TDRP	WGC	
	Before	After	Before	After	% of Increase	Before	After	Before	After		Before	After
<i>0.8 Base Demand</i>	0.18	0.10	28.6	48.6	69.8	2.59	1.14	3.20	3.36	2.32	5.79	6.82
<i>Base Demand</i>	0.16	0.08	24.0	48.3	101.2	3.39	1.17	3.97	4.13	3.25	7.36	8.54
<i>1.2 Base Demand</i>	0.14	0.08	20.8	45.2	117.4	4.23	1.31	4.66	5.05	4.52	8.89	10.88
<i>1.4 Base Demand</i>	0.15	0.07	17.8	42.2	136.6	5.18	1.51	5.52	5.89	5.84	10.70	13.24

* The values in this table are the average of the last 10 iterations.

Overall Average Speed

Table (7.4) above depicts that with the provision of *TDRP*, the overall average speed increases from 24.02 km/h to 48.33 km/h under the base demand. It is also depicted that under different demand levels there is a substantial increase in the overall average speed and as the demand level increases, the percentage of increase in the overall average speed becomes very substantial: for example, the percentages of increase under 1.2 BD and 1.4 BD are 117.4% and 136.6% respectively.

It should be recognized from the above table that the reduction in the overall average speed as a result of increasing the demand level (before *TDRP*) is not with the same proportion of the reduction in travel time. This could be because the value of speed is an average value over the whole network (not weighted) while the travel time is a weighted average.

Departure pattern and travel costs

In the analysis presented in this section, travellers do not have the option of switching to alternative transport modes or not making the journey since the travel demand is assumed to be inelastic. Therefore, travellers' response to *TDRP* is limited to switching to an alternative route (a route with no or less charge) or shifting the departure time (to avoid or minimize the charges). Thus, travellers are expected to switch from congested routes to less congested ones and simultaneously keep adjusting their departure times until the system reaches its final state of equilibrium.

Figures (7.14), (7.15) and (7.16) depict the distribution of the departure rate for before and after *TDRP* for three selected OD pairs: 2-7, 3-6 and 4-8. The three figures show that introducing *TDRP* leads to shifting the peak departure rate to the right hand side (i.e. towards late departures) and the departure rate is redistributed over the same period of non-zero departure rate before introducing *TDRP*. Unlike what was expected, the distribution of the departure rate does not spread over a longer period after introducing the charges.

Typically, the distribution of the departure rate for all movements from any individual origin under *TDRP* shows the same shifting peak phenomenon as depicted in Figures (7.17), (7.18) and (7.19) for origins 1, 2 and 5 respectively.

Figures (7.20) to (7.23) also demonstrate the shifting peak phenomenon for the system as a whole under different demand levels. Only with higher demand level (i.e. 1.4 BD) the distribution of the departure rate spreads over a longer period beside having the above phenomenon.

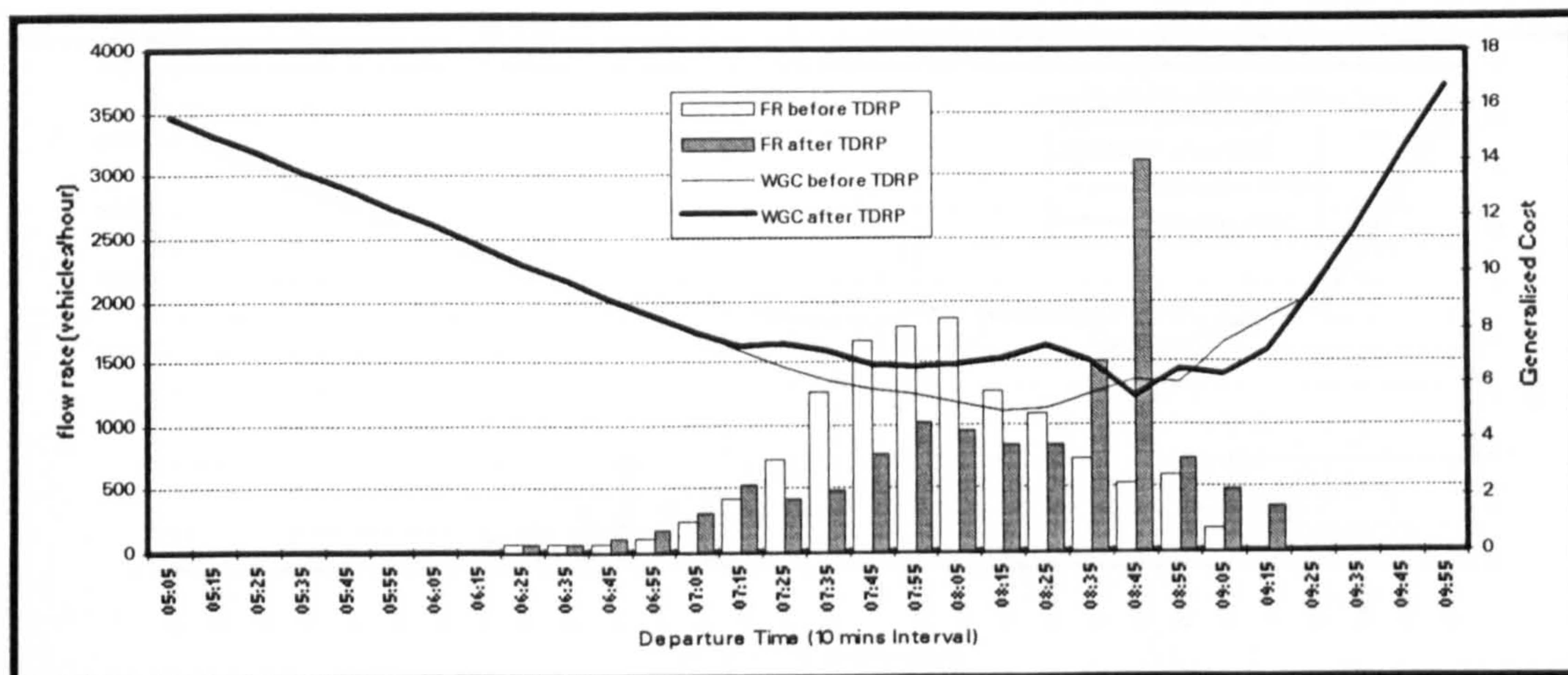


Figure (7.14): Weighted G. Cost versus Flow Rate Distribution for O2-D7 (before and after TDRP).

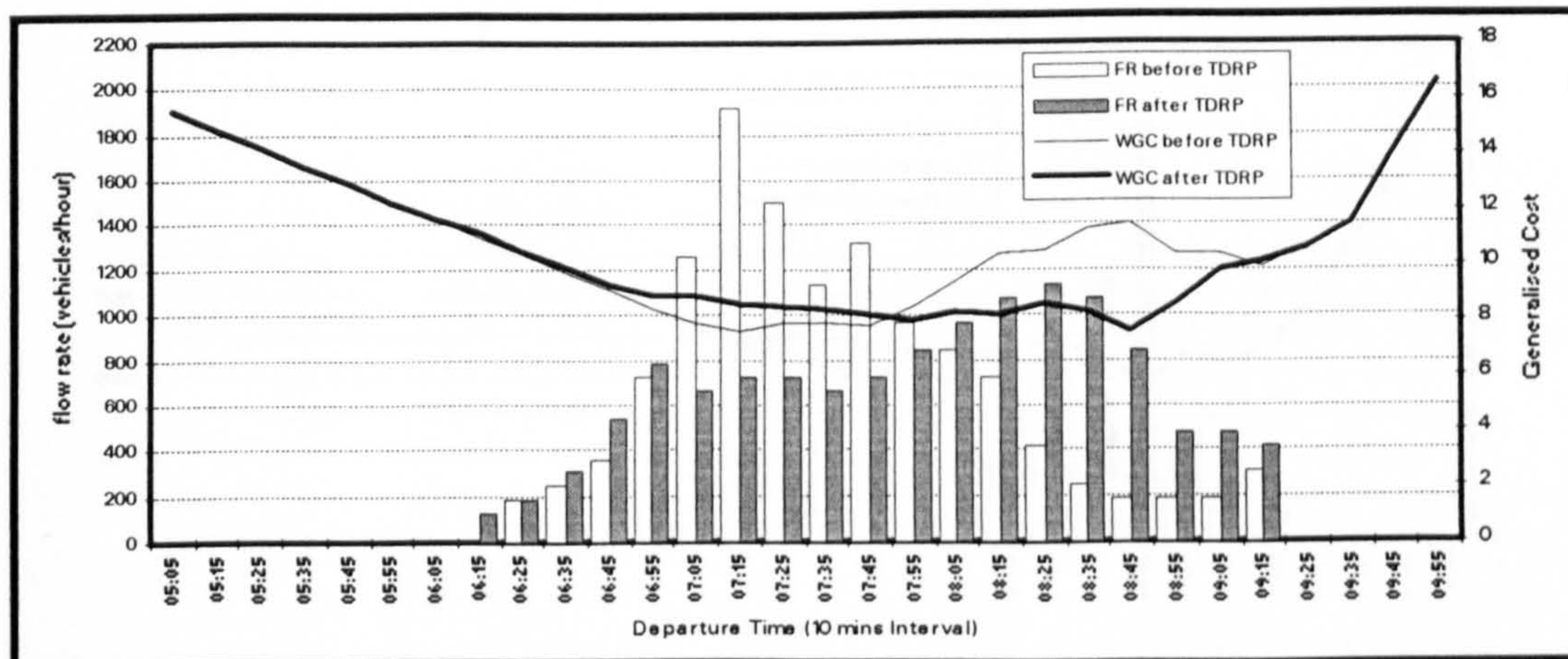


Figure (7.15): Weighted G. Cost versus Flow Rate Distribution for O3-D6 (before and after TDRP).

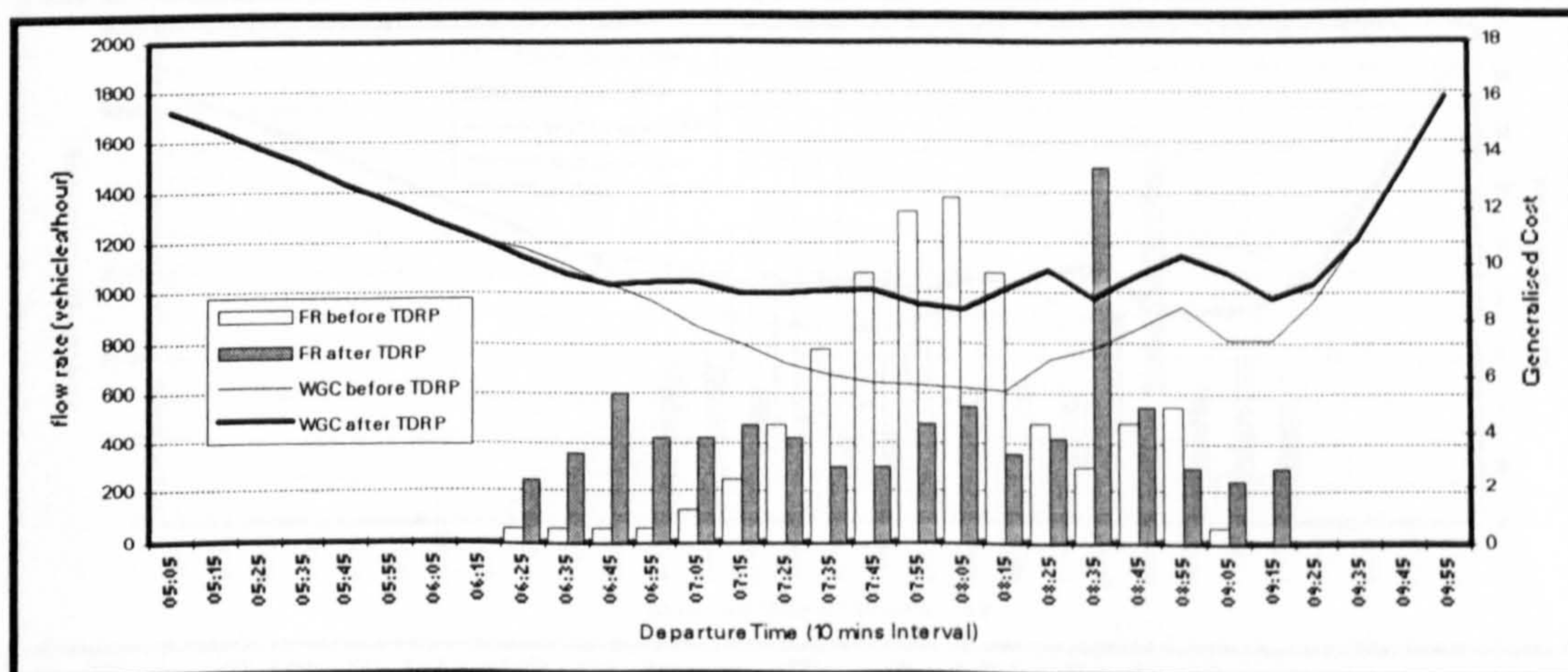


Figure (7.16): Weighted G. Cost versus Flow Rate Distribution for O4-D8 (before and after TDRP).

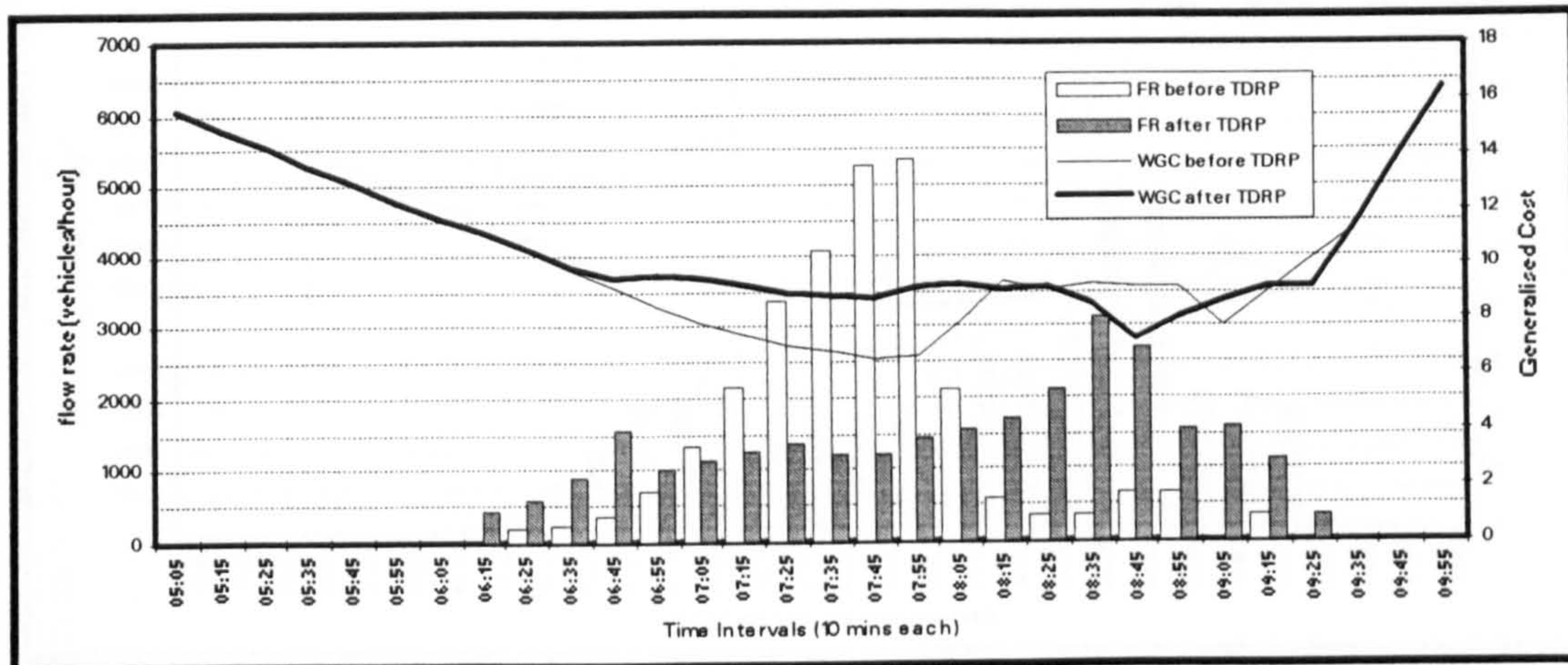


Figure (7.17): Weighted G. Cost versus Flow Rate Distribution for all movements from O1 (before and after TDRP).

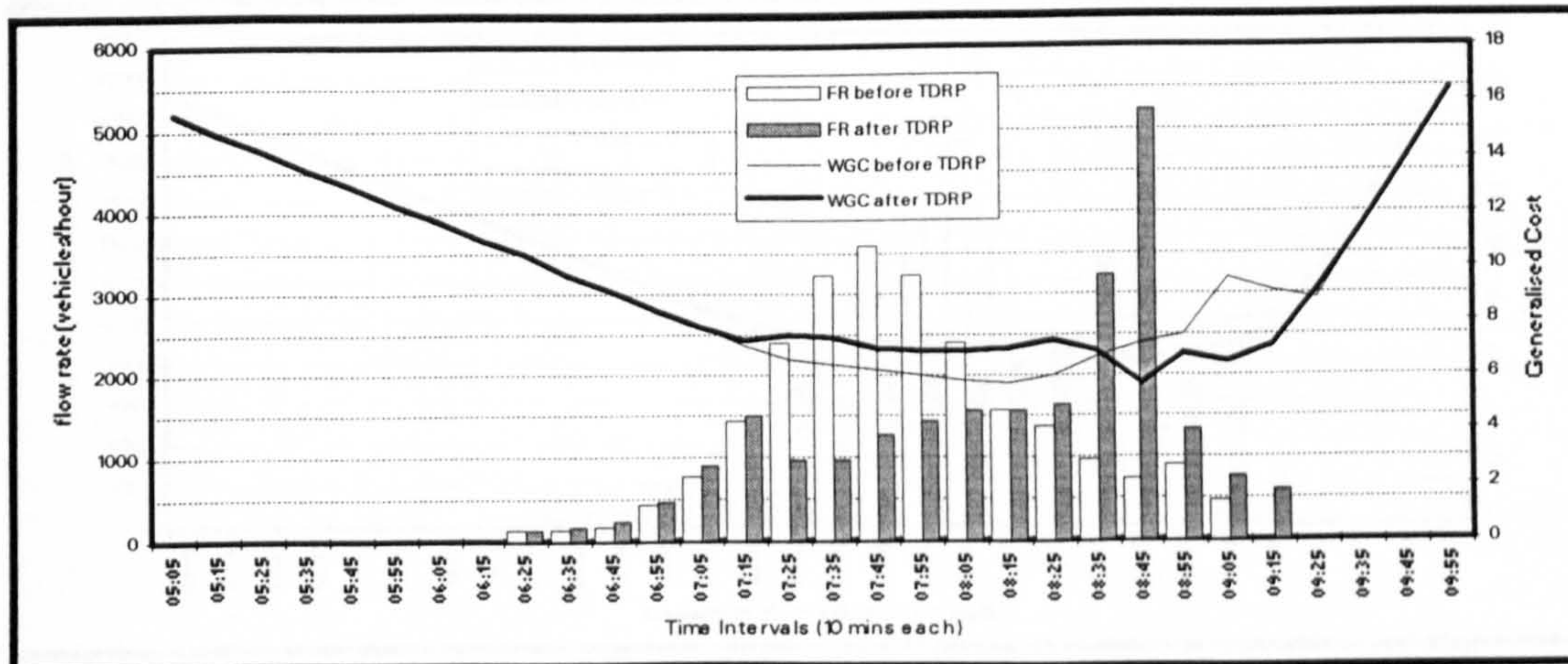


Figure (7.18): Weighted G. Cost versus Flow Rate Distribution for all movements from O2 (before and after TDRP).

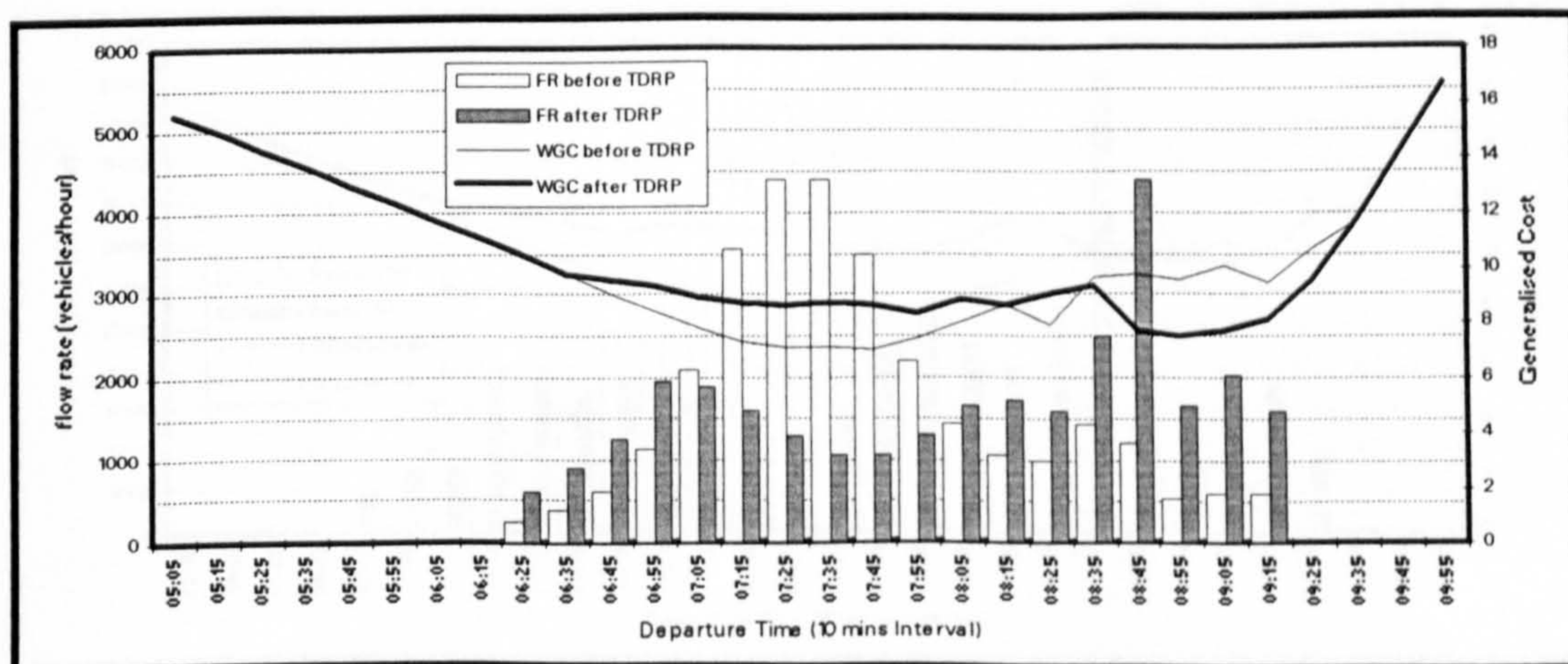


Figure (7.19): Weighted G. Cost versus Flow Rate Distribution for all movements from O5 (before and after TDRP).

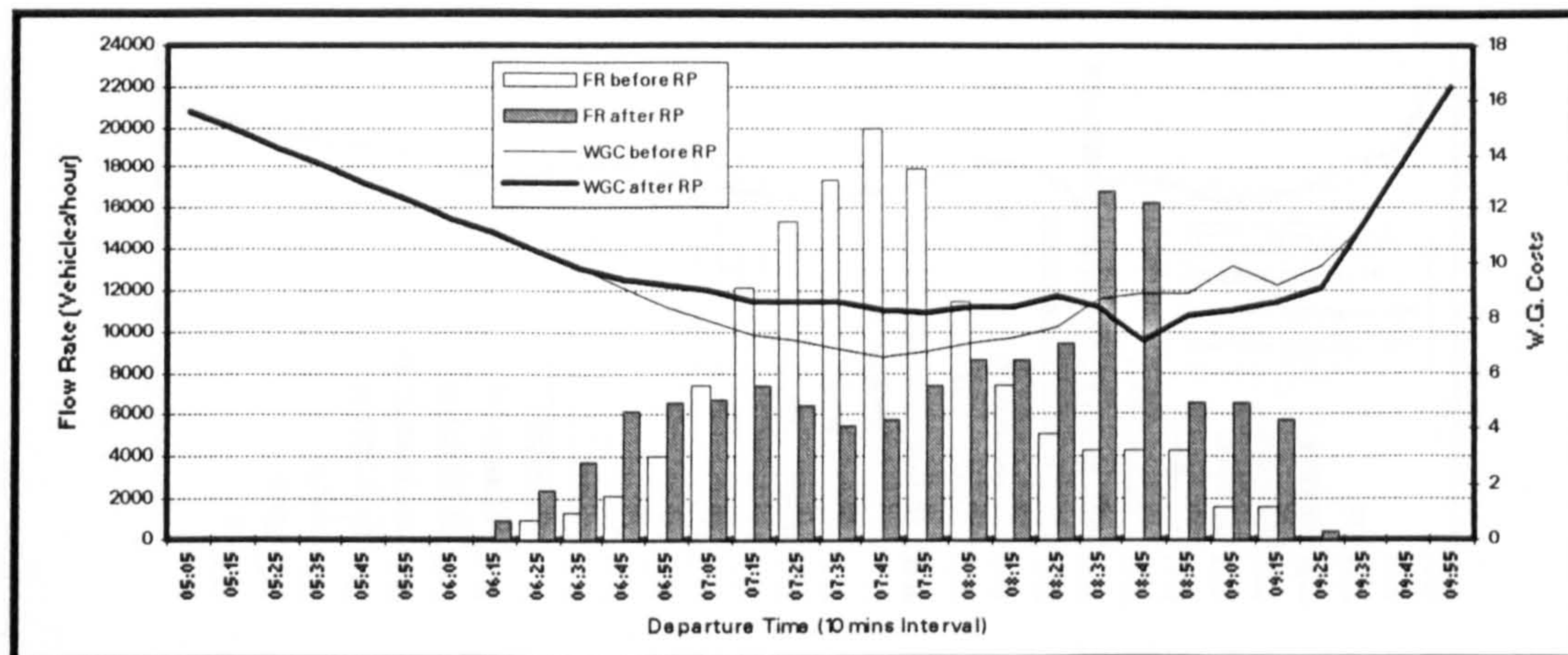


Figure (7.20): Weighted G. Cost versus Flow Rate Distribution for all movements under the base demand (before and after TDRP).

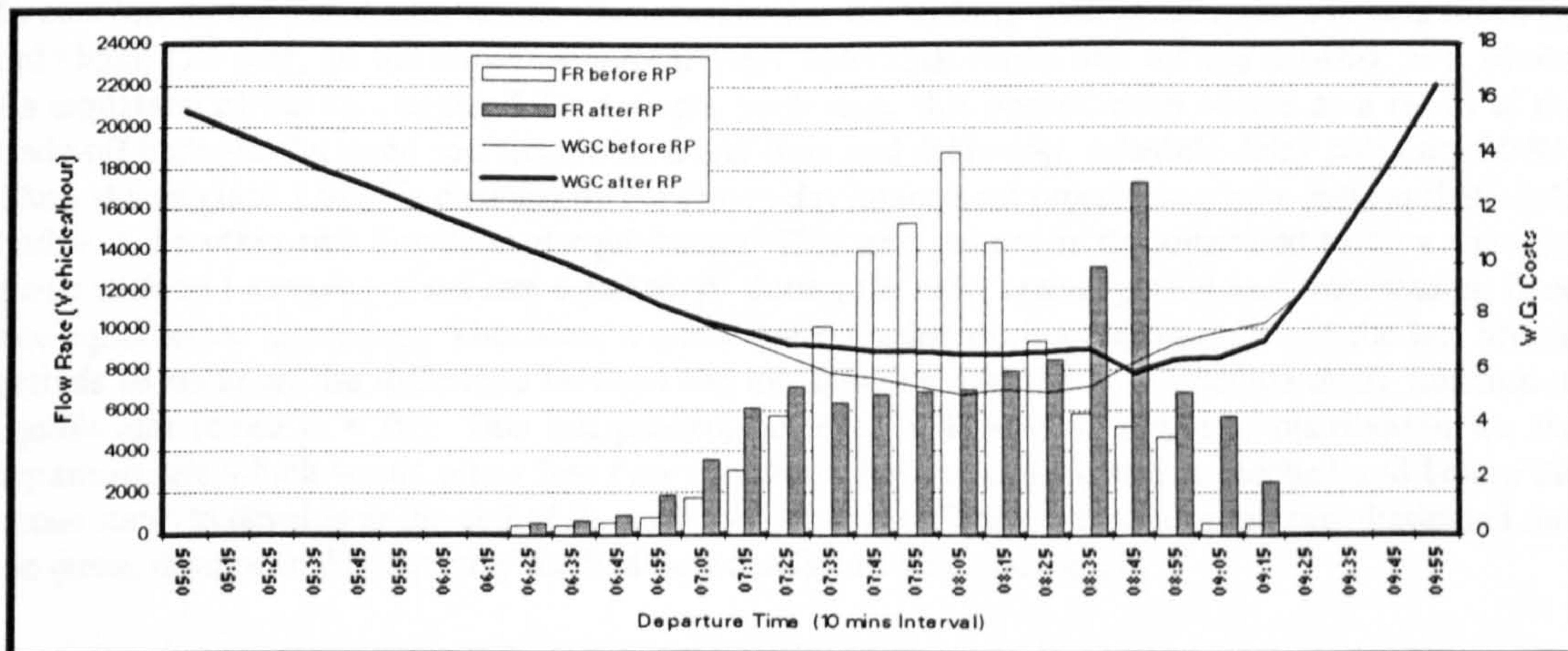


Figure (7.21): Weighted G. Cost versus Flow Rate Distribution for all movements under 0.8 the base demand (before and after TDRP).

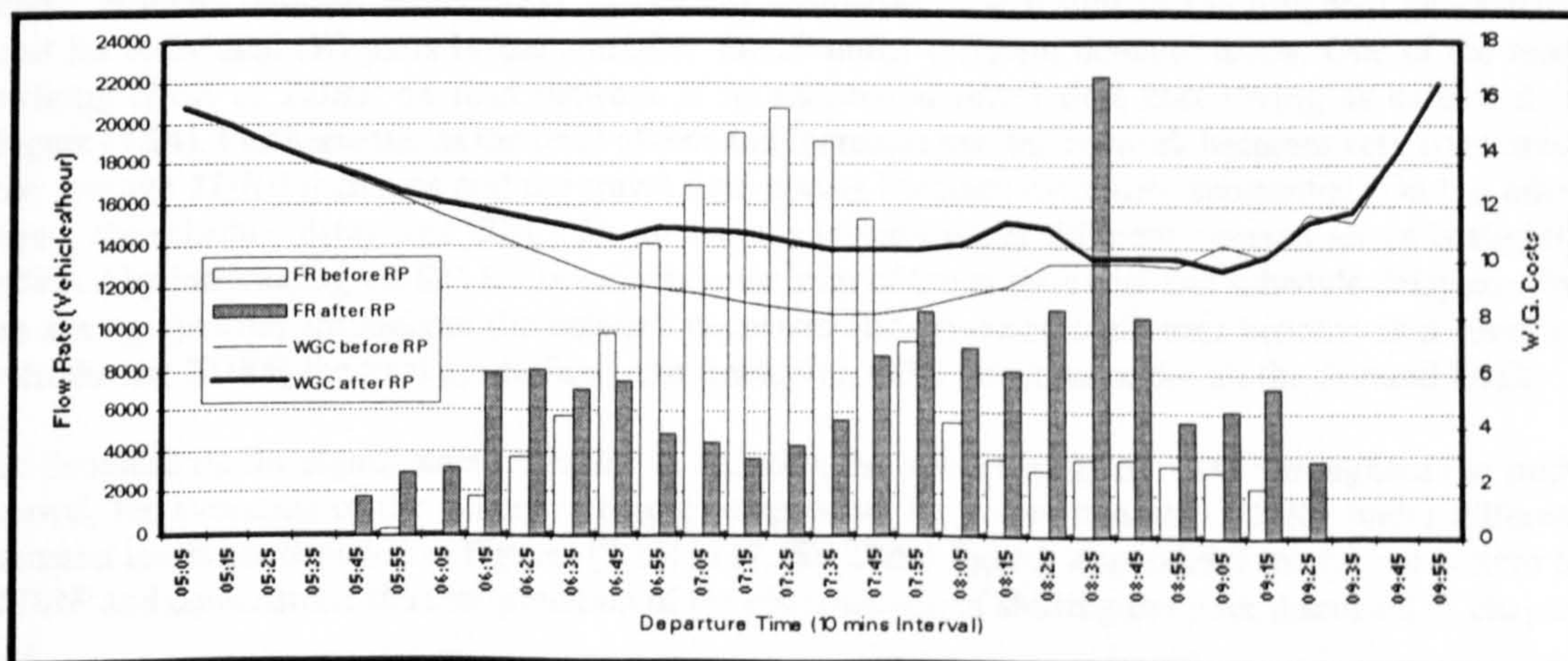


Figure (7.22): Weighted G. Cost versus Flow Rate Distribution for all movements under 1.2 the base demand (before and after TDRP).

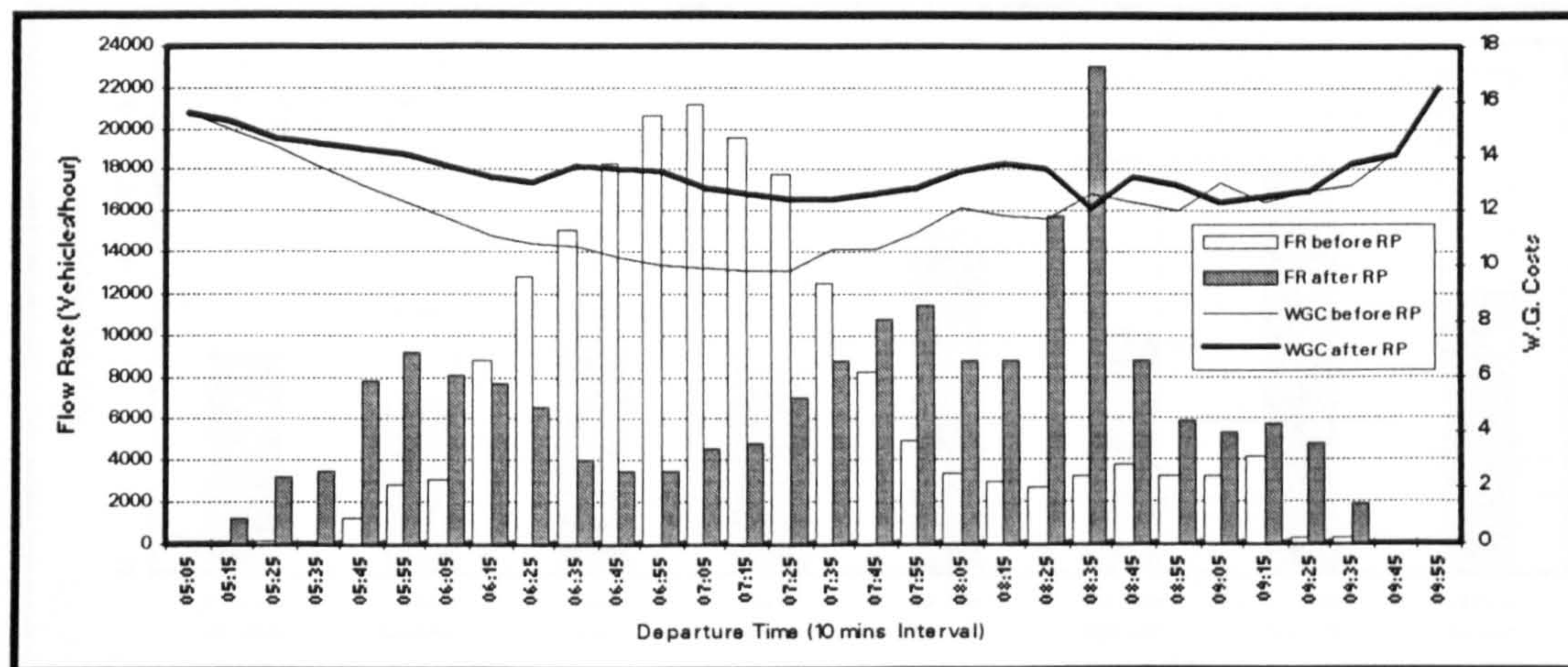


Figure (7.23): Weighted G. Cost versus Flow Rate Distribution for all movements under 1.4 the base demand (before and after *TDRP*).

In conclusion, the simulation results demonstrate that the shifting peak phenomenon occurs for every individual OD pair, all the movements from every individual origin and for the network as a whole. As explained earlier in Chapter 5 for a single bottleneck, this phenomenon occurs as a result of the trade-off between the three cost functions: travel time and delay cost, schedule delay cost and *TDRP*. Thus, the original *TDRP* is changed by the day-to-day adjustment process to a new pattern that could trade-off the other two functions at equilibrium. This new pattern is demonstrated to be a triangle-shape with an increasing fixed rate equal to ' β ' during the early arrival period and a decreasing fixed rate equal to ' α ' afterwards. Therefore, a queue must develop during the on-time and the late arrival periods to trade-off the difference between the toll function and the late schedule delay function at equilibrium (since $\alpha < \beta\gamma$). This toll pattern, however, will lead to a typical distribution for the departure rate which would allow free flow conditions before its peak rate is reached and before the queue starts to develop at the end of the early arrival period. Then, after this peak rate dissipated and the queue disappeared completely the free flow conditions take place again.

Travel costs

The values of travel cost components (travel time and delay costs, schedule delay costs and *TDRP*) for an average journey throughout the network are presented in Table (7.4) and exhibited in Figure (7.24) under different demand levels. Also Table (7.5) summarises the results of the weighted generalized cost for individual OD pairs before and after *TDRP* under different demand levels. One of the most striking effect of *TDRP* on road network is its substantial travel time cost saving as exhibited in Figure (7.24). Furthermore, as the level of demand increases and the network becomes very congested, the average *TDRP* increases and the travel time saving becomes extremely substantial. On the other hand, the schedule delay cost 'SDC' for an average journey under different demand levels is slightly affected by introducing *TDRP*. However, while the sum of travel time cost and schedule delay cost for an average journey throughout the network decreases and produces some user benefits as a result of introducing *TDRP*, the total generalised cost (including tolls) increases under all the demand levels.

To demonstrate the significance of *TDRP* in reducing the travel time cost 'TTC' throughout the study period, the evolution of the generalized cost components for before and after *TDRP* under different demand levels are depicted in Figures (7.25) to (7.28). These figures also exhibit the typical pattern of *TDRP* and demonstrate the interpretation of the phenomenon of shifting the peak discussed in chapter (5).

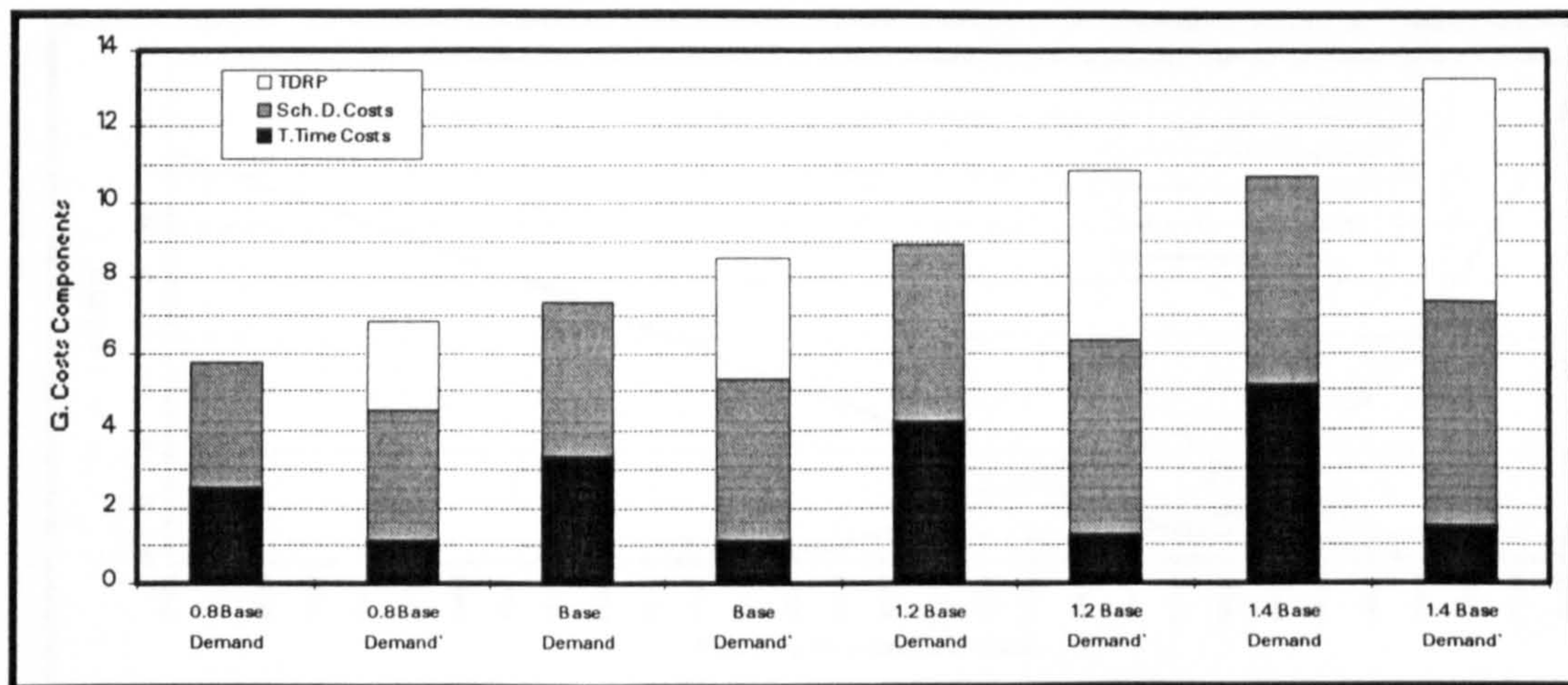


Figure (7.24): G. Cost Components under different demand levels for before and after TDRP (The values represent the average of the last 10 iterations).

Table (7.5): Weighted Generalized costs for all O-D Pairs before and after TDRP under different demand levels.

O-D Pairs	Demand Levels							
	0.8 Base Demand		Base Demand		1.2 Base Demand		1.4 Base Demand	
	Before	After	Before	After	Before	After	Before	After
1-6	5.72	7.34	7.10	8.71	8.37	11.48	10.48	13.95
1-8	6.40	7.37	7.57	9.15	8.82	11.96	10.92	14.58
2-7	4.19	5.18	5.80	6.59	7.84	8.09	9.93	10.32
2-8	5.71	5.75	7.35	7.15	9.35	9.30	11.51	12.46
3-6	6.70	6.83	8.61	8.62	10.25	11.03	11.52	13.97
3-8	6.81	6.82	8.55	8.74	10.43	11.33	11.67	14.07
4-6	4.29	6.75	6.22	8.81	6.97	10.85	8.32	12.92
4-7	6.13	7.43	7.49	9.48	8.41	11.93	10.11	13.87
4-8	4.83	7.08	6.34	9.19	7.36	11.71	8.66	13.64
5-6	5.34	6.91	7.31	8.64	8.39	10.64	10.44	12.26
5-7	6.33	6.51	8.03	7.96	9.48	9.75	11.66	11.47
5-8	6.74	7.67	8.36	9.57	10.05	11.85	12.13	14.15
Average GC	5.77	6.80	7.39	8.55	8.81	10.83	10.61	13.14

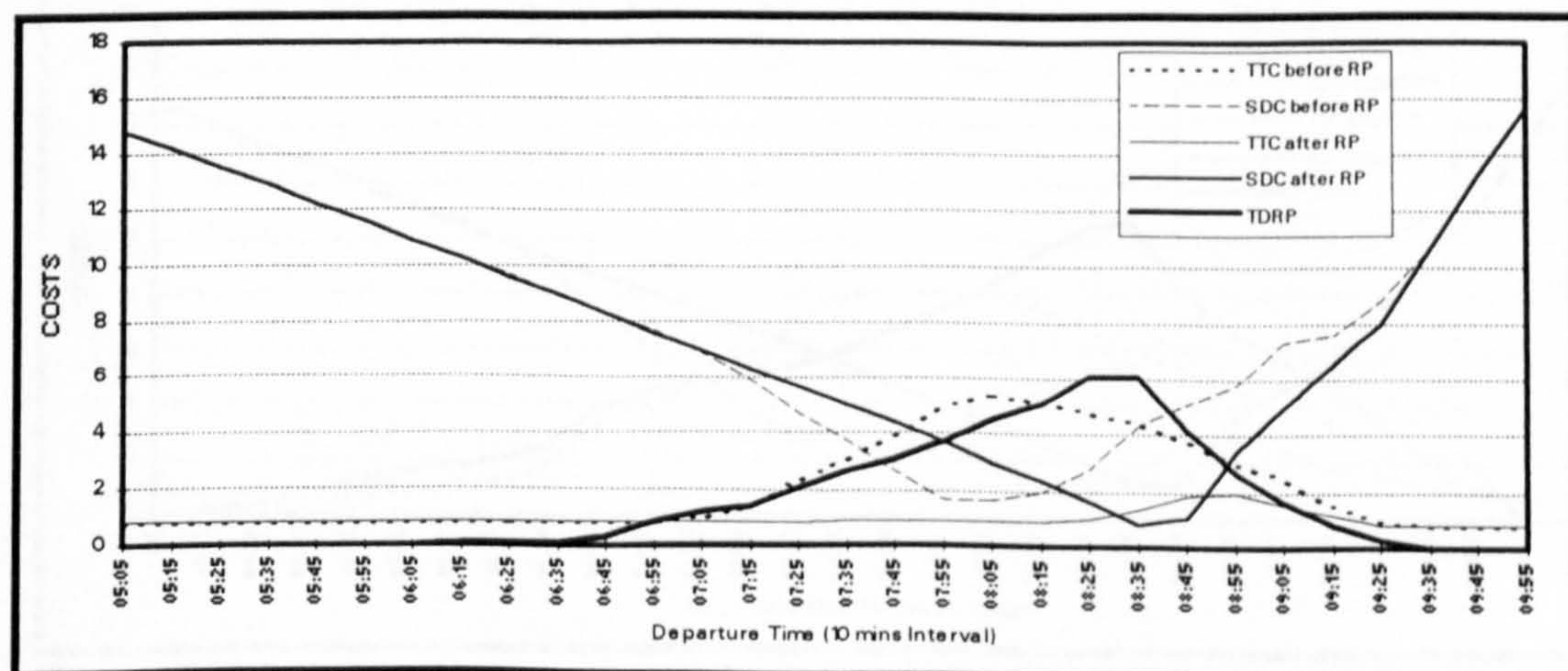


Figure (7.25): G. Cost Components under the base demand (before and after TDRP).

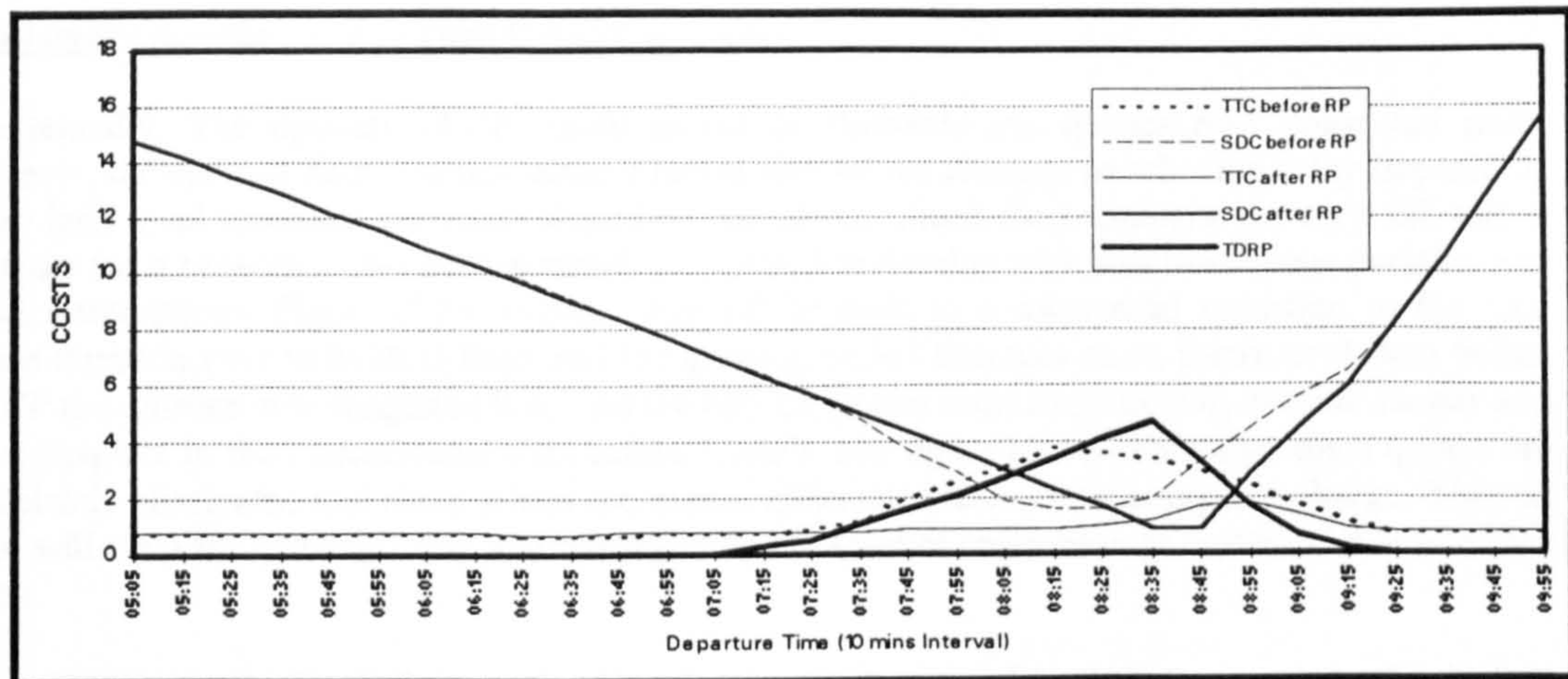


Figure (7.26): G. Cost Components under 0.8 the base demand (before and after *TDRP*).

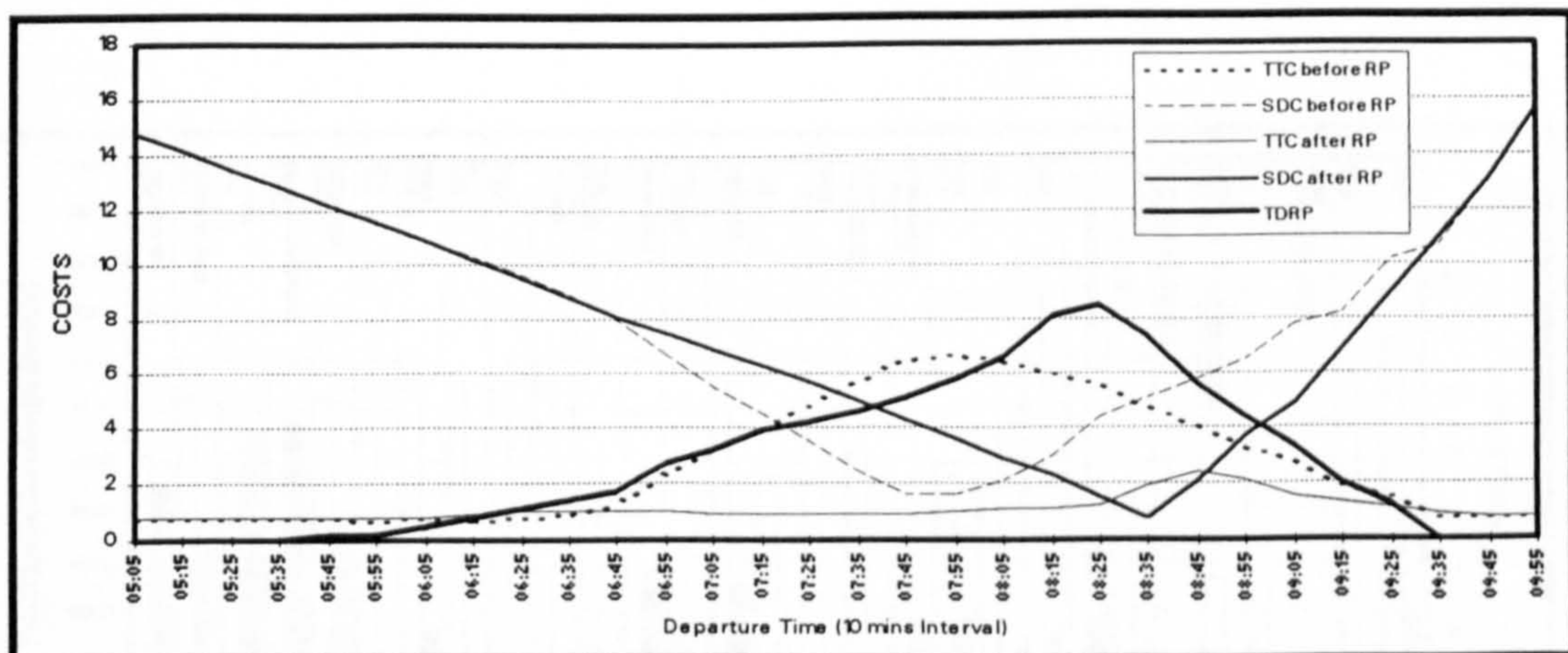


Figure (7.27): G. Cost Components under 1.2 the base demand (before and after *TDRP*).

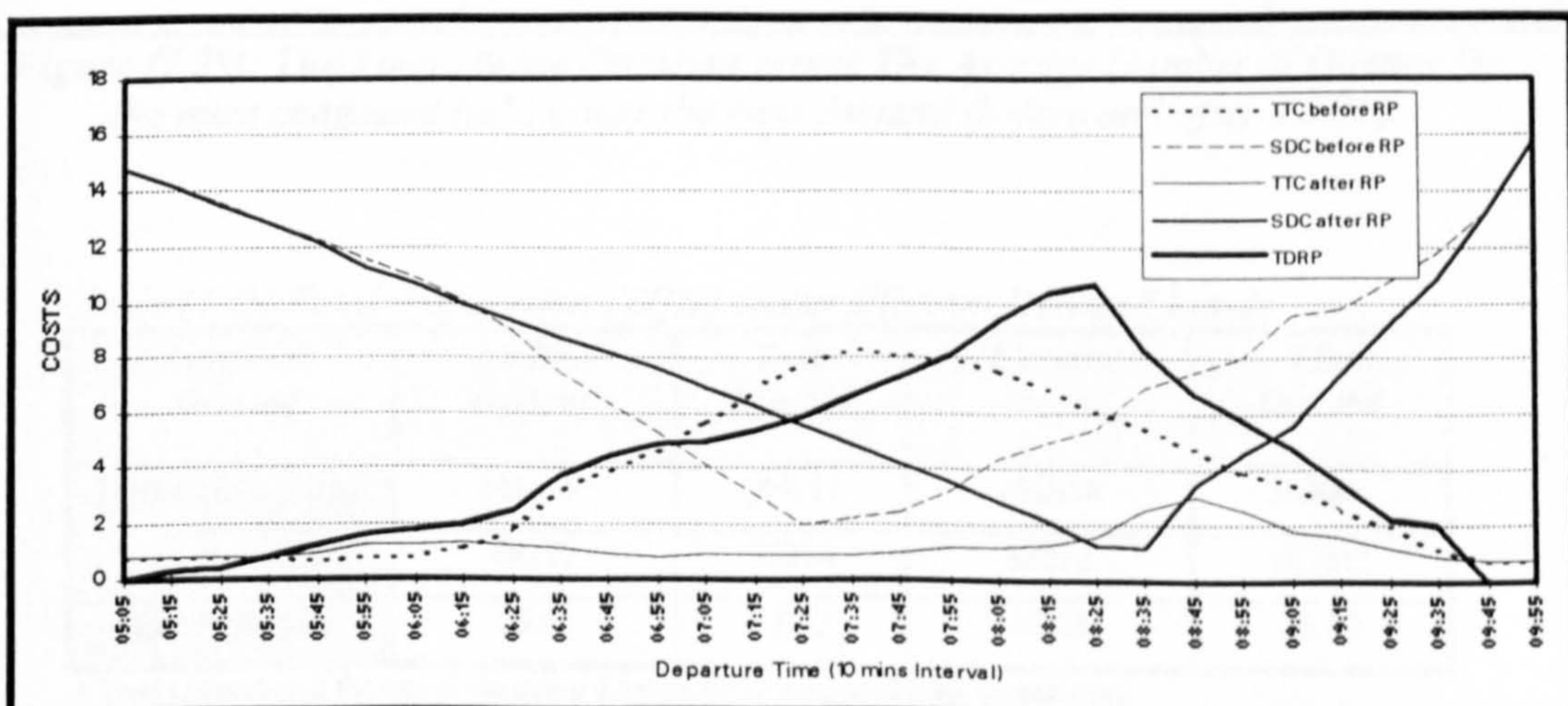


Figure (7.28): G. Cost Components under 1.4 the base demand (before and after *TDRP*).

Total queue duration and average number of queues

Theoretically, The optimal *TDRP* would expect to eliminate the queues over congested links. However, the optimal *TDRP* in that context means that all the changes in schedule delay imposed by every individual travellers on others should be considered. Since these changes are very difficult to consider for a network, some queues would be expected to develop with less total queue duration and more interruptions. Figure (7.29) exhibits that *TDRP* leads to a substantial reduction in the total queue duration over individual links and the queuing period becomes more interrupted than before *TDRP* (i.e. queues over congested links and the very congested links in particular, become shorter and more frequent in their occurrences than before *TDRP*). The figure also exhibits that some queues are completely eliminated and some others are almost eliminated after introducing the charge. This, in turn, will keep the queuing delay and consequently the level of congestion at bottlenecks at very low levels.

Table (7.6) presents the total queue duration for the whole network and the percentage of reduction due to introducing *TDRP*, while Table (7.7) summarises the total queue duration and the average number of queues before and after *TDRP* for the most congested links under different demand levels. The results in Table (7.6) explore that the percentage of reduction in the total queue duration does not necessarily increase as the level of demand increases.

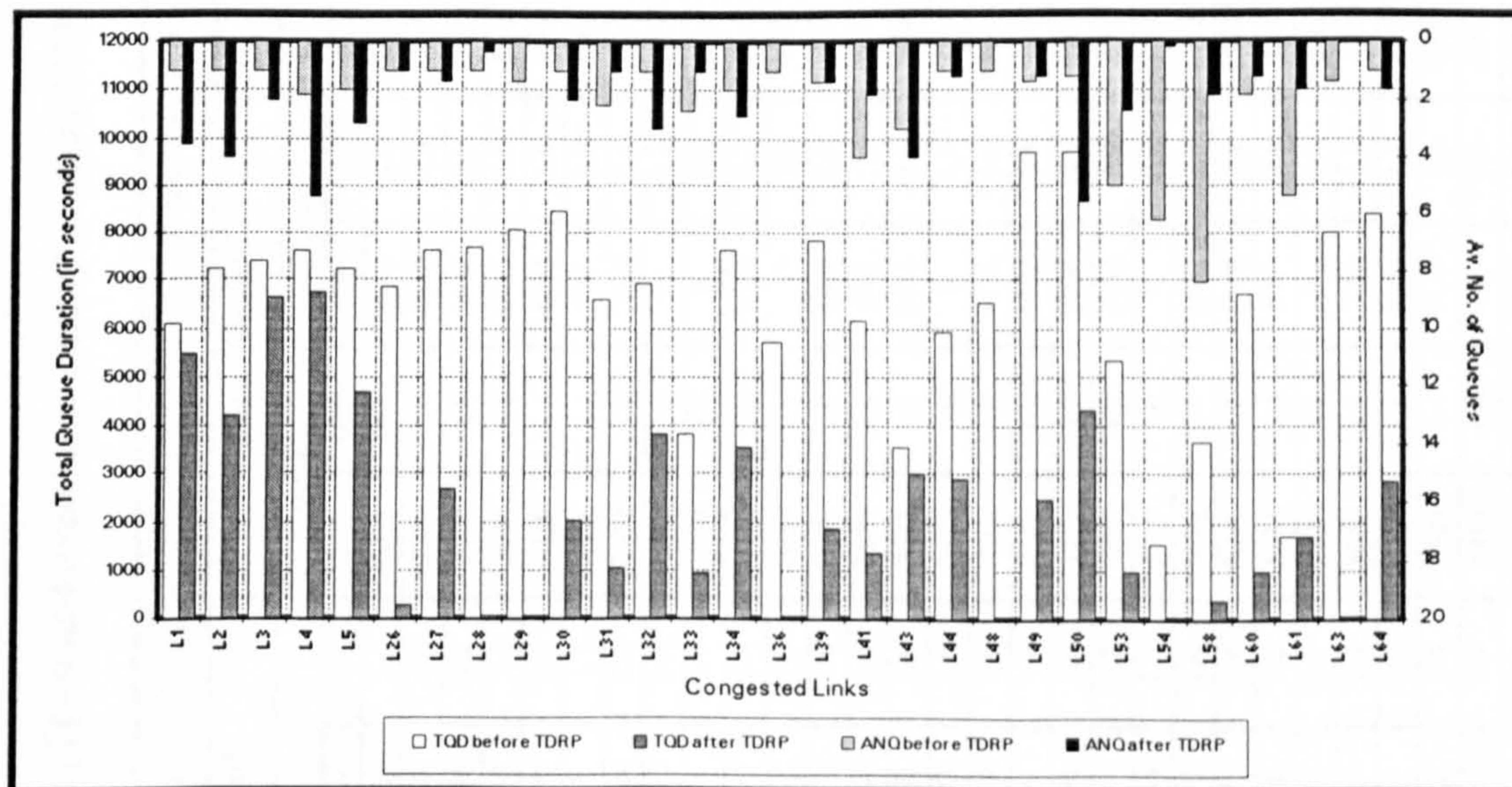


Figure (7.29): The Total Queue Duration versus The Average Number of Queues for the most congested links under the base demand (before and after *TDRP*).

Table (7.6): Total Q-Duration (TQD) under different Demand Levels.

Levels of Demand	0.8 Base Demand	Base Demand	1.2 Base Demand	1.4 Base Demand
<i>TQD Before TDRP</i>	141839*	189911	232856	283006
<i>TQD After TDRP</i>	48779	65284	88292	127012
<i>% of Reduction</i>	65.61	65.62	62.08	55.12

* Total Q-duration is the sum of the queue duration for all congested links (in seconds).

Table (7.7): Total queue duration (TQD) and Average number of queues (ANQ) for the most congested links under different demand levels (before and after TDRP).

Congested Links	Demand Levels																	
	0.8 Base Demand						Base Demand						1.2 Base Demand					
	Before			After			Before			After			Before			After		
	TQD	ANQ		TQD	ANQ		TQD	ANQ		TQD	ANQ		TQD	ANQ		TQD	ANQ	
L1	4512*	1.0		4426	2.4		6072	1.0		5497	3.6		8431	1.2		6955	4.6	
L2	5592	1.0		3920	3.6		7224	1.0		4210	4.0		10332	1.2		4762	2.4	
L3	5868	1.0		5206	4.2		7404	1.0		6648	2.0		9228	1.0		8038	4.4	
L4	5693	1.6		4274	5.6		7614	1.8		6745	5.4		8040	1.0		6181	3.4	
L5	4416	1.0		2500	3.6		7219	1.6		4715	2.8		8592	1.0		4116	2.0	
L26	5376	1.0		907	1.2		6864	1.0		270	1.0		8460	1.0		151	0.6	
L27	6308	1.4		744	1.0		7632	1.0		2696	1.4		9780	1.0		1311	3.0	
L28	5699	1.8		29	0.4		7692	1.0		29	0.4		9576	1.0		72	0.6	
L29	5352	1.0		0	0.0		8064	1.4		0	0.0		9547	1.8		0	0.0	
L30	6689	1.2		1709	2.4		8436	1.0		2036	2.0		10260	1.0		3086	2.8	
L31	4739	2.2		1464	1.0		6600	2.2		1044	1.0		6405	4.8		2076	1.0	
L32	5112	1.0		2496	1.0		6900	1.0		3840	3.0		8232	1.0		6334	2.6	
L33	2452	2.0		972	2.0		3828	2.4		948	1.0		2878	2.2		1248	2.0	
L34	5776	1.6		3062	1.6		7642	1.6		3541	2.6		9896	1.8		6916	6.4	
L36	2383	1.8		1603	1.6		5676	1.0		0	0.0		7337	1.2		0	0.0	
L39	5453	1.6		2659	1.6		7846	1.4		1873	1.4		9518	1.2		2002	1.2	
L41	4742	2.6		1901	1.6		6141	4.0		1372	1.8		7817	4.2		2696	3.4	
L43	2899	1.6		1020	2.0		3563	3.0		3000	4.0		3858	2.8		0	0.0	
L44	4810	1.2		12	0.2		5892	1.0		2902	1.2		7560	1.2		4186	4.0	
L48	4908	1.0		1534	1.2		6540	1.0		0	0.0		7968	1.0		0	0.0	
L49	6120	1.0		3609	4.4		9694	1.4		2441	1.2		11933	1.6		2012	2.6	
L50	7603	1.2		77	0.4		9691	1.2		4348	5.6		12748	2.0		2688	1.6	
L53	4632	3.6		504	1.2		5331	5.0		979	2.4		7415	6.8		6728	6.8	
L54	1440	5.4		7	0.2		1508	6.2		7	0.2		2123	7.2		2386	1.4	
L58	3303	6.8		602	2.4		3659	8.4		374	1.8		4720	9.2		245	1.2	
L60	5281	1.4		10	0.2		6674	1.8		1008	1.2		8154	3.2		1993	6.4	
L61	2188	3.6		420	1.4		1687	5.4		1712	1.6		2529	7.4		2347	1.2	
L63	5592	1.0		0	0.0		8014	1.4		0	0.0		9014	1.2		3179	3.4	
L64	6360	1.0		3112	2.6		8376	1.0		2822	1.6		10176	1.0		3206	3.2	

* Total queue duration is in seconds.

In conclusion, *TDRP* could lead to a very substantial saving in queuing duration and travel time and could keep the system running very close to the free flow condition. Therefore, it has the potential to achieve an efficient use of the network facilities throughout the peak period under different levels of congestion (demand).

7.5.2 Comparison between the different *TDRP* scenarios

To explore the significance of charging some links and exempting others, this section compares between the results of the base case under the three different *TDRP* scenarios defined earlier.

The fluctuation of the weighted coefficient of variation of the generalized cost of the base demand under the three scenarios of *TDRP* is depicted in Figure (7.30). The convergence of the results under scenario (I) exhibits the lowest coefficient of variation, followed by scenario (II). The coefficient of variation of scenario (III) is similar to that before *TDRP*, but with less oscillation.

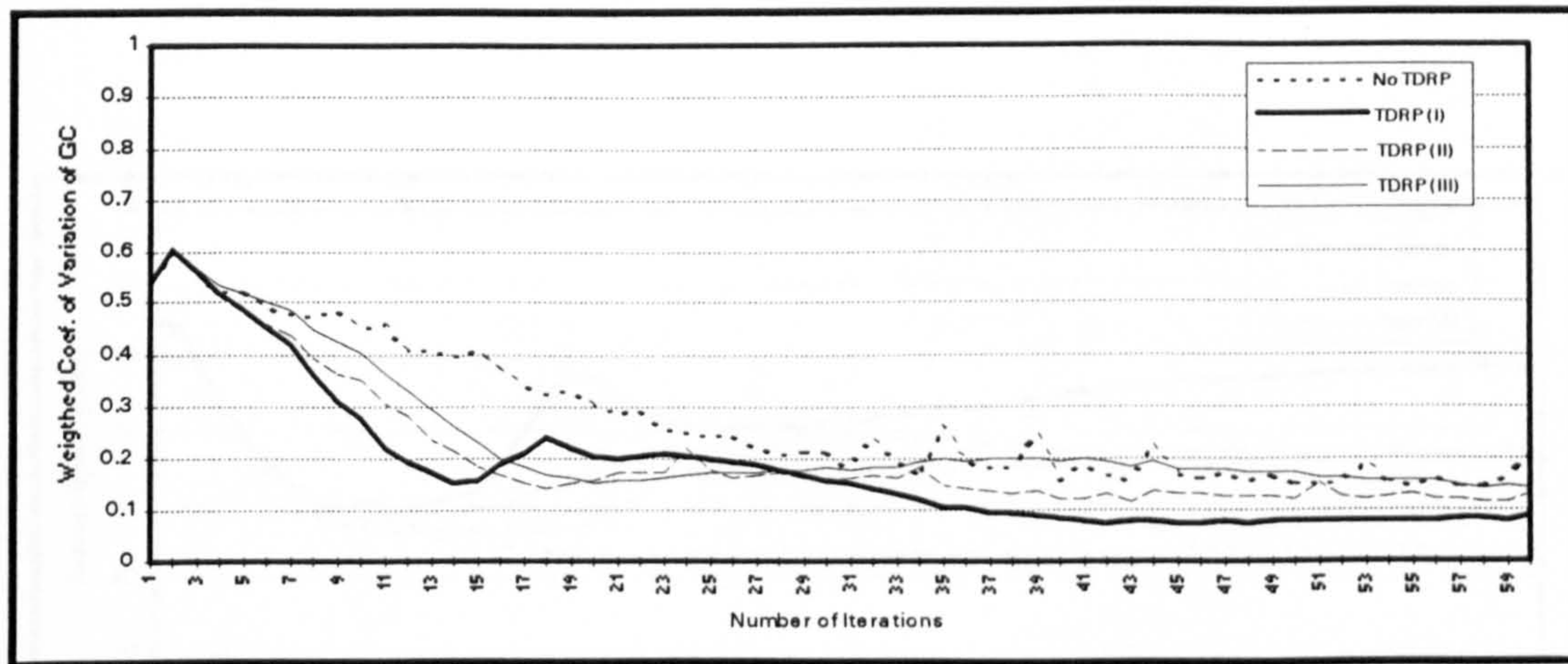


Figure (7.30): Fluctuation of the Weighted Coefficient of Variation of G. Cost under the base demand and the three *TDRP* Scenarios.

Table (7.8) summarises the base case results, while Figure (7.31) depicts the average generalized cost components under the three *TDRP* scenarios. The figure depicts that the three scenarios give a very substantial travel time savings and scenario (I) gives the highest value of *TDRP* and travel time saving. The travel time cost saving experienced under scenario (I) is around 60% higher than that experienced under the other two scenarios. It is also depicted that while the value of *TDRP* for scenario (II) is higher than that for scenario (III), the travel time savings given by the two scenarios are almost the same. Moreover, the figure shows that there is no significant difference in the average schedule delay between before and after *TDRP* under the three scenarios. Thus, as a result of the travel time saving, the travel speed significantly increases from 24.02 km/h to 48.33 km/h, 36.51 km/h and 37.39 km/h under scenarios (I), (II), and (III) respectively as depicted in Figure (7.32) and Table (7.8).

Table (7.8): Summary of the base case results under different *TDRP* scenarios.

<i>TDRP</i> Scenarios	Coeff. of Var.	Speed (km/h)	T. Time Cost	S. Delay Cost	<i>TDRP</i>	WGC	Total Q-duration	% of reduction in Q-duration
No <i>TDRP</i>	0.163	24.02	3.39	3.97	0.00	7.36	189911*	-----
Scenario (I)	0.080	48.33	1.17	4.13	3.25	8.54	65284	65.62
Scenario (II)	0.127	36.51	1.96	4.13	2.21	8.31	119992	36.82
Scenario (III)	0.155	37.39	1.91	3.87	1.97	7.75	102914	45.81

* Total queue duration is measured in seconds.

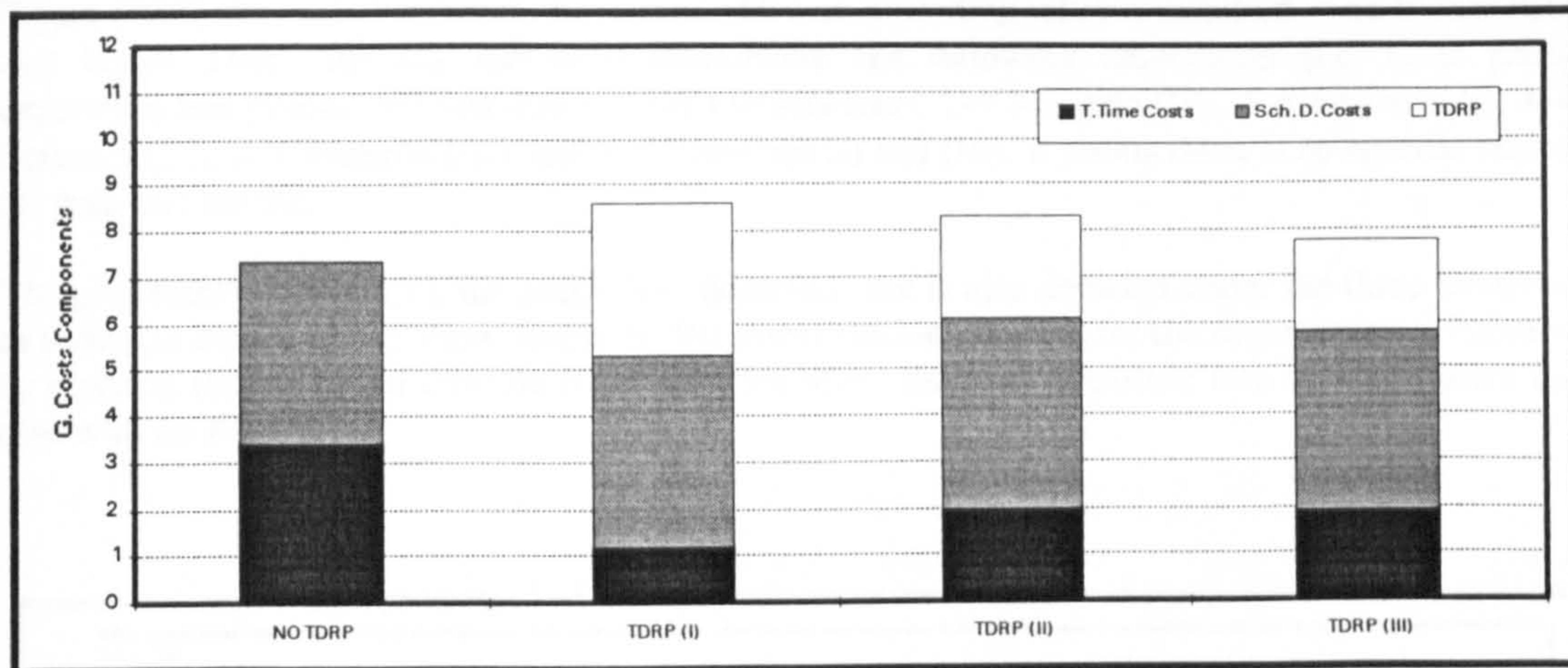


Figure (7.31): The Average G. Cost Components under the base demand and the three TDRP Scenarios.

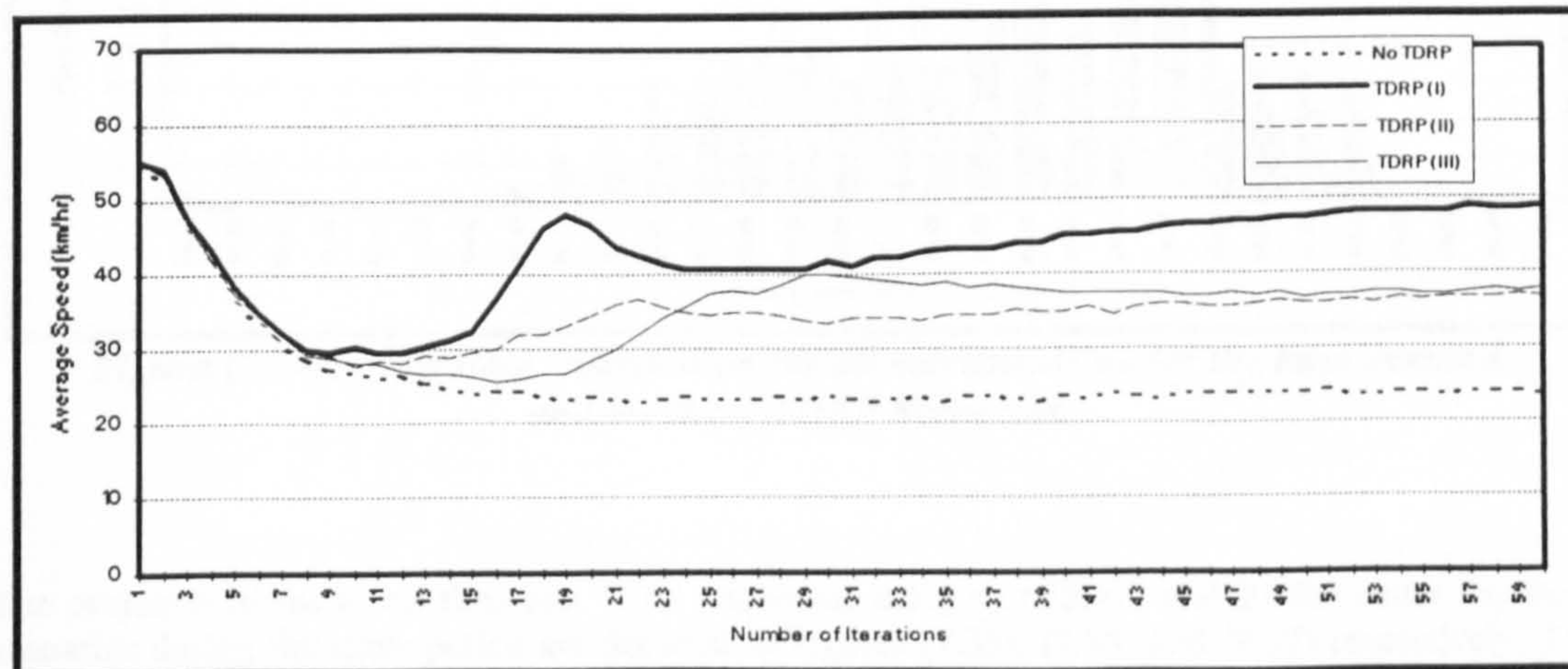


Figure (7.32): Fluctuation of the Average Speed under the base demand and the three TDRP Scenarios.

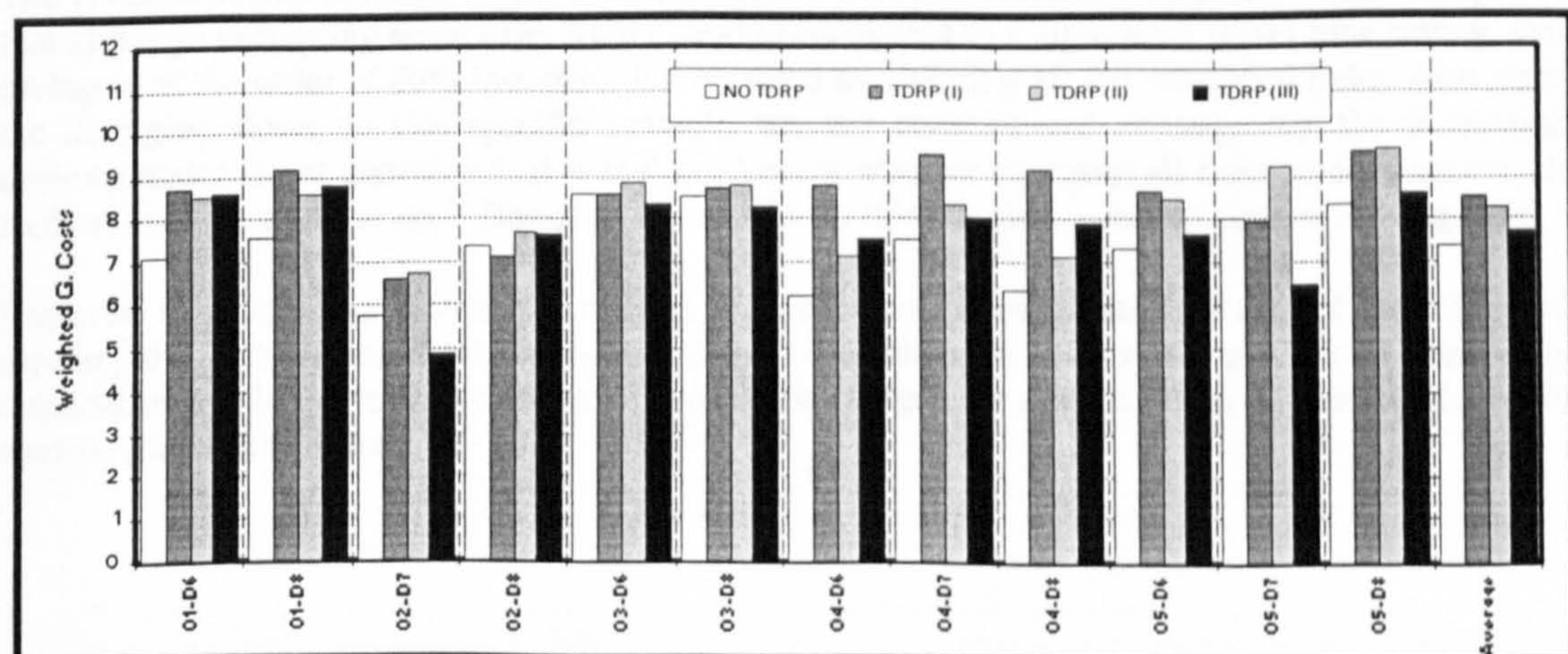


Figure (7.33): Weighted Generalized Cost for individuals OD pairs under the base demand and the three TDRP Scenarios.

Figure (7.33) shows that although the generalized cost (including tolls) is expected to be higher than that before *TDRP* for any individual movements, the following movements (i.e. O-D pairs) experience less generalized cost under one or two scenarios: 2-7 Scenario (III), 2-8 Scenario (I), 3-6 Scenario (III), 3-8 scenario (III), and 5-7 Scenarios (I) and (III). It seems there is no specific reason for such odd results.

The phenomenon of shifting the peak of the departure rate is also depicted under the three scenarios as shown in Figure (7.34) below. Scenario (III) shows the highest peak for the departure rate, followed by scenario (II), and then scenario (I) which goes below the peak departure rate depicted under the case with no *TDRP*.

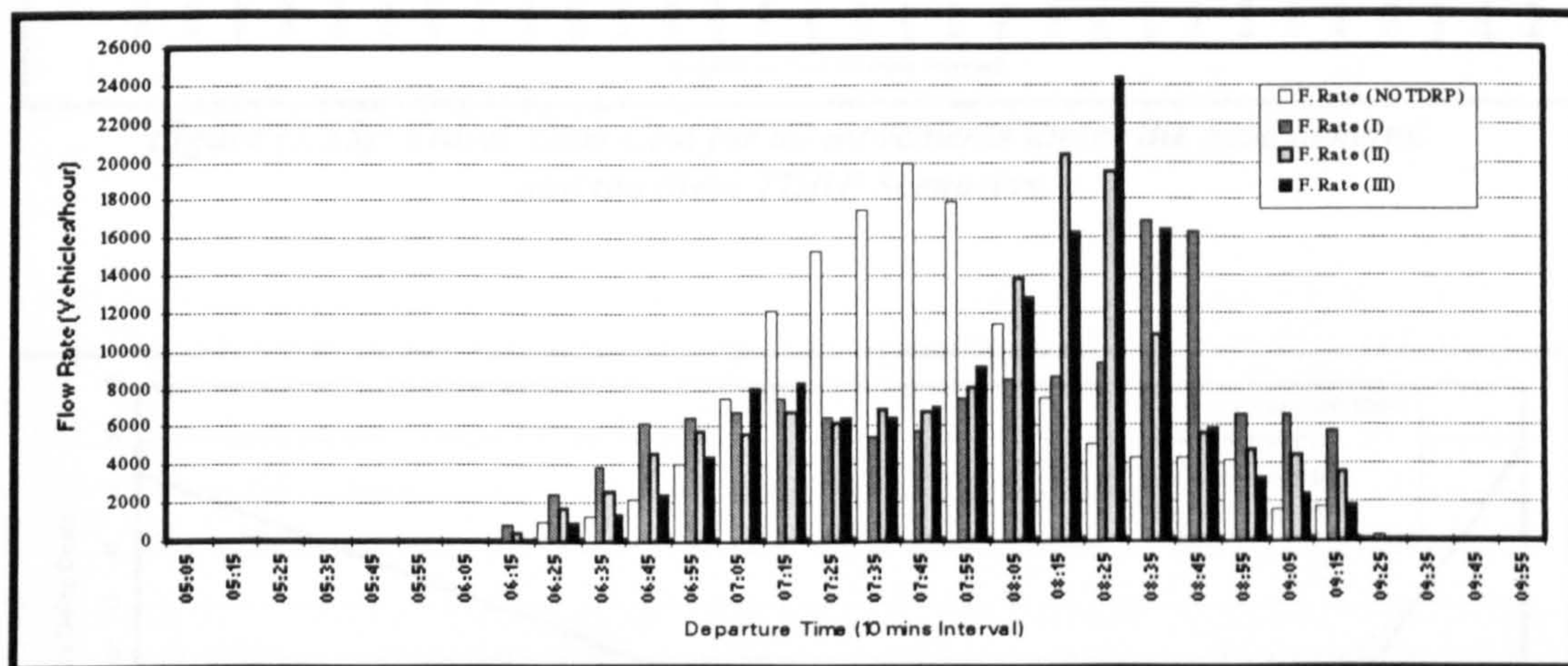


Figure (7.34): Flow Rate Distribution for all movements under the base demand and the three *TDRP* Scenarios.

The evolution of the travel time cost 'TTC', schedule delay cost 'SDC' and *TDRP* under the three scenarios during the study period are depicted in Figures (7.35), (7.36) and (7.37) respectively. The three figures show that each of the above generalized cost components has the same pattern under the three *TDRP* scenarios. Figure (7.35) indicates that the travel conditions under scenario (I) is very close to the free flow conditions. The typical pattern of the *TDRP* shown in Figure (7.37) is a triangle-shape which is similar to what has been demonstrated for a single bottleneck in chapter (5).

The conclusion that could be drawn from the above discussion and based on this specific example, is that although exempting some links from charging could lead to a significant travel time saving, that saving is of the order of 60% less than that obtained by charging all the congested links. Also since the diverging flows, in this specific example, are not constant and consequently the interaction between nodes is not considered, it is still not known whether charging all links under scenario (I) leads to over-charging or not? However, the answer for this question is left for further investigation.

However, in practice, as the number of links to be charged increases, the total cost of installing and running the system would substantially increase. Therefore, a cost/benefit analysis is essentially required to decide the optimal number of links to be charged for every specific case and under every specific traffic conditions.

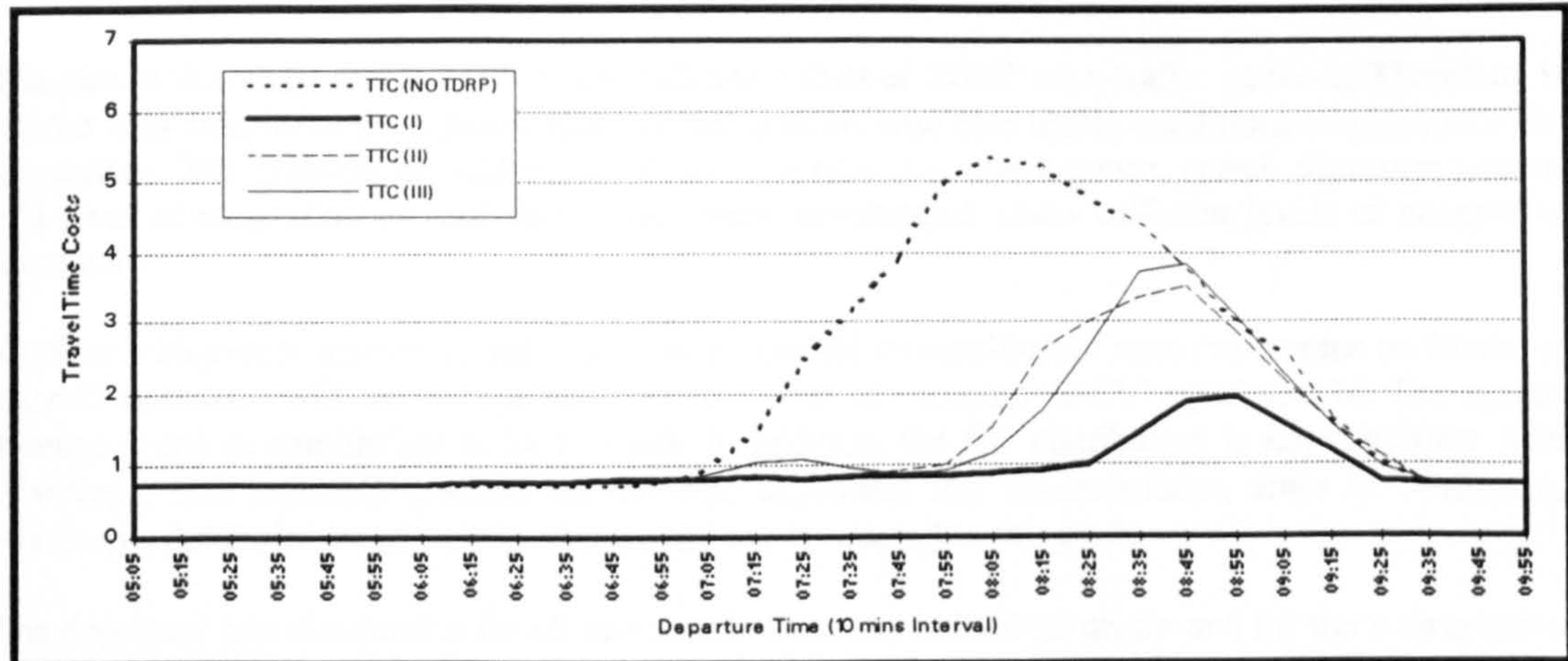


Figure (7.35): Travel Time Cost for all movements under the base demand and the three TDRP Scenarios.

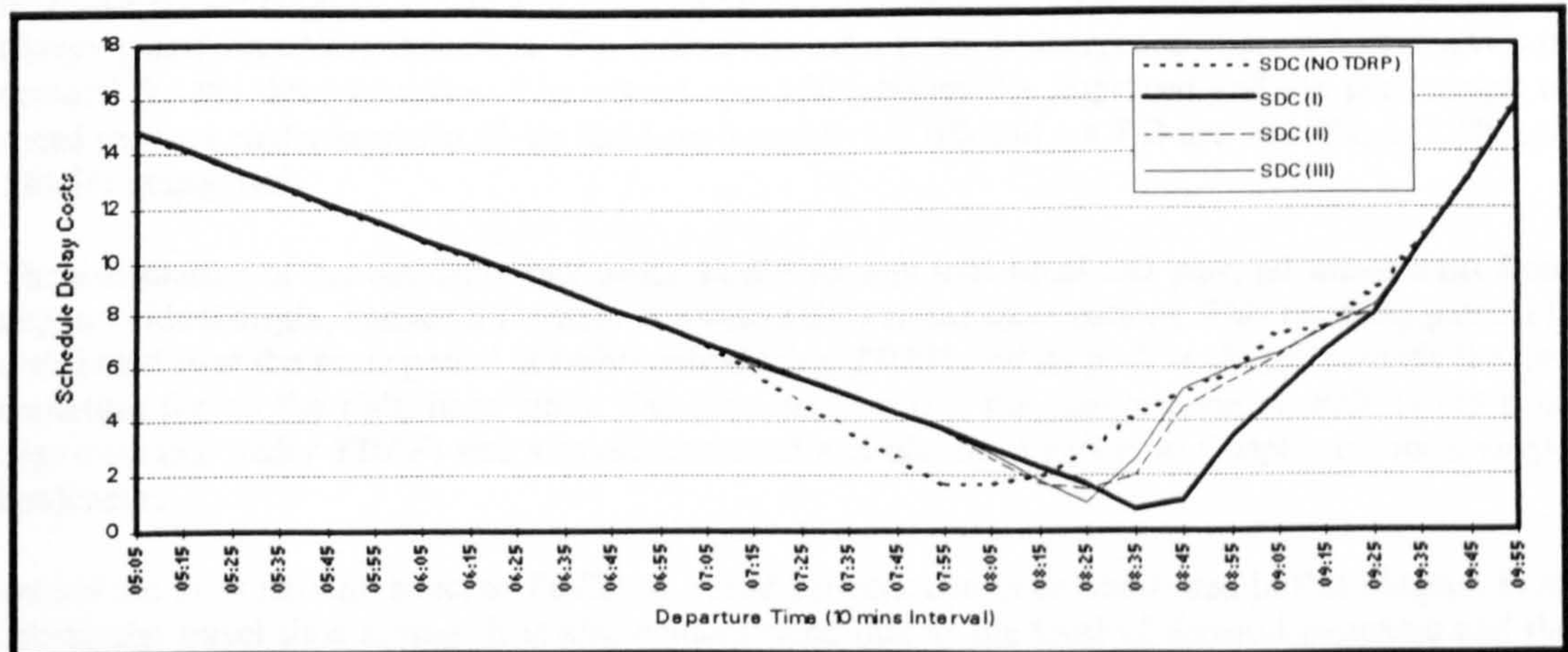


Figure (7.36): Schedule Delay Cost for all movements under the base demand and the three TDRP Scenarios.

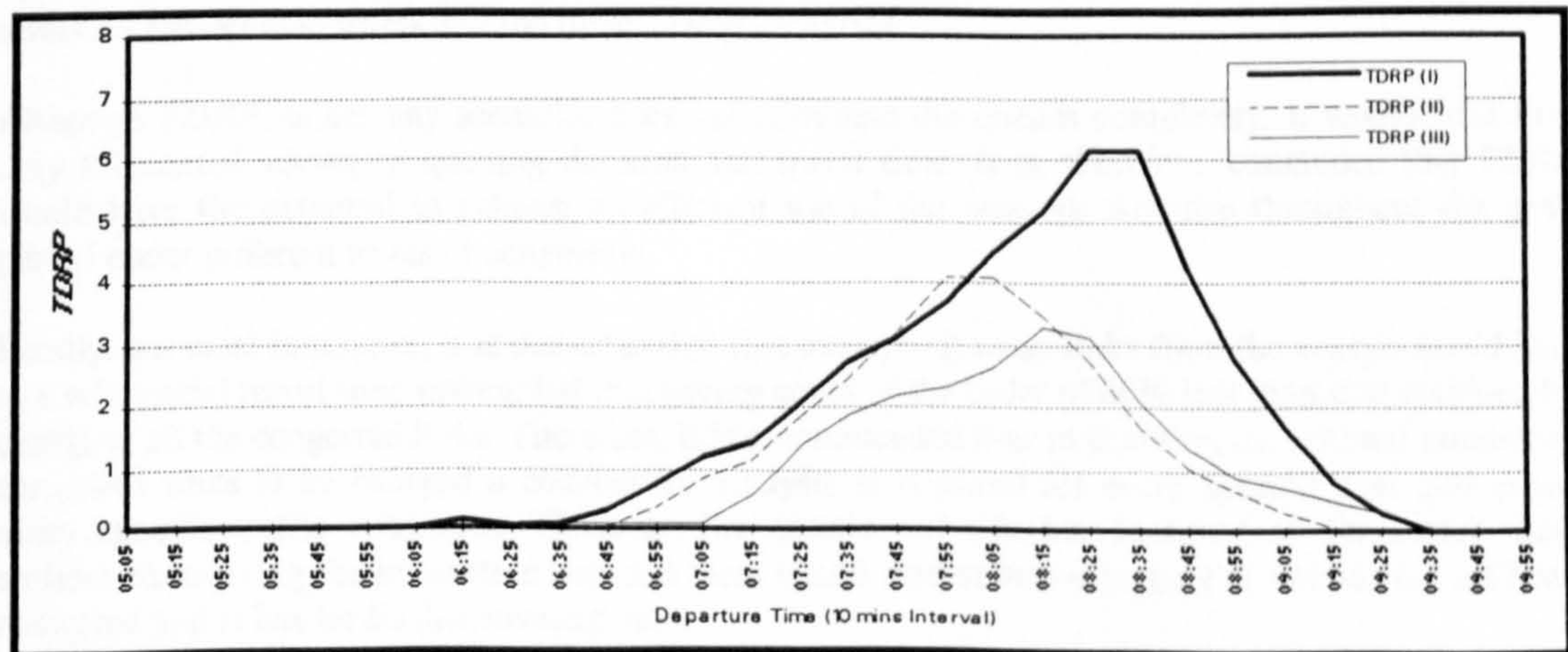


Figure (7.37): The average value of TDRP for all movements under the base demand and the three TDRP Scenarios.

7.6 Summary

The aim of this chapter is to evaluate the different effects of *TDRP* on a traffic network. Therefore, it started with describing the typical traffic network and the base case traffic conditions employed for the simulation. The different characteristics of the base case (i.e. convergence, speed, departure pattern and level of congestion on individual links) were investigated under different levels of congestion (demand).

Starting with pseudo-stationary state, the system reached its equilibrium state (with some oscillations) after 45 iterations, with an average weighted coefficient of variation 'WCV' equal to 0.16. The overall average speed at equilibrium is 24.02 km/h. In addition, the GC distribution is not consistent with Wardrop's first principle adjusted for the time dependent trip characteristics, since in contrast to Wardrop's deterministic approach, a stochastic one is adopted in this study.

The departure rate distribution for all movements from any individual origin and for the network as a whole showed the same pattern, and the peak departure rate of this pattern is skewed towards early departures. Moreover, under the base demand, the number of the most congested links is 24 links (out of 64 links in total).

In this chapter, three different scenarios for *TDRP* were suggested to investigate the different effects of *TDRP* on the network system and to explore the significance of considering some links (nodes) for charging and exempting the others. The system showed a better convergence under different levels of demand for the three scenarios. The overall speed is substantially improved and the percentages of speed increase under scenario (I) for the base demand, 1.2 BD and 1.4 BD are 101.2%, 117.4% and 136.6% respectively.

The distribution of the departure rate under *TDRP* for any individual OD pair, all movements from any individual origin, and for the system as a whole showed the same pattern. This resulting pattern is distributed over the same period of before introducing *TDRP*, and its peak is shifted towards the late departure (i.e. to the right hand side). This phenomenon (i.e. the phenomenon of shifting the peak departure rate under *TDRP*) was also demonstrated and discussed earlier in Chapter (5) for a single bottleneck.

One of the most striking effect of *TDRP* on a road network that is demonstrated in this chapter, is its substantial travel time saving. It is also demonstrated that as the level of demand increases and the network becomes very congested, the average toll of *TDRP* increases and the travel time saving becomes substantial. On the other hand, the schedule delay cost for an average journey under different demand levels is slightly increased by introducing *TDRP*. However, while the sum of travel time cost and schedule delay cost for an average journey throughout the network decreases and produces some user benefits as a result of introducing *TDRP*, the total generalised cost (including tolls) increases under all the demand levels and the three *TDRP* scenarios.

Although *TDRP*, under any scenario, does not eliminate the queues completely, it would lead to a very substantial saving in queuing duration and travel time. It is, therefore, concluded that *TDRP* would have the potential to achieve an efficient use of the network facilities throughout the peak period under different levels of congestion.

Finally, but most important, it is demonstrated that exempting some links from the charge would lead to a substantial travel time saving, but that saving could be of the order of 60% less than that obtained by charging all the congested links. Therefore, it is recommended that in deciding the optimal number of congested links to be charged a cost/benefit analysis is required for every specific case and under every specific traffic conditions. However, the question of whether charging all (or some) links without considering the interaction between them would lead to over-charging or not has not yet been answered and is left for further investigation.

CHAPTER EIGHT

8. Comparative and Sensitivity Analysis

8.1 Introduction

8.2 Comparative Analysis

- 8.2.1 Comparing *TDRP* with Distance-Based Charging System.
- 8.2.2 Comparing *TDRP* with Time-Based Charging System.
- 8.2.3 Comparing *TDRP* with Delay-Based Charging System.
- 8.2.4 Comparing *TDRP* with Flat Charging System.
- 8.2.5 *TDRP* vs. alternative charging systems under different demand levels.

8.3 Sensitivity Analysis

- 8.3.1 The effects of changing the work-time flexibility on the results.
- 8.3.2 The effects of changing the values of the schedule delay parameters.

8.4 Summary

8.1 Introduction

The aim of this chapter is to demonstrate the significance of *TDRP* system and its implementation under different circumstances. Therefore, the chapter is divided into two parts. In the first part, a comparative analysis with other systems of charging is conducted to show the distinctive features of *TDRP*. The systems of charging used for this comparative analysis include:

- Distance-based charging system;
- Time-based charging system;
- Delay-based charging system; and
- Uniform (or flat) charging system.

The second part is specified for checking the sensitivity of the results under *TDRP* to different travel aspects: e.g. level of demand, work start time flexibility, and the key parameters of the user cost function.

8.2 Comparative Analysis

To carry out the comparative analysis, the above charging systems are considered under different levels of complexity, as shown in Table (8.1) below, and under three different demand levels: base demand, 1.2 and 1.4 base demand. As depicted in Table (8.1), the process of comparing between different charging systems is a three-dimension one. The first dimension represents the basis of charging (i.e. charging strategy); the second dimension, is the level of complexity of charging; and the third dimension is the level of congestion under consideration. The last dimension is very important, as for the same network, the optimal charging system could differ under different levels of congestion.

Table (8.1): Alternative charging systems and their levels of complexity.

<i>Levels of Complexity</i>	<i>Basis of charging systems</i>					
	<i>TDRP</i>		<i>Time-Based</i>	<i>Delay-Based</i>	<i>Distance-Based</i>	<i>Uniform</i>
	Node Interaction	No Node Interaction	<i>toll</i>	<i>toll</i>	<i>toll</i>	<i>toll</i>
1. All links are charged all times on an individual vehicle-basis.	--	--	*	*	--	--
2. All links are charged all times on a time-interval basis.	--	*	--	--	--	*
3. Only some links are charged all times on an individual vehicle-basis.	--	--	--	--	*	--
4. Only some links are charged all times on a time interval-basis.	--	*	--	--	--	--
5. Only some links are charged for some times on an individual vehicle-basis.	--	--	--	--	*	--
6. Only some links are charged for some times on a time interval-basis.	--	--	--	--	--	--

* The cases of charging considered in the scope of this study.

Initially, the four alternative charging systems and their implementations into the simulation model are briefly described each in turn below.

- **Distance-Based Charging**

Here each vehicle is charged according to the distance travelled, by that vehicle, in the charge area. This encourages drivers to make greater use of regions that have no charge, and it will also have the effect of reducing congestion within the charge area. On the other hand, it also encourages drivers to use shorter distance routes within the charge area, and this will have the effect of increasing congestion within this area.

In implementing distance-based charging in this study, a penalty in seconds per kilometre (s/km) is added to the generalised cost of traversing each link in the charge area.

- **Time-Based Charging**

Here each vehicle is charged in proportion to the time spent travelling in the charging area. Delay and free-flow travel time are charged equally. Drivers are expected to choose routes which reduce their own journey time in the charge area. This may perhaps be easier to implement than congestion pricing; but it has the drawback that uncongested and congested travel are equally priced and so departs very significantly from the delay approximation to marginal cost pricing described in Smith and Ghali (1992).

In implementing time-based charging in this study, a penalty equivalent to a multiple of the travel time is added to the generalised cost of traversing each link in the charge area.

- **Delay-Based Charging**

Here each vehicle is charged according to the component of travel time spent in delay (mostly in traffic queues) in the charge area, thereby encouraging the vehicle to avoid congestion points and so to choose routes that reduce queuing delays. This charge would be expected to vary at only a slow rate both within each day and from day to day. In this way the charge is predictable and so may be expected to influence drivers' choices in a beneficial way.

In implementing delay-based charging in this study, a penalty equivalent to a multiple of the delay experienced over each link, is added to the generalised cost of traversing that link in the charge area.

- **Uniform charging system (Flat Toll)**

The flat charge suggested for the comparative analysis in this study is an approximation of the time-dependent road pricing system. The value of the charge is calculated for every individual link from the *TDRP* on that link throughout the simulation period as given by Equation (8.1) below:-

$$FT_L = \frac{1}{(T - T_0)} \cdot \int_{T_0}^T TDRP_L(t) \cdot dt \quad \dots\dots\dots (8.1)$$

Where

FT_L	is the flat toll over link L.
$TDRP_L(t)$	is the component of <i>TDRP</i> at time t over link L.
T_0 & T	are the starting and ending times of the simulation period.
t	is the time interval (10 seconds).

In implementing this system of charging, *TDRP* is calculated and adjusted using the day-to-day adjustment process as explained earlier. Then, the different values of *TDRP* due over every individual link throughout the study period are summed up and flattened to give a single flat charge over every individual link throughout the next iteration. Thus, the charge during any iteration will be based on the queuing conditions of the iteration preceding the previous one (i.e. there is always a two-days lag between the charges incurred and the queuing conditions, unlike *TDRP* where the charge is based on the queuing conditions of the previous day).

The following sub-sections evaluate and compare between *TDRP* and every individual charging system. Then, a summary of the comparative analysis between *TDRP* and all the alternative charging systems under different demand levels is discussed.

8.2.1 Comparing *TDRP* with Distance-Based Charging System

The designated area for charging under distance-based charging system includes all the links in the city centre as well as the links leading to it. The total number of links considered for charging is 25 links (*L26, L27, L28, L29, L30, L32, L36, L37, L39, L48, L49, L50, L51, L52, L53, L54, L55, L57, L58, L59, L60, L61, L62, L63 and L64*). Initially, these links are charged all the time during the study period. The average generalized cost components are examined for before and after distance-based charges under five different charging rates per kilometre traversed: 60 s/km, 120 s/km, 240 s/km, 300 s/km and 600 s/km. The results are depicted in Figure (8.1). The Figure shows that distance-based charging fails to give any significant benefits under the different charging rates considered. This is because distance-based charges encourage travellers to divert to shorter and probably more congested routes, thus aggravating congestion. Also, as the charging rate increases the level of congestion increases and consequently all the travel cost components increase.

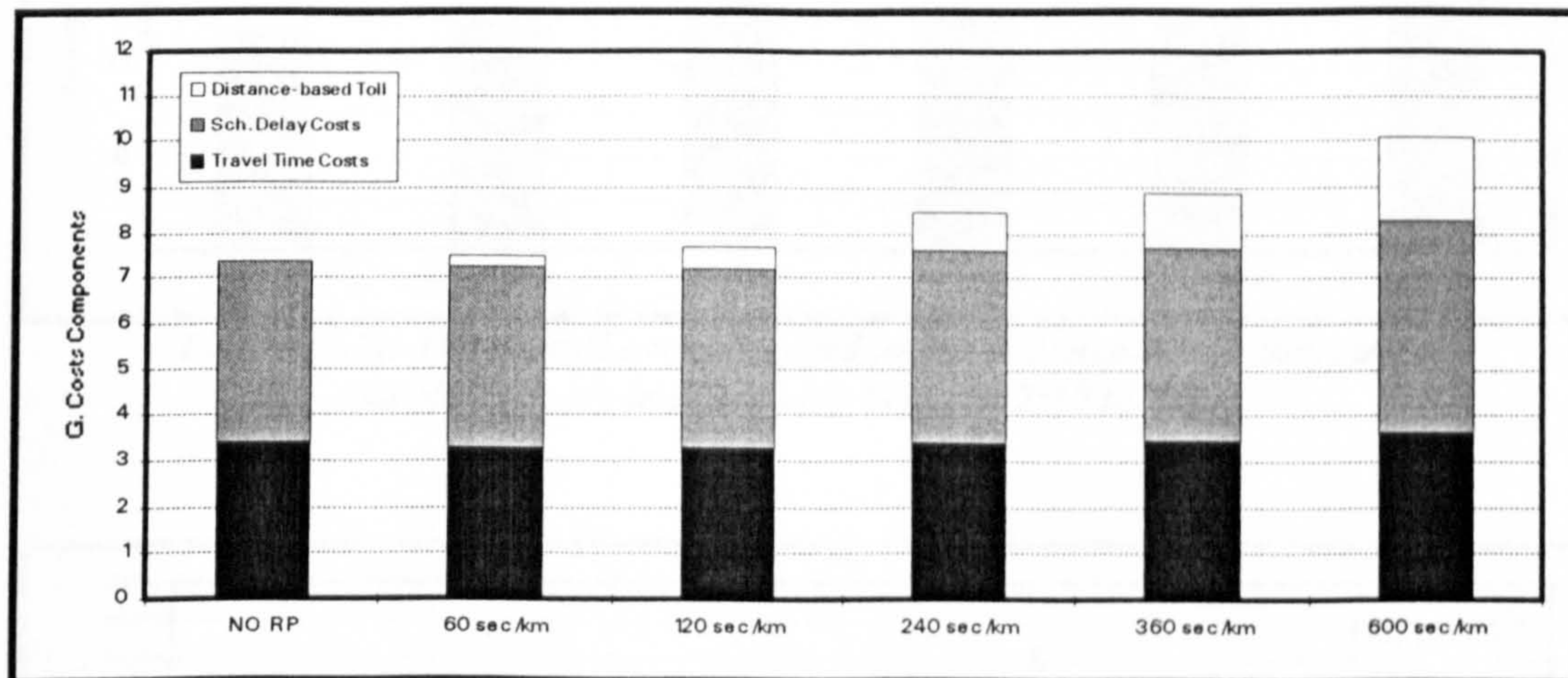


Figure (8.1): The average generalized cost components under distance-based charging (the charges are imposed over the whole study period).

However, distance-based charging system is examined again for the same set of links but only during the most congested period (between 7:30 a.m. to 8:30 a.m.). The results of the average generalized cost components, in this case, versus *TDRP* are depicted in Figure (8.2). The Figure exhibits very small benefits under the distance-based charges which are distinctively very much less than that depicted under *TDRP*. In addition, Figures (8.2) and (8.3) exhibit that as the charging rate increases, the travel time saving as well as the overall average speed fall, and the charging rate of 240 s/km exhibits the highest travel time saving and overall average speed compared with that depicted under the other charging rates.

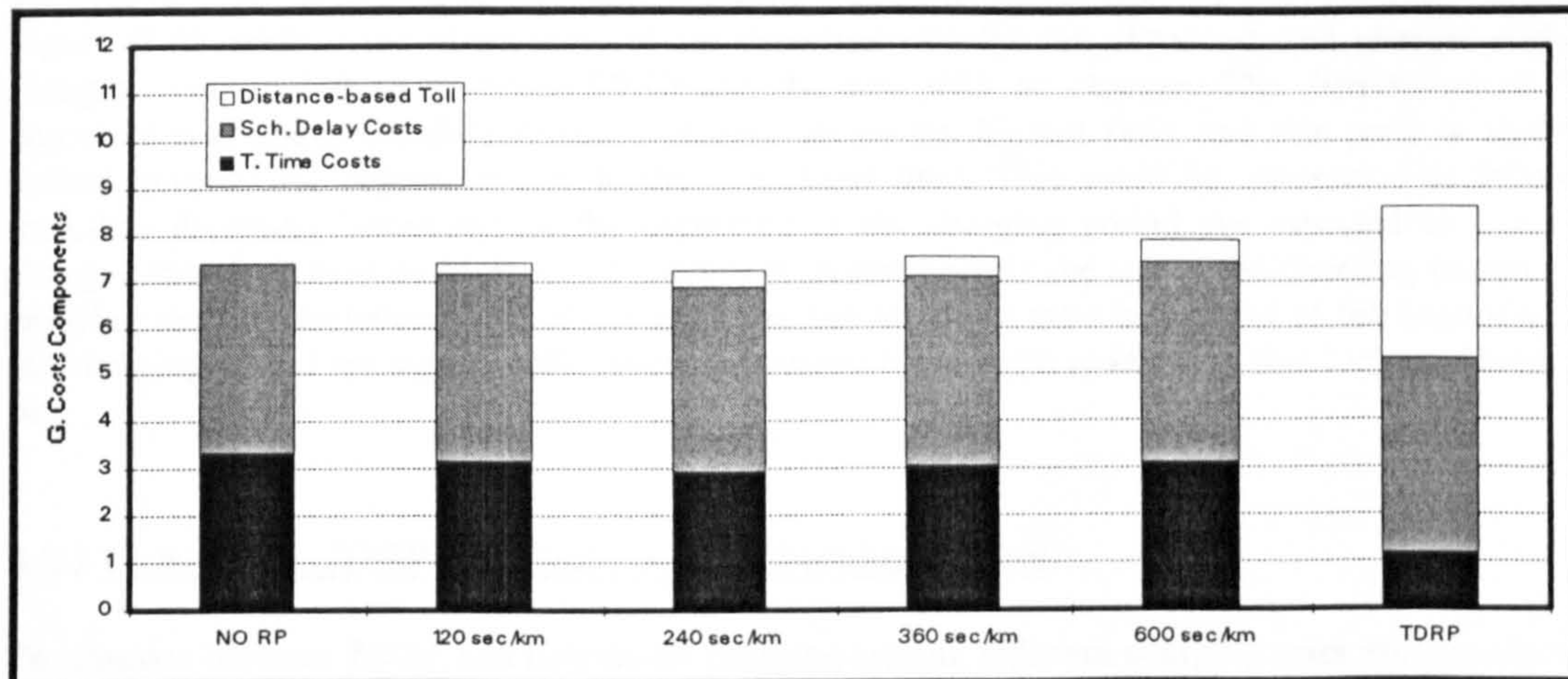


Figure (8.2): The average generalized cost components under distance-based charging (the charges are imposed only between 7:30 to 8:30).

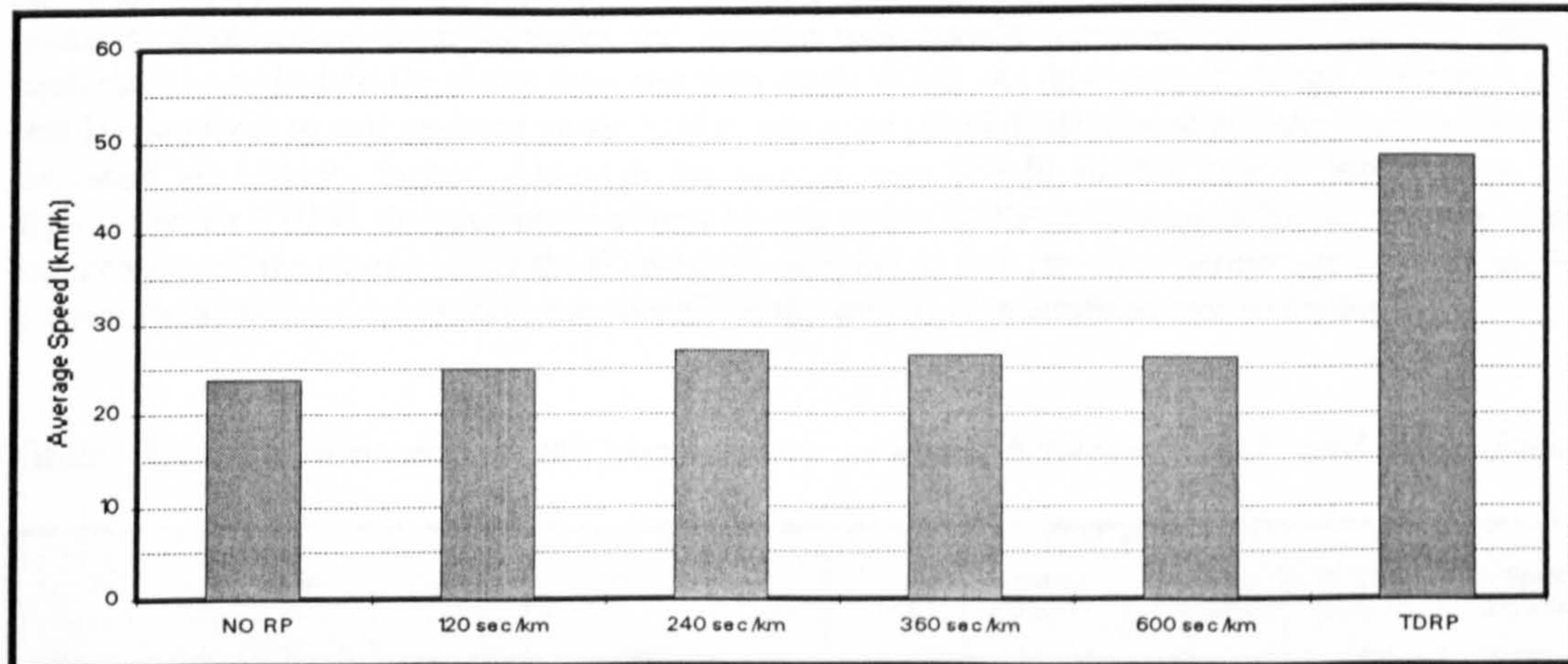


Figure (8.3): The overall average speed under distance-based charging (the charges are imposed only between 7:30 to 8:30).

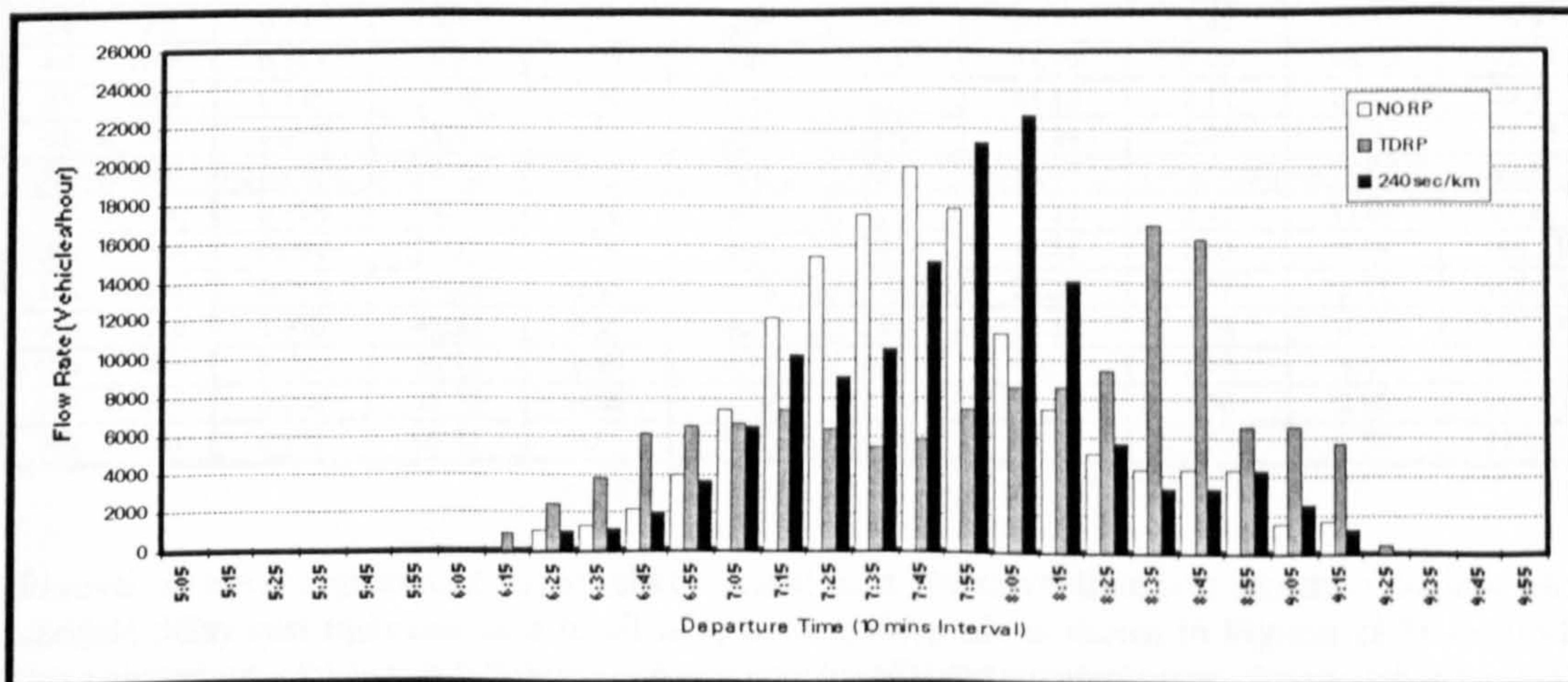


Figure (8.4): The departure time distribution for all movements under distance-based charging (the charges are imposed only between 7:30 to 8:30).

To illustrate the influences of the charges during a certain time period on rescheduling journeys, Figure (8.4) exhibits the distribution of the departure rate for the distance-based charges with a charging rate of 240 s/km versus *TDRP* and the case with no charges. The distribution of the departure rate under the distance-based charges shows the highest peak and this peak is slightly shifted towards late departures (i.e. to the right hand side). This could be interpreted as follows: travellers departing before and at the beginning of the charging period are expected to face the charges, therefore, they try to avoid the charges by departing near the end of the charging period and enjoying most of the journey free of charge. Thus, the departure rates before and at the beginning of the charging period are significantly fallen and pushed to the right to build up this highest departure rate.

8.2.2 Comparing *TDRP* with Time-Based Charging System.

To compare between *TDRP* and time-based charging system, different charging rates are considered: 0.5, 1.0, 1.5, 2.0, ..., and 10.0 the travel time experienced over every individual link. The results show that as the charge rate increases the user benefits, in terms of travel time and schedule delay saving and the overall speed increase as shown in Figure (8.5). In order to demonstrate whether that increase in user benefits continues with increasing the charge rate, higher multiples of travel time are also explored as presented in Table (8.2) and depicted in Figure (8.6). Table (8.2) reveals that as the multiple of the travel time increases the average user benefit increases till the charges become equivalent to eight-fold the travel time and then starts to fall. At this level of charge, the travel time cost is equivalent to that depicted under *TDRP*, while the schedule delay cost and the average speed in the latter are slightly higher. Although the average user benefit in this case is higher than that depicted under *TDRP*, the percentage of user benefit under *TDRP* is very much higher because of the higher value of the charge under the time-based charging in this case (the percentage of user benefits is equal to the ratio of the average user benefit to the weighted generalized cost including tolls).

Table (8.2): A summary of the comparative analysis between *TDRP* and Time-Based charging.

Tolls	Weighted Coeff. of Variation	Weighted G. Cost	T. Time Cost	Schedule D. Cost	Average Toll	Average Speed	Av. User Benefits	% User Benefits	% Speed Increase
No RP	0.160	7.36	3.39	3.97	0.00	24.02	—	—	—
<i>TDRP</i>	0.082	8.54	1.17	4.13	3.25	48.33	2.06	24.12	101.21
0.5 T. Time	0.202	7.99	2.71	3.93	1.35	27.90	0.72	9.01	16.15
1.0 T. Time	0.137	8.55	2.34	3.87	2.34	30.41	1.15	13.45	26.60
1.5 T. Time	0.168	8.75	1.93	3.92	2.90	34.59	1.51	17.26	44.00
2.0 T. Time	0.138	8.98	1.70	3.88	3.40	36.69	1.78	19.82	52.75
3.0 T. Time	0.137	10.31	1.60	3.92	4.80	38.32	1.84	17.85	59.53
4.0 T. Time	0.153	10.69	1.35	3.92	5.41	41.70	2.09	19.55	73.61
5.0 T. Time	0.143	11.70	1.30	3.92	6.48	43.18	2.14	18.29	79.77
6.0 T. Time	0.135	13.28	1.33	3.94	8.00	42.24	2.09	15.74	75.85
8.0 T. Time	0.121	14.55	1.17	4.01	9.37	44.96	2.18	14.98	87.18
10.0 T. Time	0.110	16.69	1.15	4.04	11.50	45.37	2.17	13.00	88.88
20.0 T. Time	0.084	25.37	1.00	4.31	20.06	48.81	2.05	8.08	103.21
30.0 T. Time	0.064	33.29	0.93	4.53	27.83	51.29	1.90	5.71	113.53
40.0 T. Time	0.070	41.82	0.91	4.66	36.25	52.18	1.79	4.28	117.24
50.0 T. Time	0.060	49.92	0.89	4.74	44.29	53.00	1.73	3.47	120.65
60.0 T. Time	0.068	58.55	0.88	4.78	52.89	53.30	1.70	2.90	121.90
70.0 T. Time	0.051	66.41	0.87	4.83	60.71	53.76	1.66	2.50	123.81

However, as the charge exceeds eight-fold the travel time, the travel time cost decreases slightly while schedule delay cost increases as a result of spreading the peak as shown in Figures (8.7) and (8.8). This consequently leads to a fall in the average user benefit and its percentage. On the other hand, the overall average speed continues in its increase with an increasing level of charge and would approach the free-flow speed with a very high multiple of travel time as shown in Figure (8.9).

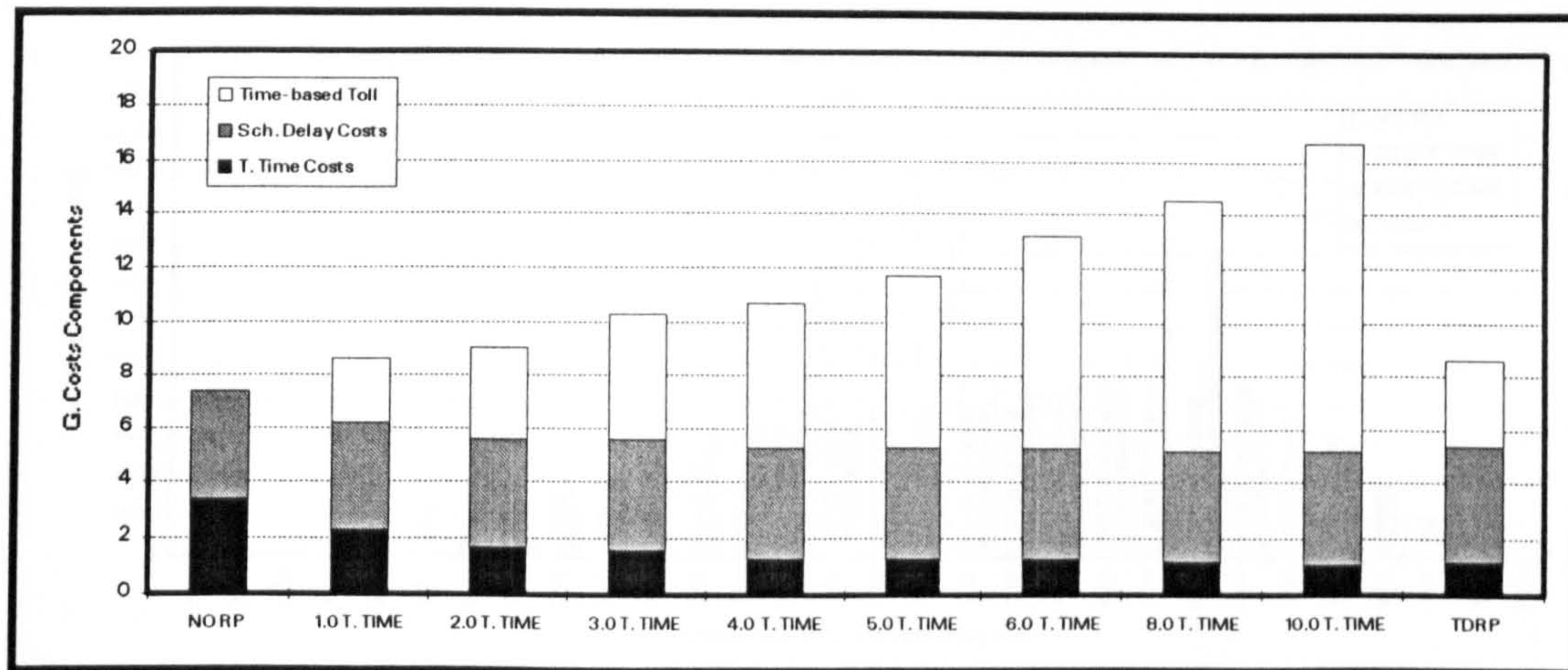


Figure (8.5): The average generalized cost components under time-based charging (for a multiple of travel time from 1.0 to 10.0).

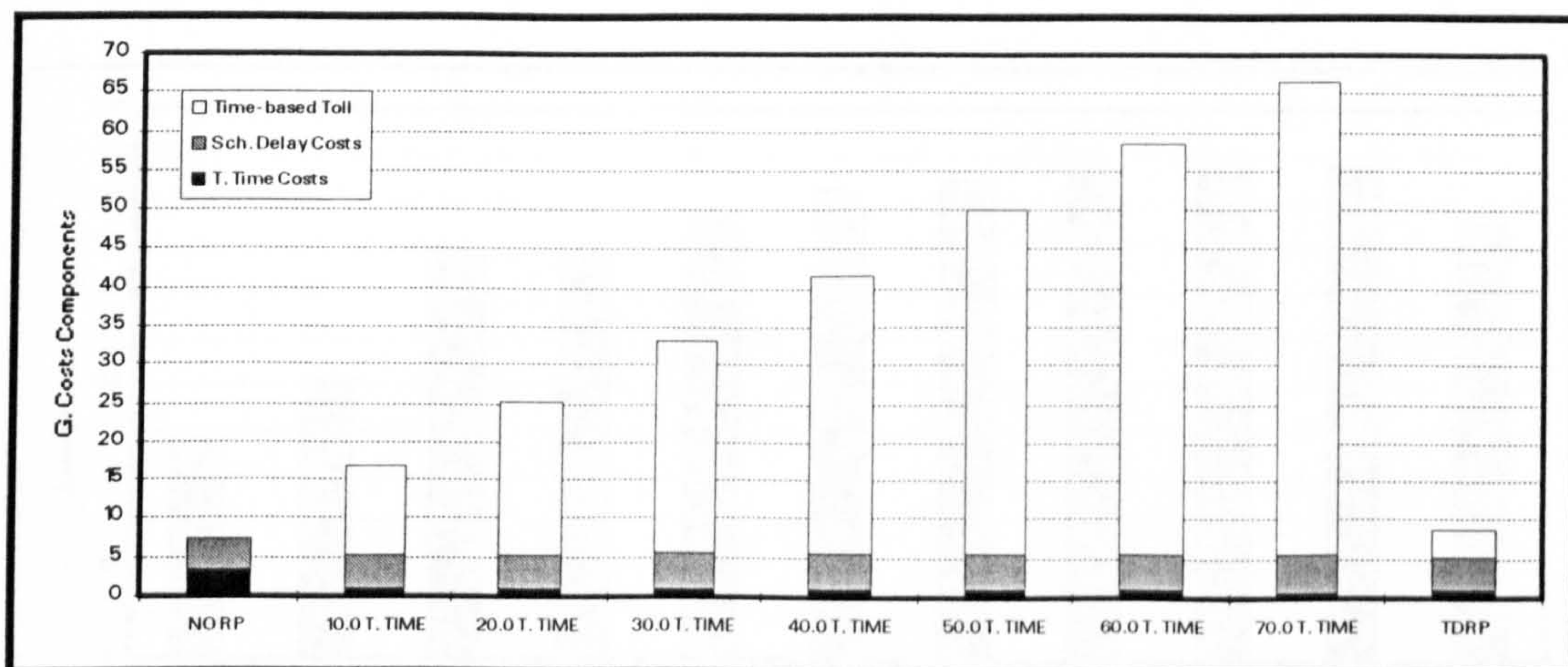


Figure (8.6): The average generalized cost components under time-based charging (for a multiple of travel time between 10.0 to 70.0).

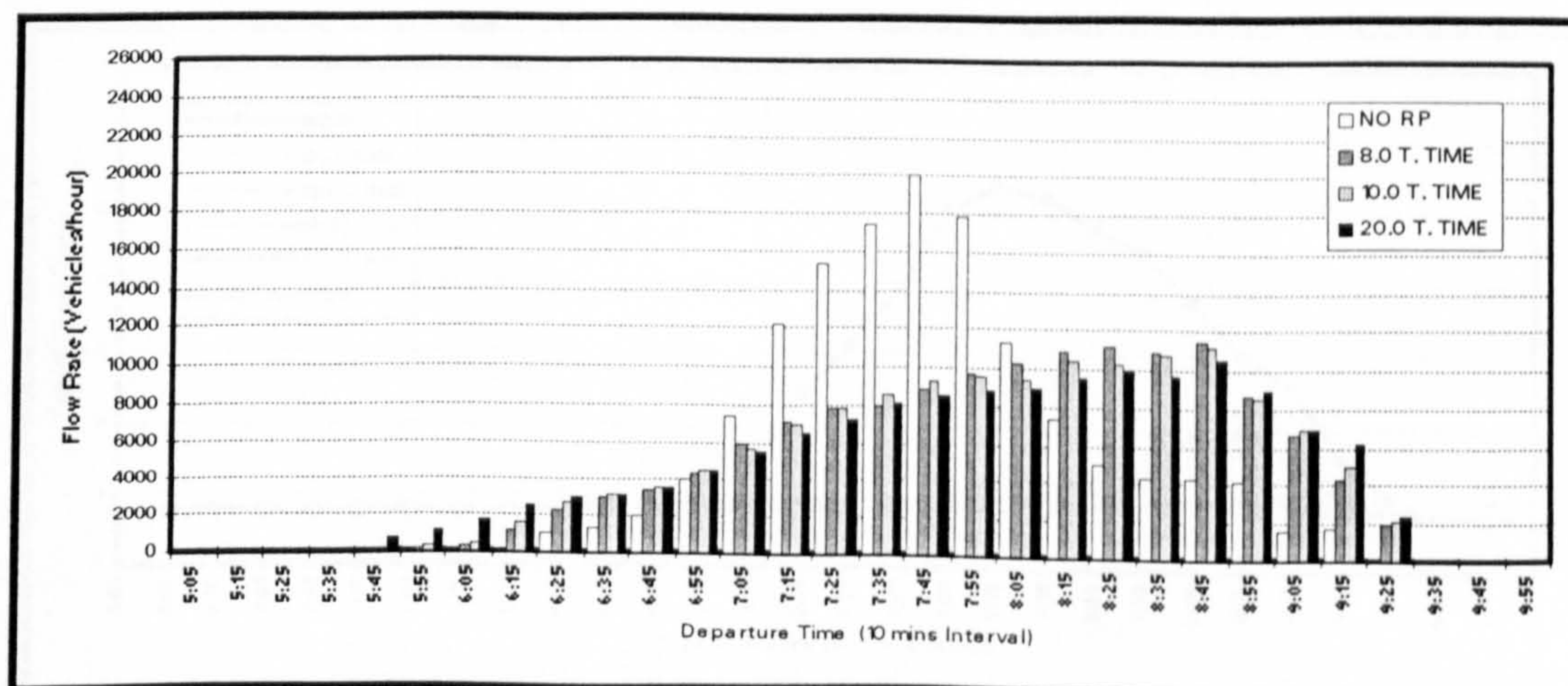


Figure (8.7): The departure time distribution for all movements under time-based charging (for a multiple of travel time equal to 8.0, 10.0 and 20.0).

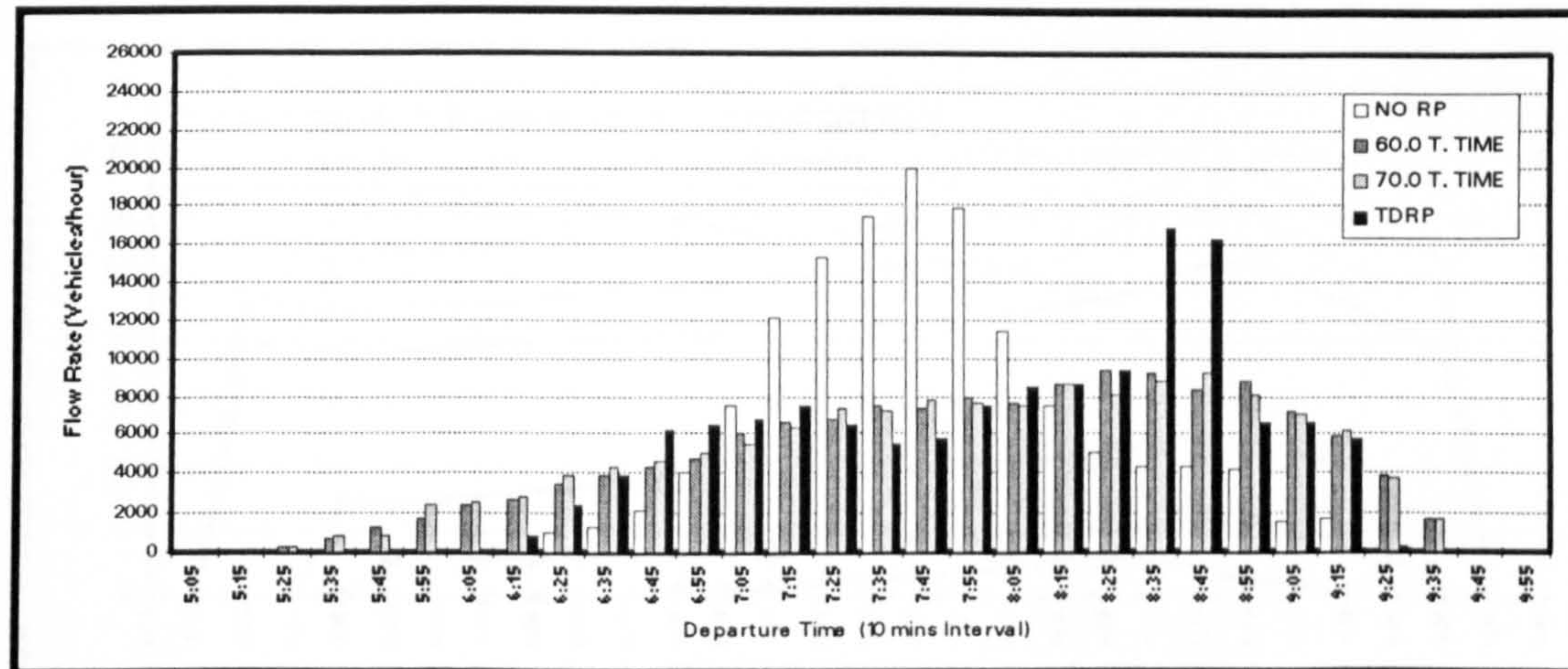


Figure (8.8): The departure time distribution for all movements under time-based charging (for a multiple of travel time equal to 60.0 and 70.0).

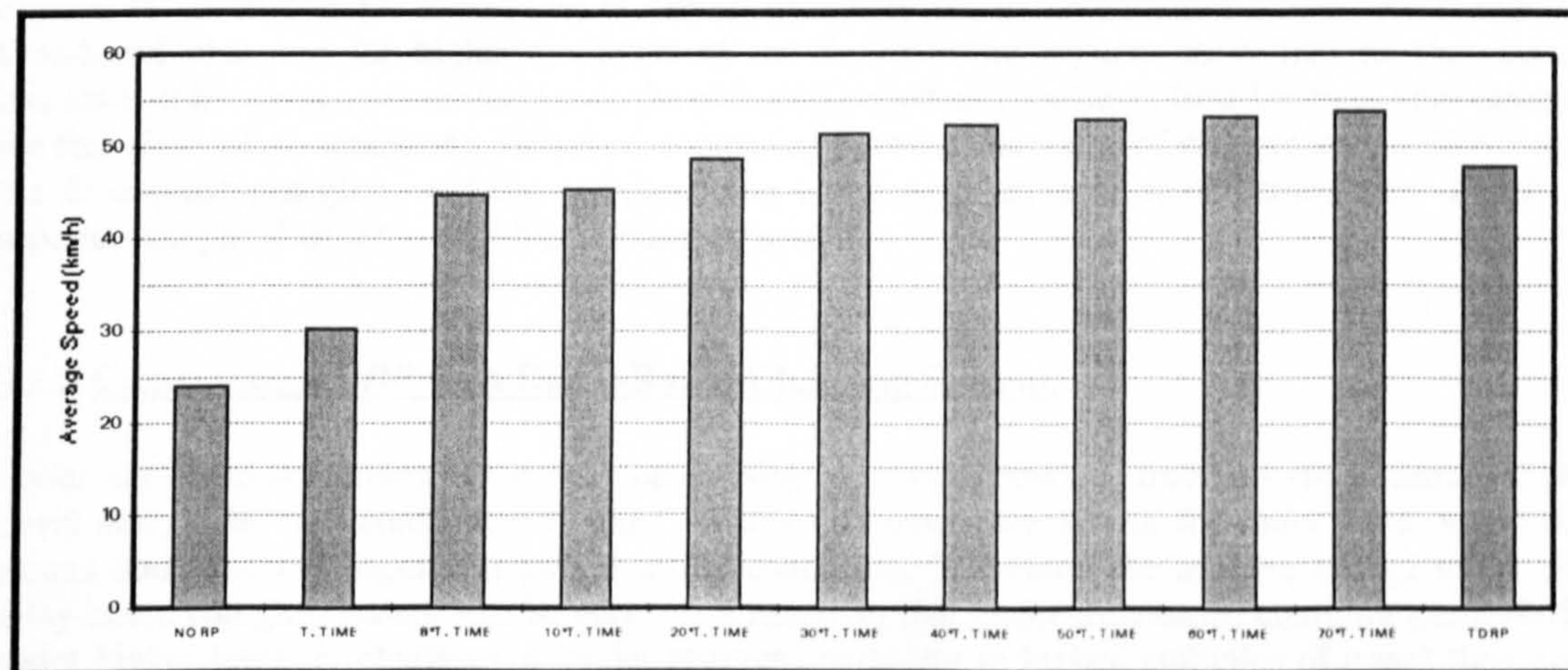


Figure (8.9): The overall average speed under time-based charging.

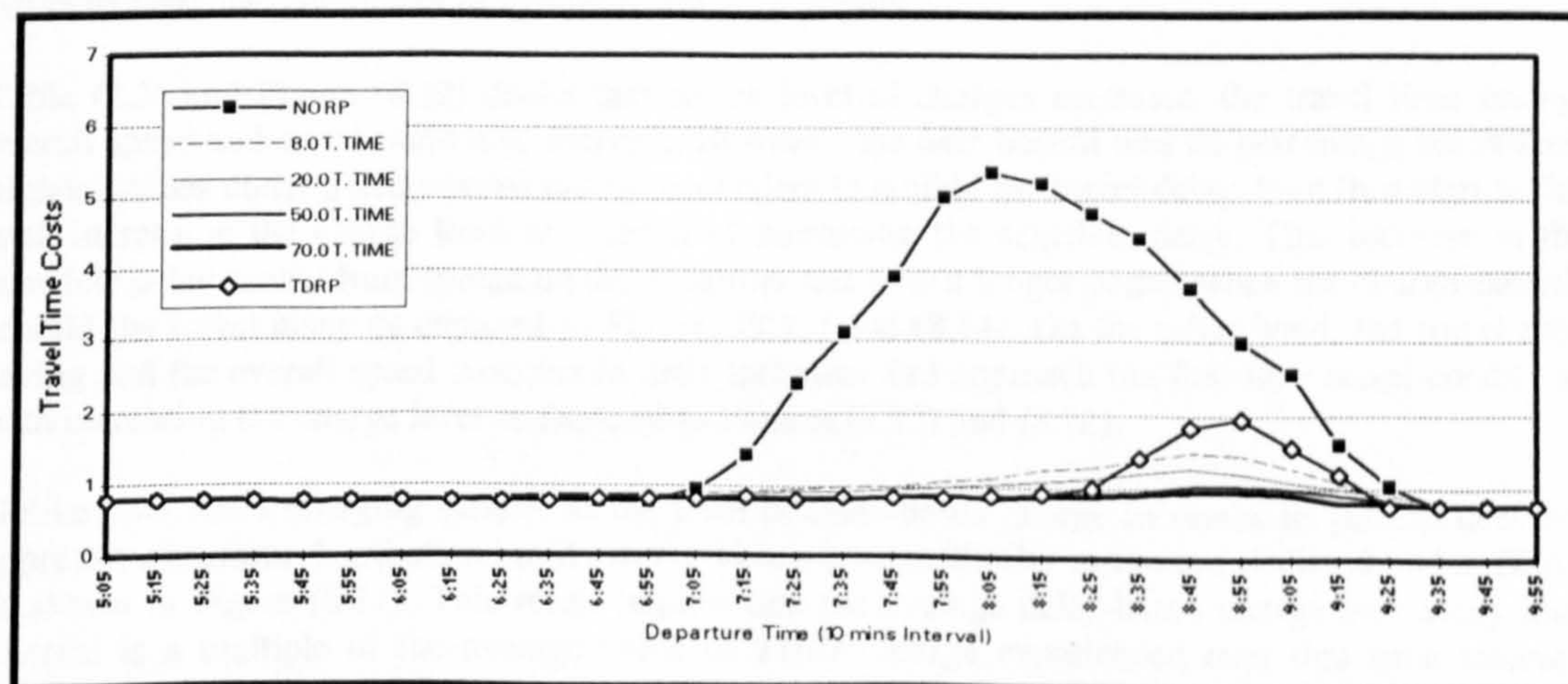


Figure (8.10): Travel time costs for all movements under time-based charging (base-demand).

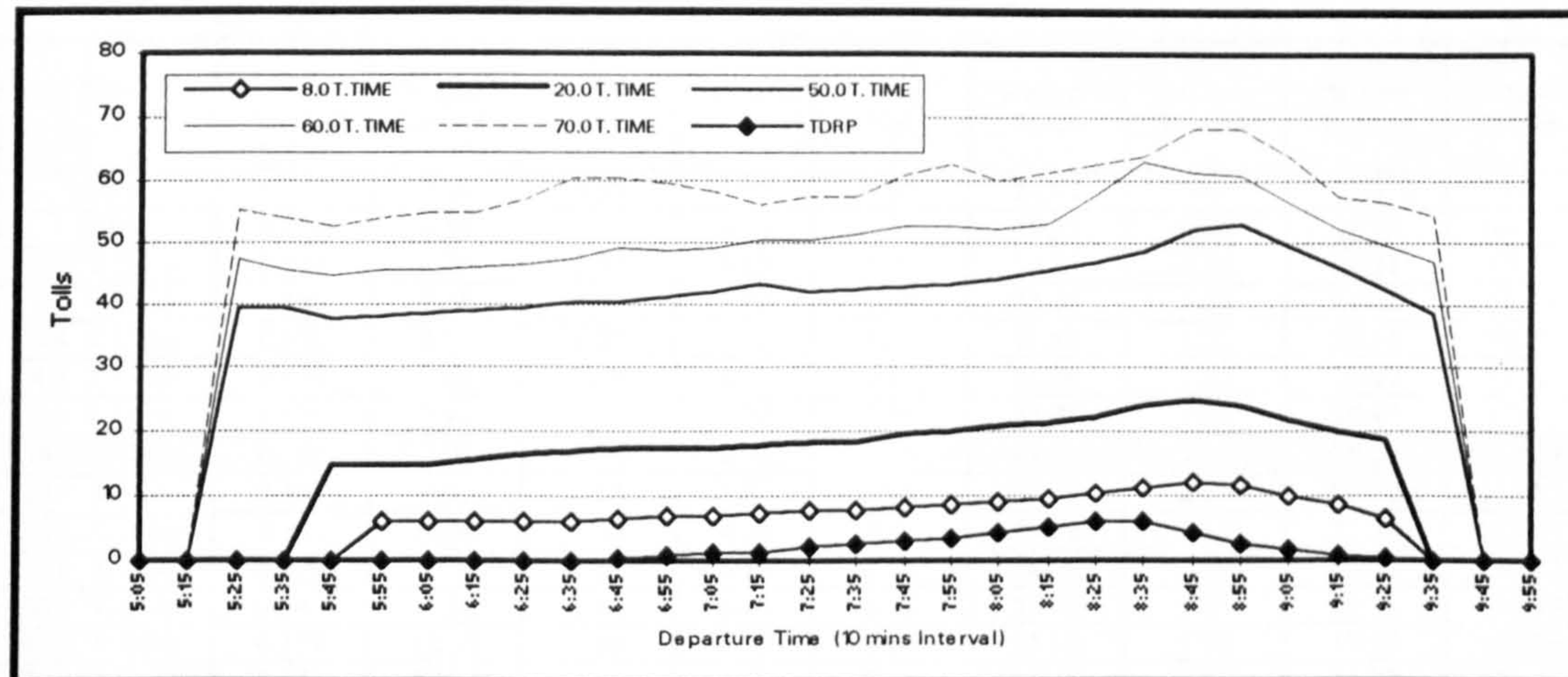


Figure (8.11): The distribution of the average toll for all movements under time-based charging.

Figures (8.10) and (8.11) depict that the travel time cost and the toll pattern under *TDRP* and the time-based charging for higher multiples of travel time. The Figures show that as the charge increases, it becomes more uniform over the charging period and the travel time becomes very close to the free-flow travel conditions. However, comparing between the value of charges under *TDRP* and the time-based charging, and the corresponding travel time savings, demonstrates that *TDRP* is superior compared with the time-based charging system.

8.2.3 Comparing *TDRP* with Delay-Based Charging System.

Under any travel conditions, the travel time consists of two components: free flow travel time and the travel delay. The first component is fixed for any OD movement taking the same route, while the second component is dependent on the traffic conditions. Therefore, the average charge under the delay-based charging system will be very much less than that under time-based charging particularly under higher levels of charges (i.e. under charges equivalent to higher multiples of travel time and delay respectively). However, the results of the comparison between *TDRP* and delay-based charging are summarised in Table (8.3). Comparing between the weighted generalized costs 'WGC' and its coefficient of variation in Tables (8.2) and (8.3) shows that for higher multiples of time and delay, the 'WGC' under delay-based charging is very much less than that depicted under time-based charging, while its weighted coefficient of variation 'WCV' is higher.

Table (8.3) and Figure (8.12) depict that as the level of charges increases, the travel time saving, overall speed and user benefit also increase. However, the user benefit and its percentage reach their highest values under a delay-based charge equivalent to tenfold the travel delay, then they start to fall with increasing the charge level as a result of increasing the schedule delay. This increase in the schedule delay results from spreading the departure rate over a longer period when the charge exceeds tenfold the travel delay as depicted in Figures (8.13) and (8.14). On the other hand, the travel time saving and the overall speed continue in their increases and approach the free-flow travel conditions with increasing the charge level as depicted in Figures (8.15) and (8.16).

Unlike time-based charging system, as the level of delay-based charge increases its pattern does not approach a uniform distribution but it always takes a pattern similar to the one depicted under *TDRP* as shown in Figure (8.17). This result implies that the average delay-based charge over every time interval is a multiple of the average value of *TDRP* charge experienced over that time interval. However, this multiple factor is not the same for all time intervals throughout the study period as it is clearly depicted in Figure (8.17) under charges equivalent to ten and twenty-fold the travel delay.

Table (8.3): A summary of the comparative analysis between TDRP and Delay-Based charging.

Tolls	Weighted Coeff. of Variation	Weighted G. Cost	T. Time Cost	Schedule D. Cost	Average Toll	Average Speed	Av. User Benefits	% User Benefits	% Speed Increase
No RP	0.160	7.36	3.39	3.97	0.00	24.02	----	----	----
TDRP	0.082	8.54	1.17	4.13	3.25	48.33	2.06	24.12	101.21
1.0 T. Delay	0.169	6.85	1.94	3.83	1.08	34.89	1.59	23.21	45.25
2.0 T. Delay	0.146	7.78	1.92	3.80	2.06	36.18	1.64	21.08	50.62
4.0 T. Delay	0.176	8.00	1.57	3.78	2.66	41.86	2.01	25.13	74.27
6.0 T. Delay	0.179	7.98	1.39	3.73	2.87	45.20	2.24	28.07	88.18
8.0 T. Delay	0.150	7.94	1.28	3.74	2.92	47.47	2.34	29.47	97.63
10.0 T. Delay	0.136	7.96	1.22	3.73	3.01	49.02	2.41	30.28	104.08
20.0 T. Delay	0.123	9.01	1.13	3.90	3.99	52.32	2.33	25.86	117.82
30.0 T. Delay	0.140	10.00	1.09	4.08	4.82	53.88	2.19	21.90	124.31
40.0 T. Delay	0.129	10.71	1.07	4.28	5.36	54.99	2.01	18.77	128.93
50.0 T. Delay	0.147	11.63	1.05	4.46	6.11	55.53	1.85	15.91	131.18
60.0 T. Delay	0.151	12.39	1.05	4.58	6.76	55.98	1.73	13.96	133.06
70.0 T. Delay	0.149	12.97	1.04	4.78	7.16	56.45	1.54	11.87	135.01
100.0 T. Delay	0.172	15.46	1.04	5.20	9.22	56.99	1.12	7.24	137.26

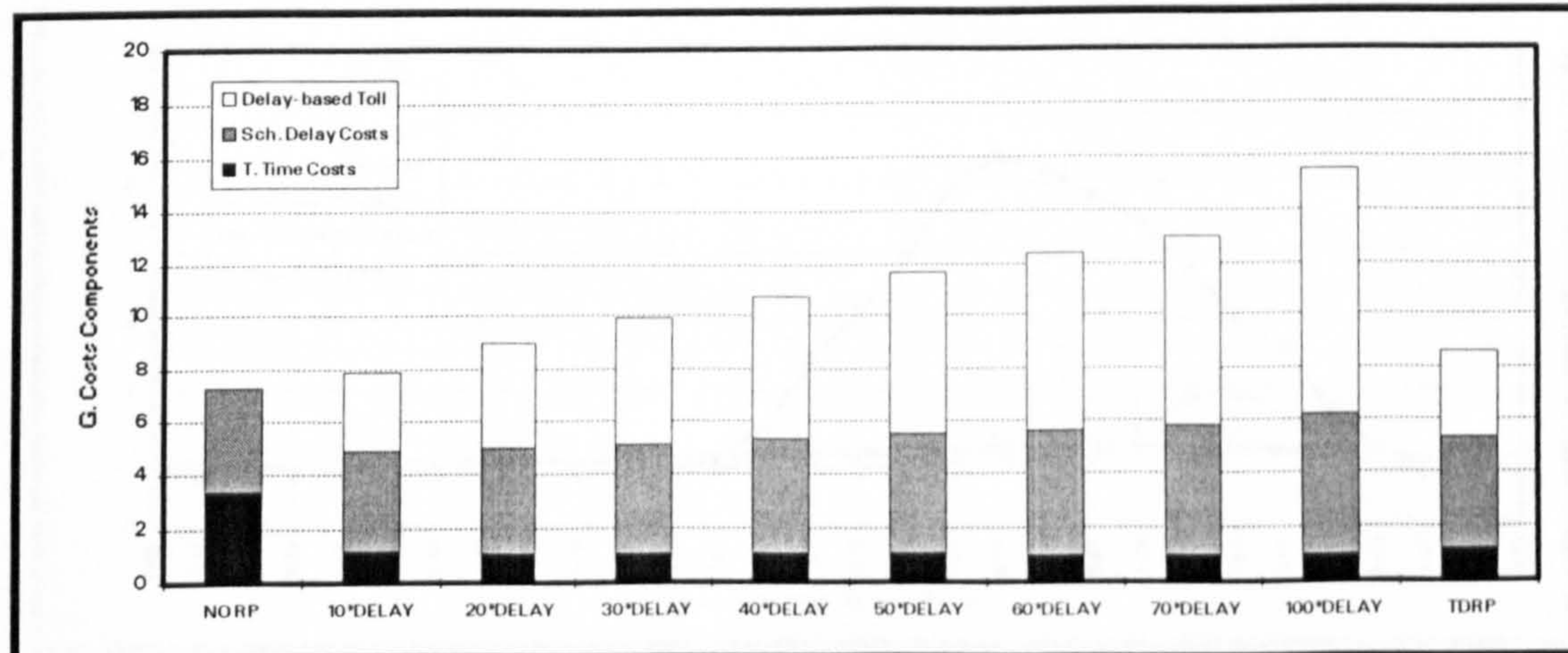


Figure (8.12): The average generalized cost components for all movements under delay-based charging.

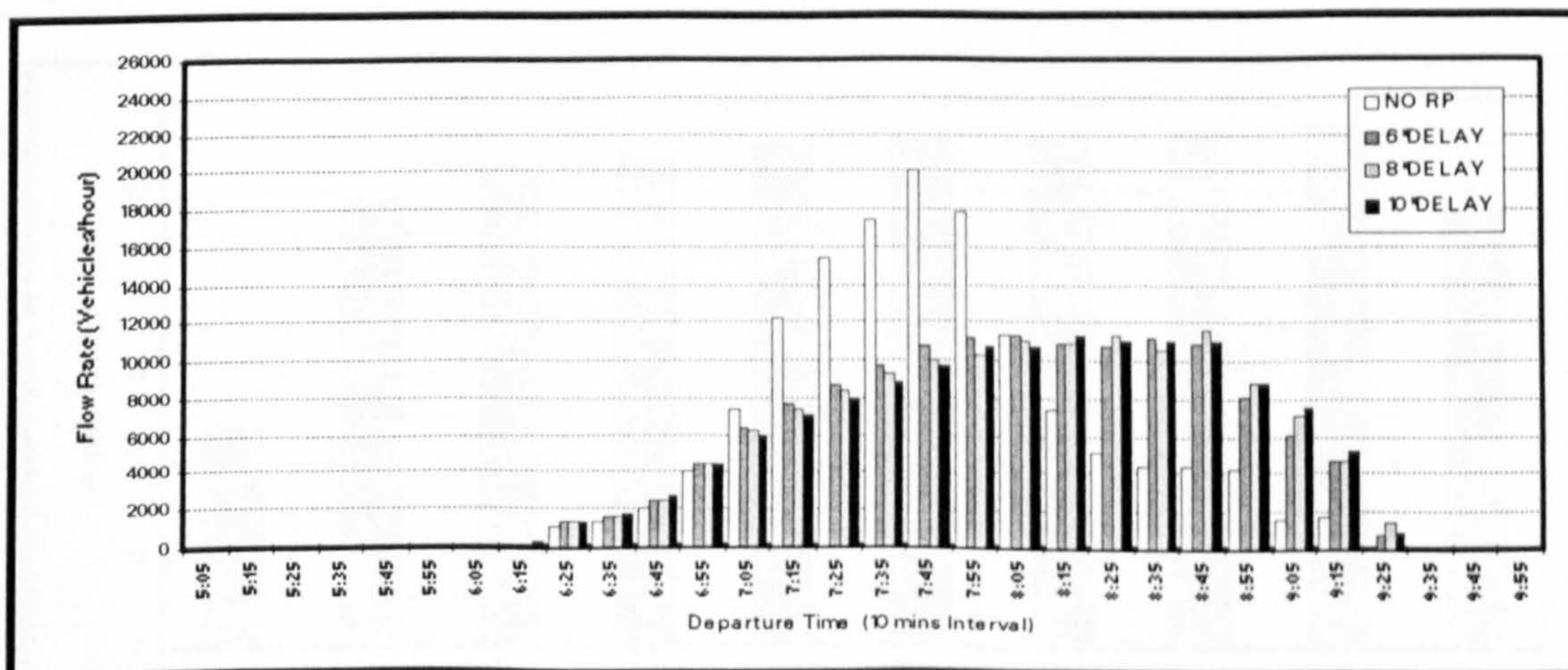


Figure (8.13): The departure time distribution for all movements under delay-based charging (for a multiple of travel delay equal to 6.0, 8.0 and 10.0).

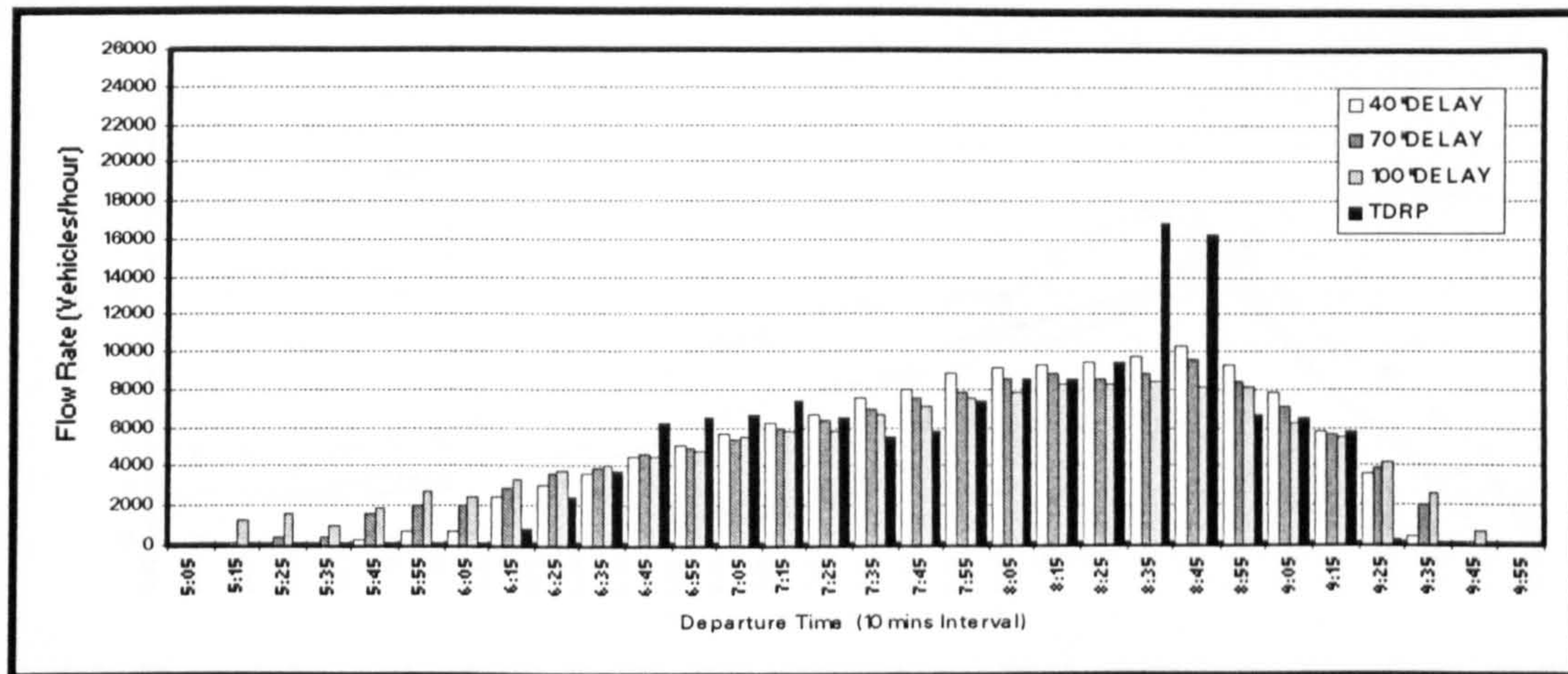


Figure (8.14): The departure time distribution for all movements under delay-based charging (for a multiple of travel delay equal to 40.0, 70.0 and 100.0).

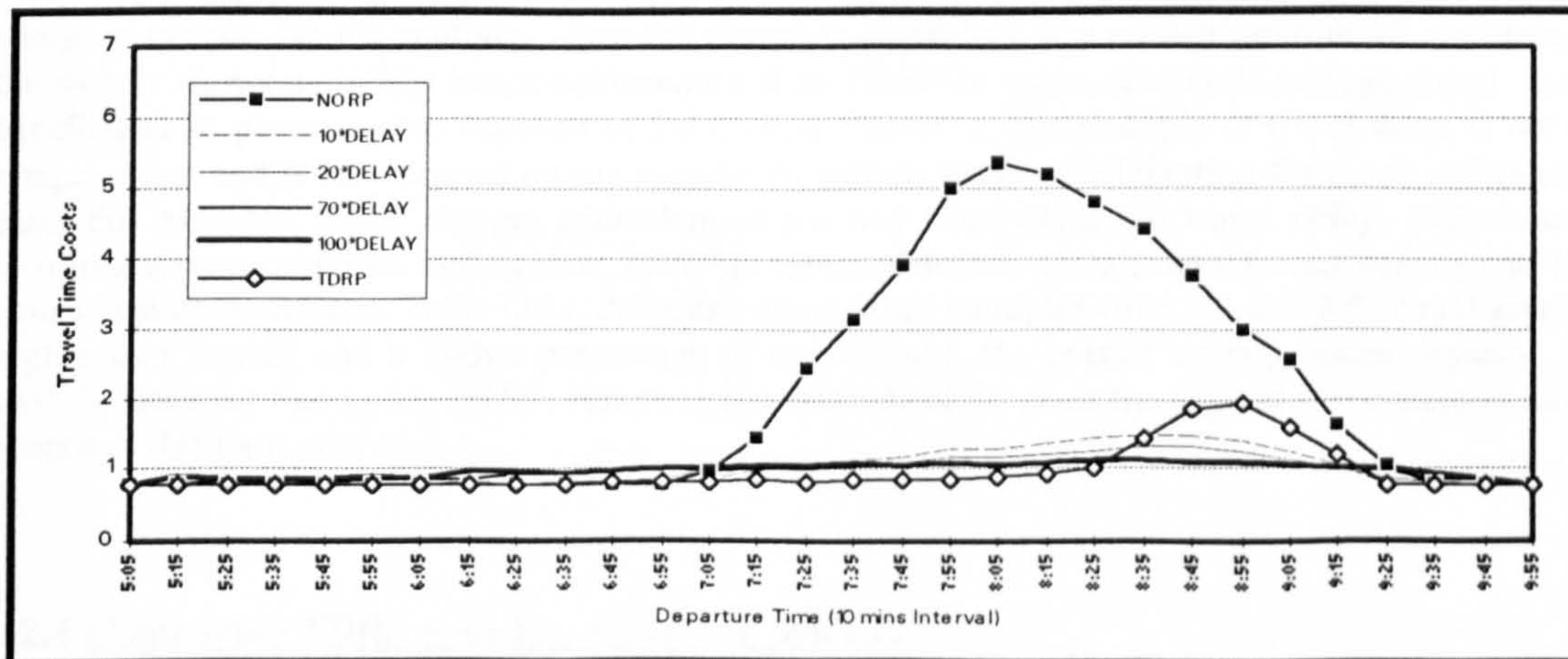


Figure (8.15): Travel time costs for all movements under delay-based charging (base-demand).

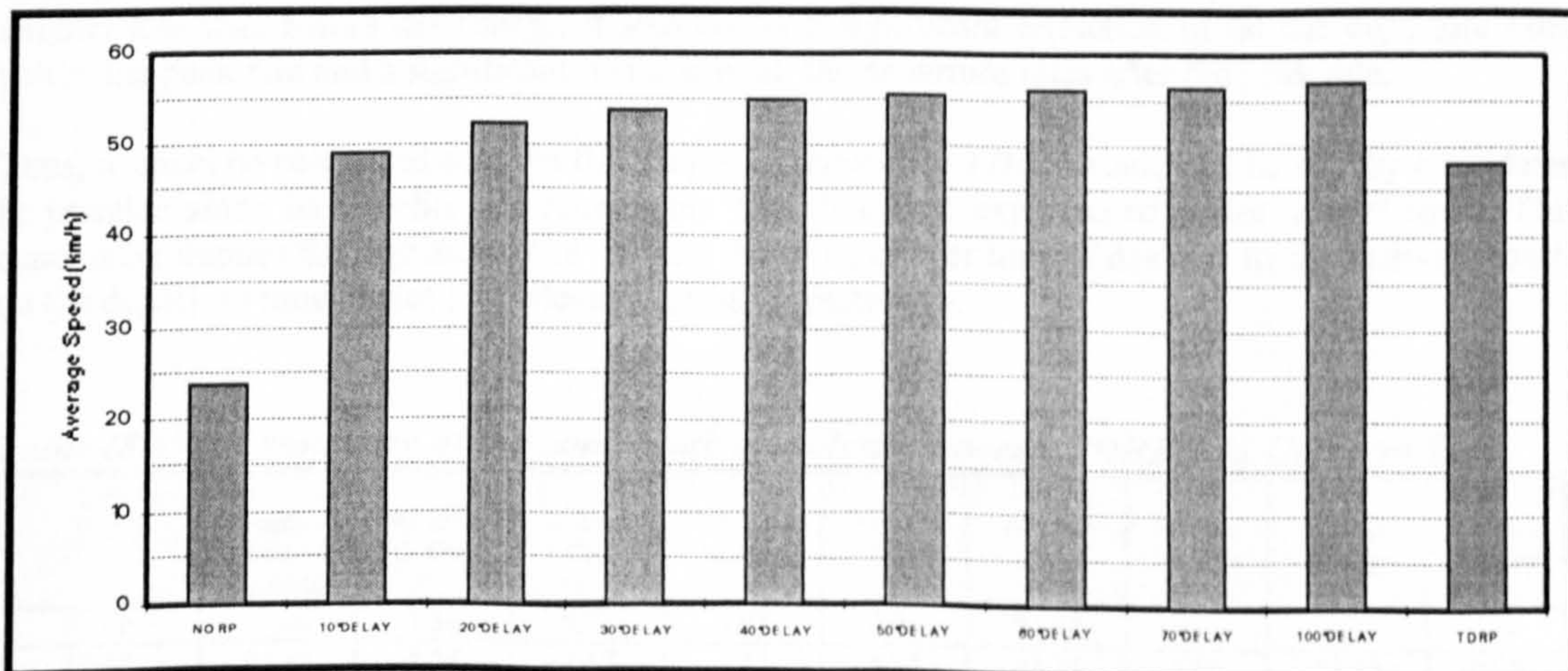


Figure (8.16): The overall average speed under delay-based charging.

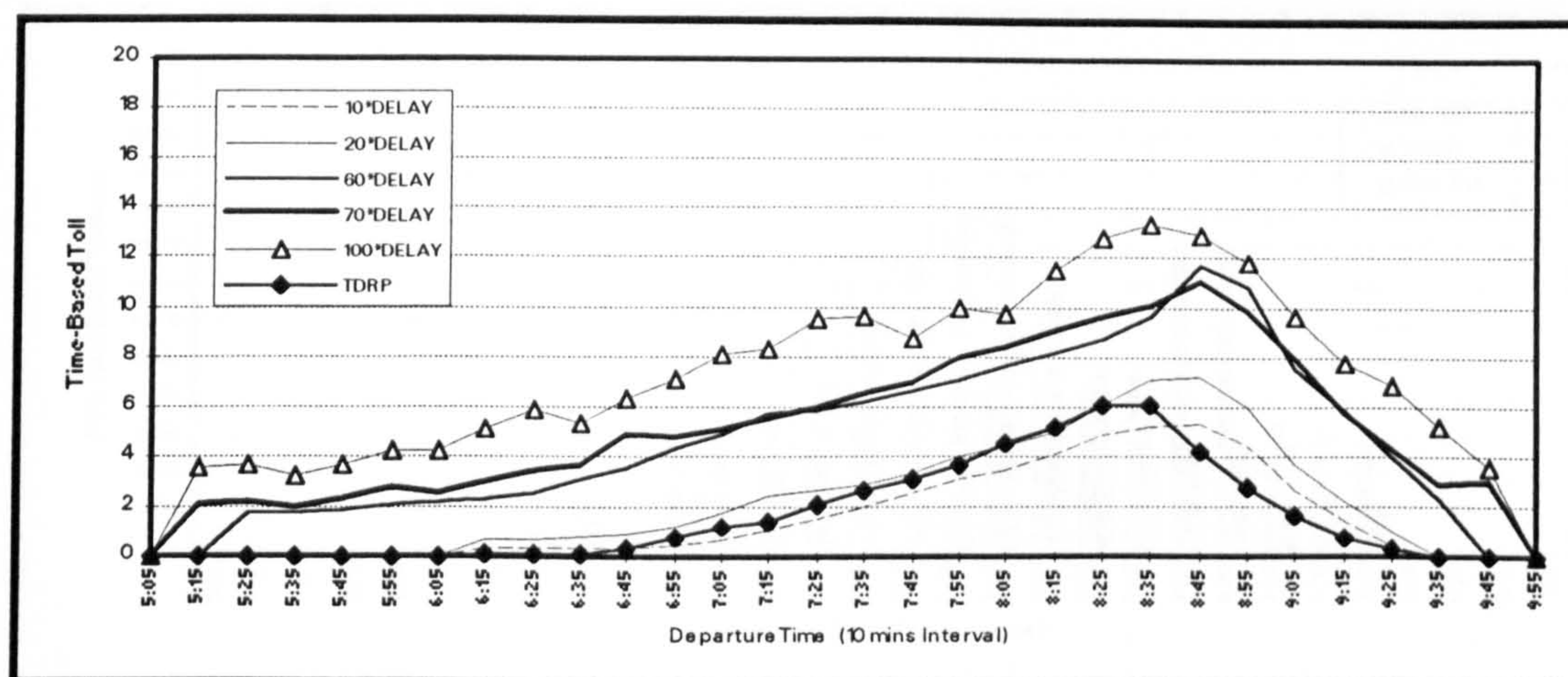


Figure (8.17): The distribution of the average toll for all movements under delay-based charging.

Overall, it appears from the above results that as long as the delay-based charge is equivalent to the delay experienced by the charged vehicles, *TDRP* shows a better performance than delay-based charging system. On the contrary, when the charge is equivalent to a certain multiple of that delay, the system shows a slightly better performance than *TDRP* in terms of overall average speed, user benefit and its percentage as depicted in Table (8.3). However, this multiple of travel delay is not a unique value and it may depend on the specific conditions under consideration for every individual case. For example, under charges equivalent to ten and twenty-fold the travel delay, delay-based charging system performs well against *TDRP* in terms of overall average speed, user benefit and its percentage as depicted in Table (8.3). Although some other multiples (like 6.0 and 8.0) could give a higher user benefit and a higher percentage of user benefit, the overall average speed depicted is slightly less than that under *TDRP*. However, the multiple of 10 gives the highest percentage of user benefit in this particular case.

8.2.4 Comparing *TDRP* with Flat Charging System.

The comparison between *TDRP* and the uniform (or flat) toll is summarised in Table (8.4) below. The table shows that although the uniform toll leads to some user benefits, these benefits are very much less than that obtained under *TDRP*. The reason for this little influence on the user benefit is because the uniform toll has a very little effect on the departure rate distribution as depicted in Figure (8.18). The figure shows a very little shift for the peak departure rate towards late departures and it becomes slightly less than before the charge. It also shows a significant reduction in all the departure rates before the peak rate and a significant increase in all the departure rates after the peak rate.

Thus, it could be concluded that the flat charge equivalent to *TDRP* would not be worthy considered in practice since its benefits are remarkably less than that experienced under *TDRP* itself. This conclusion implies the significance of varying the charges over time of day and its distinctive impacts on the departure time choice and alleviating traffic congestion.

Table (8.4): A summary of the comparative analysis between *TDRP* and Uniform Toll.

Tolls	Weighted Coeff. of Variation	Weighted G. Cost	T. Time Cost	Schedule D. Cost	Average Toll	Average Speed	Av. User Benefits	% User Benefits	% Speed Increase
No RP	0.160	7.36	3.39	3.97	0.00	24.02	----	----	----
<i>TDRP</i>	0.082	8.54	1.17	4.13	3.25	48.33	2.06	24.12	101.21
Uniform Toll	0.127	9.38	2.77	3.83	2.78	28.25	0.76	8.10	17.61

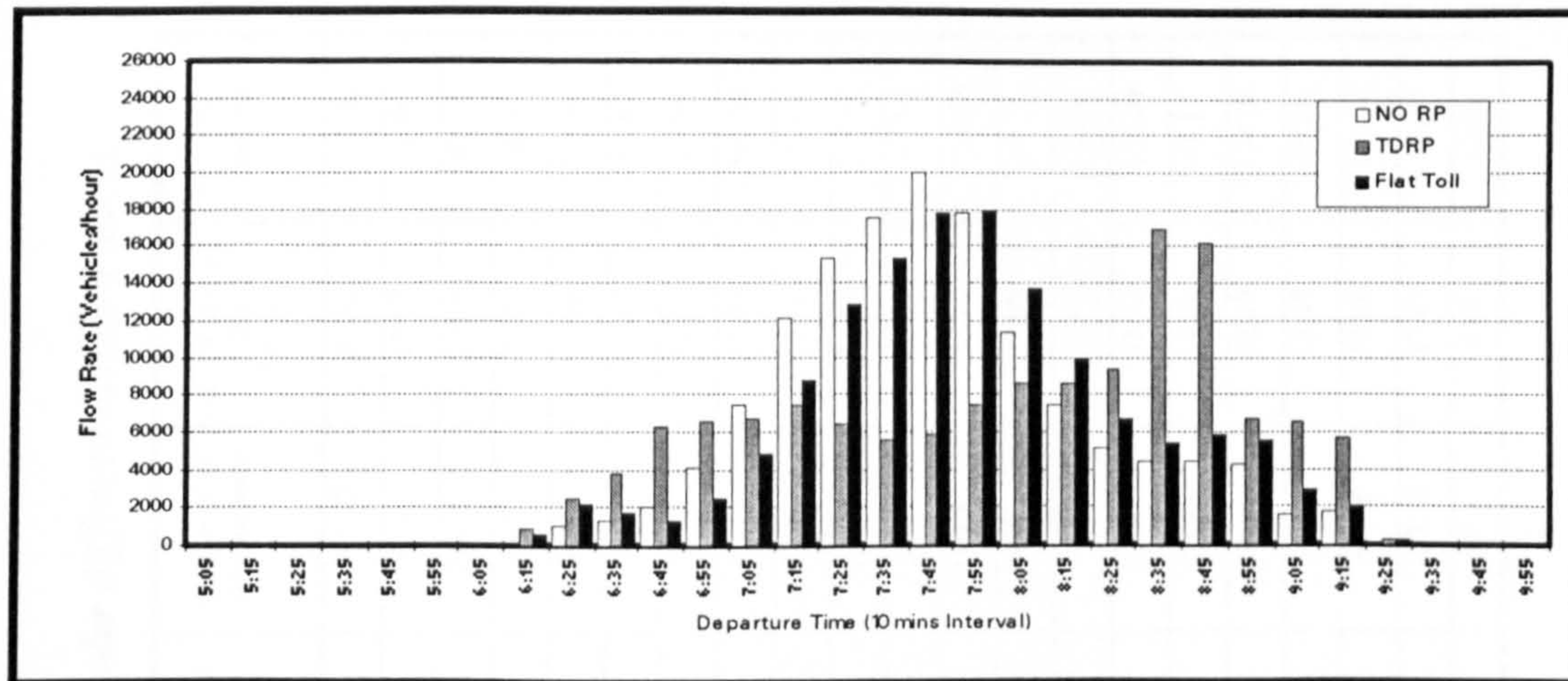


Figure (8.18): The departure time distribution for all movements under the flat toll and TDRP.

8.2.5 TDRP vs. alternative charging systems under different demand levels.

In this section the comparative analysis is conducted between *TDRP* and the four charging systems under different demand levels: base demand, 1.2 and 1.4 base demand. The results are summarised in Table (8.5) and presented in Figures (8.19) to (8.24). In addition to the multiple of unity, the multiples of 8 and 10 for time-based and delay-based charging systems respectively are also considered in examining the results under the different demand levels.

Under the three demand levels, the tenfold delay-based charge is judged to perform well against others in terms of the percentage of speed increase, the percentage of time and schedule delay savings and the total user benefit as shown in Figures (8.19) to (8.22). Also *TDRP* and eight-fold time based charge perform well and achieve benefits very close to that depicted under the tenfold delay-based charge. Figure (8.21) exhibits that the range of the percentage of schedule delay change is [-10%, 10%] and while *TDRP* exhibits the highest increase in schedule delay, tenfold delay-based charges exhibit the highest schedule delay saving under the three demand levels and the base-demand in particular.

The main drawback of the eight-fold time-based charge is the higher value of charge incurred under this system as depicted in Figure (8.23). Figures (8.22) and (8.23) also show that although the charge under tenfold delay-based charge is very much less than that incurred under *TDRP* for 1.4 base-demand, the total user benefit under the former is much higher.

In terms of reduction in the total queue duration, *TDRP* shows the highest percentage of reduction in queue duration under base-demand and 1.2 base-demand, while the tenfold delay-based charge shows the highest percentage of reduction in queue duration under 1.4 base-demand as depicted in Figure (8.24).

To summarise, the comparison between *TDRP* and other charging systems demonstrates that on the one hand, *TDRP* is a superior charging system compared with other charging systems since it leads to distinctive user benefit, speed increase and reduction in queue duration under a reasonable amount of charge. On the other hand, if the time-based charge and delay-based charge are multiplied by a certain factor (greater than one), they could lead to a comparable or better user benefit than *TDRP* regardless of the value of charge incurred (However, increasing the value of charge will consequently increase the total travel cost and then lead to both modal split and suppression of some journeys from travelling neither of which is considered in the scope of this study). Thus, the *TDRP* function employed for a network throughout this study does not lead to system optimization since it does not eliminate queues completely and since some other charging systems could lead to a better user benefit.

Table (8.5): A summary of the comparative analysis between TDRP and different charging systems under different demand levels.

Demand Level	Charging System	Weighted Coeff. of Variation	Weighted G. Cost	Average Toll	% of Total Travel Time Saving	% of Total Schedule Delay Saving	Relative % of User Benefit	Relative % of Total Revenue	% of Reduction in Queue Duration	% of Speed Increase
Base-Demand	TDRP	0.082	8.54	3.25	66.02	-3.09	100.00	100.00	65.62	101.21
	Time-Based Toll	0.137	8.55	2.34	33.40	3.64	60.40	70.19	13.67	26.60
	8*Time-Based Toll	0.121	14.55	9.37	66.22	0.41	106.92	284.85	59.09	87.18
	Delay-Based Toll	0.196	6.85	1.08	43.07	3.89	76.36	33.15	21.91	45.25
	10*Delay-Based Toll	0.136	7.96	3.01	64.00	6.82	115.43	95.09	61.86	104.08
	Flat-Toll	0.127	9.38	2.78	20.65	4.94	42.40	86.51	18.12	17.61
1.2 Base-Demand	Distance-Based Toll	0.178	7.23	0.38	14.35	2.59	27.87	11.93	15.37	13.53
	TDRP	0.075	10.88	4.52	69.42	-8.42	100.00	100.00	62.08	117.36
	Time-Based Toll	0.147	9.37	2.38	44.90	0.53	75.73	51.35	21.00	43.61
	8*Time-Based Toll	0.159	16.30	10.27	68.12	-2.36	109.04	237.89	54.56	107.69
	Delay-Based Toll	0.134	8.61	1.58	42.94	1.52	74.30	33.56	18.09	45.19
	10*Delay-Based Toll	0.214	10.04	4.19	66.41	2.79	115.73	104.85	55.30	124.86
1.4 Base-Demand	Flat-Toll	0.135	11.98	4.31	28.54	-0.48	46.63	96.72	23.98	31.25
	Distance-Based Toll	0.192	9.03	0.33	5.30	-1.62	5.83	7.35	5.37	2.60
	TDRP	0.074	13.24	5.84	71.34	-6.62	100.00	100.00	55.12	136.59
	Time-Based Toll	0.147	11.45	3.01	42.23	1.09	67.60	50.75	14.02	43.10
	8*Time-Based Toll	0.163	17.85	10.96	72.54	0.02	113.00	193.14	52.10	134.34
	Delay-Based Toll	0.129	10.43	2.04	46.17	1.15	73.83	32.01	15.47	49.05
1.4 Base-Demand	10*Delay-Based Toll	0.219	11.98	5.22	73.85	3.72	121.25	69.02	59.88	151.96
	Flat-Toll	0.157	14.79	5.96	34.23	-0.06	53.21	103.29	27.89	50.00
	Distance-Based Toll	0.211	11.03	0.29	-0.34	-0.79	-1.85	5.01	4.80	-2.69

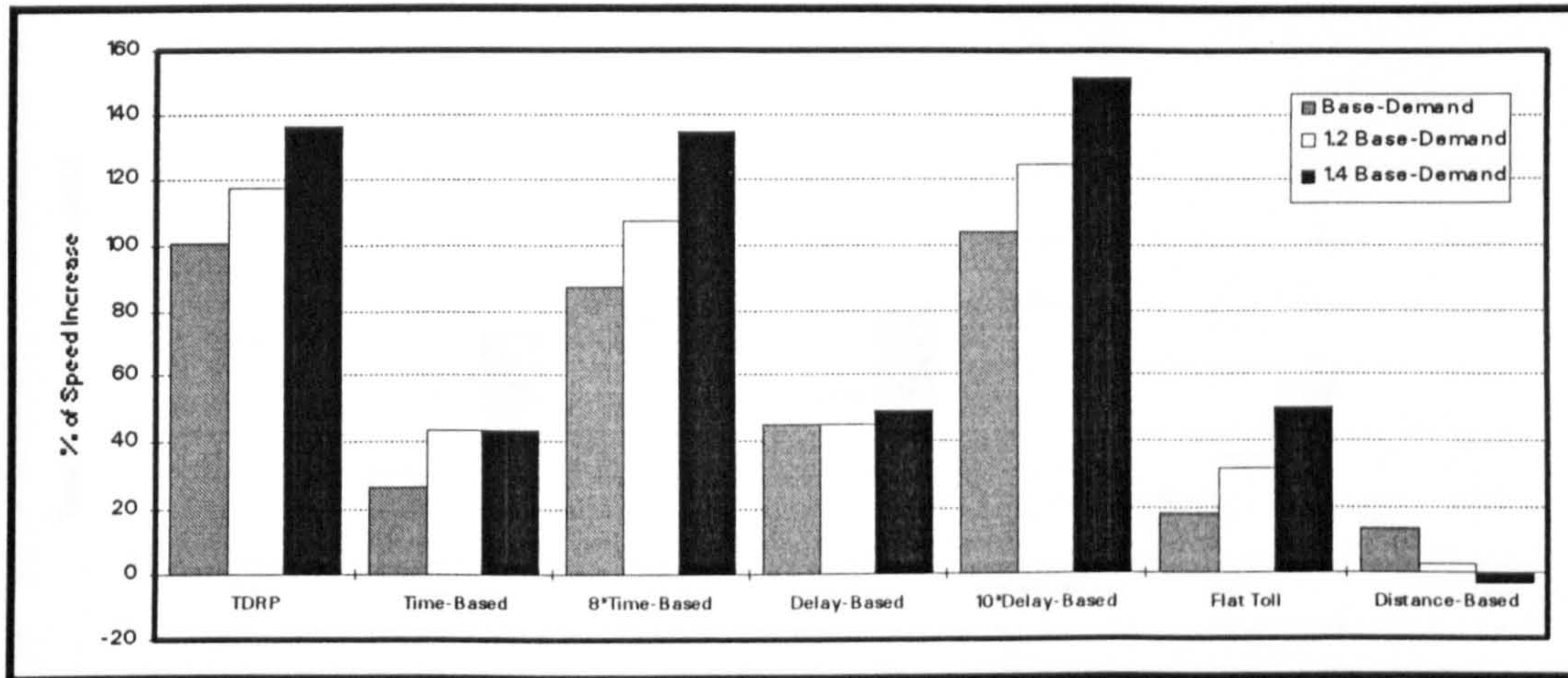


Figure (8.19): The percentage of speed increase under different charging systems and different demand levels.

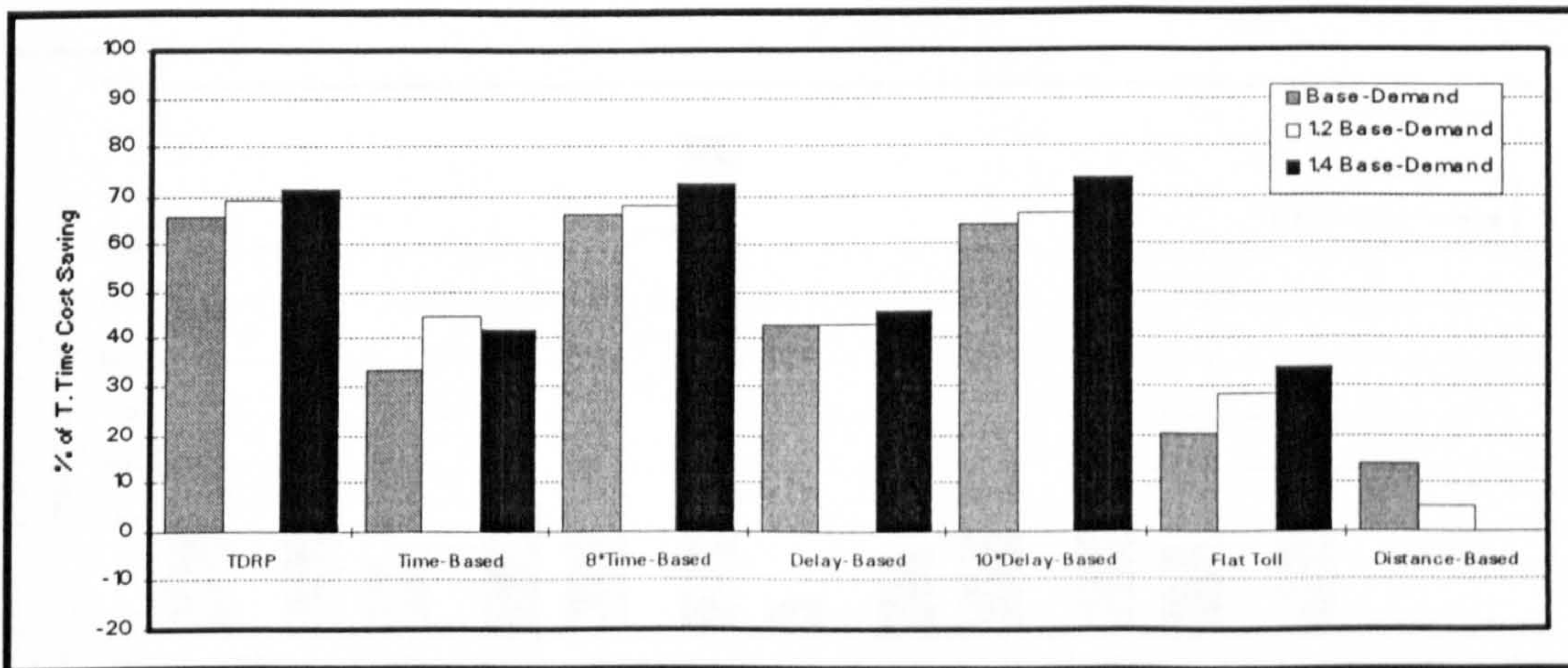


Figure (8.20): The percentage of the total travel time saving under different charging systems and different demand levels.

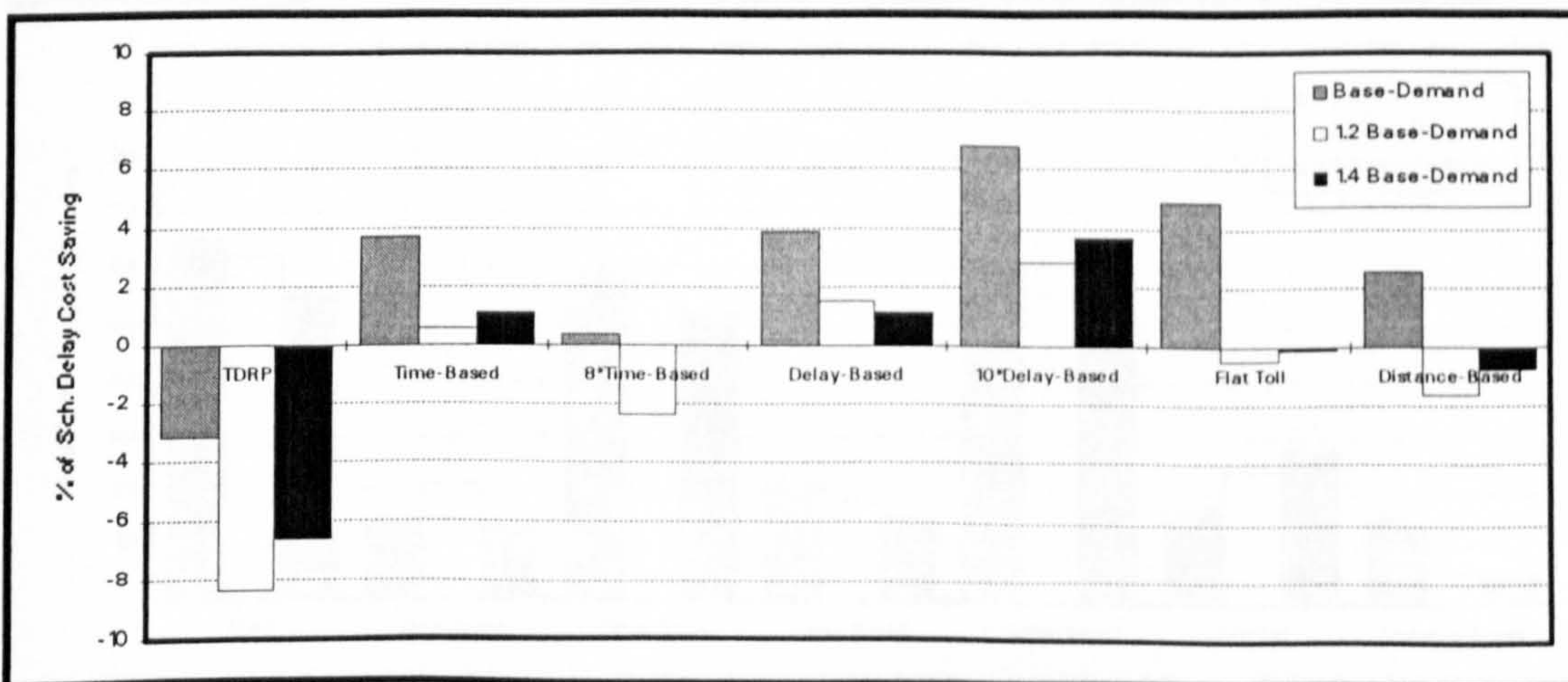


Figure (8.21): The percentage of the schedule delay cost saving under different charging systems and different demand levels.

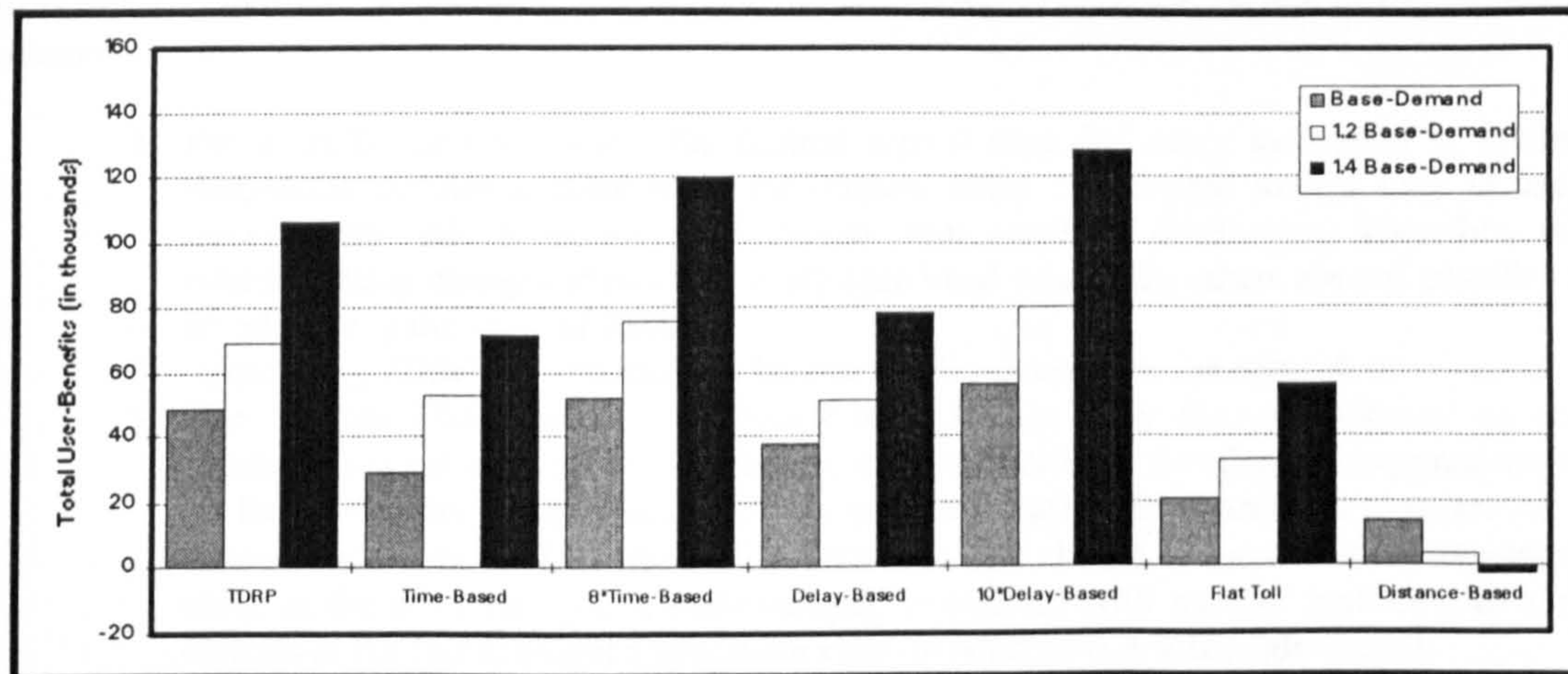


Figure (8.22): The total user-benefits under different charging systems and different demand levels.

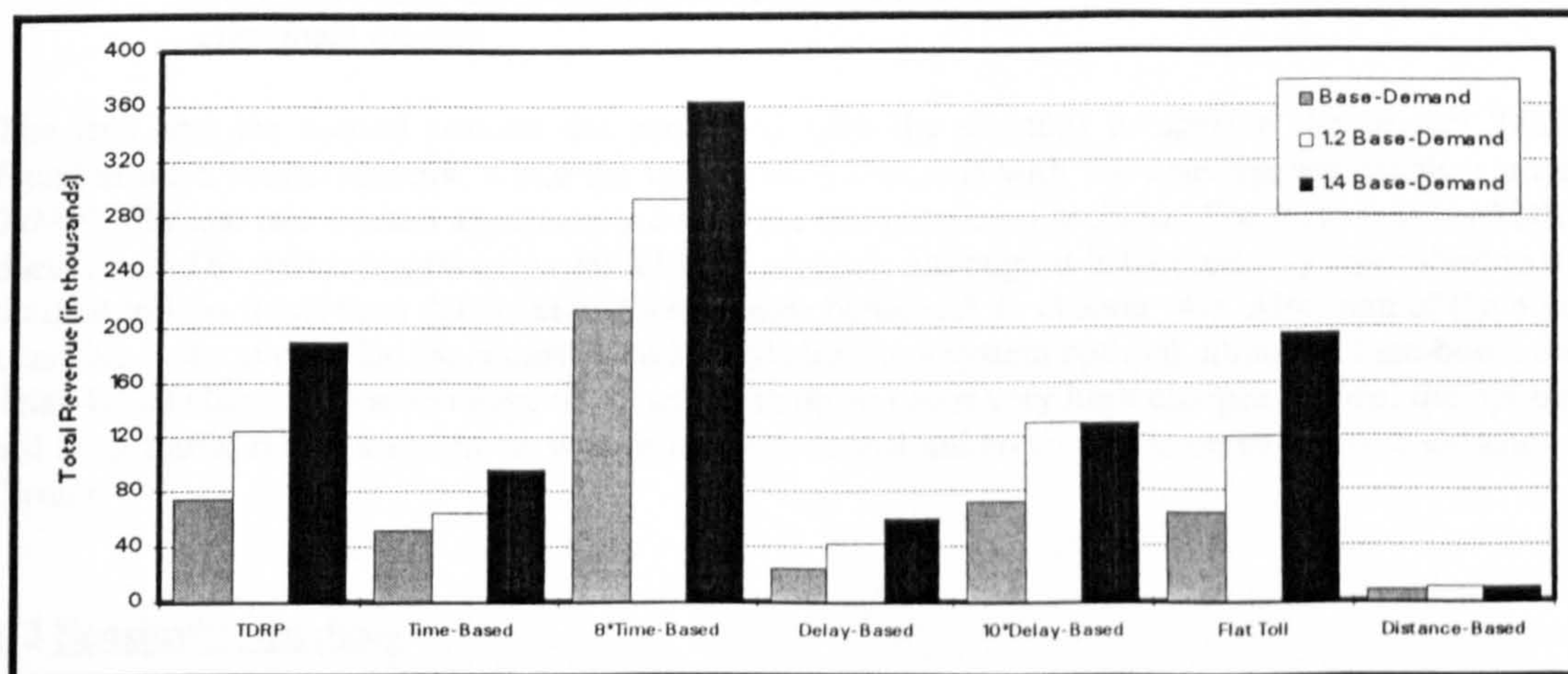


Figure (8.23): The total revenue under different charging systems and different demand levels.

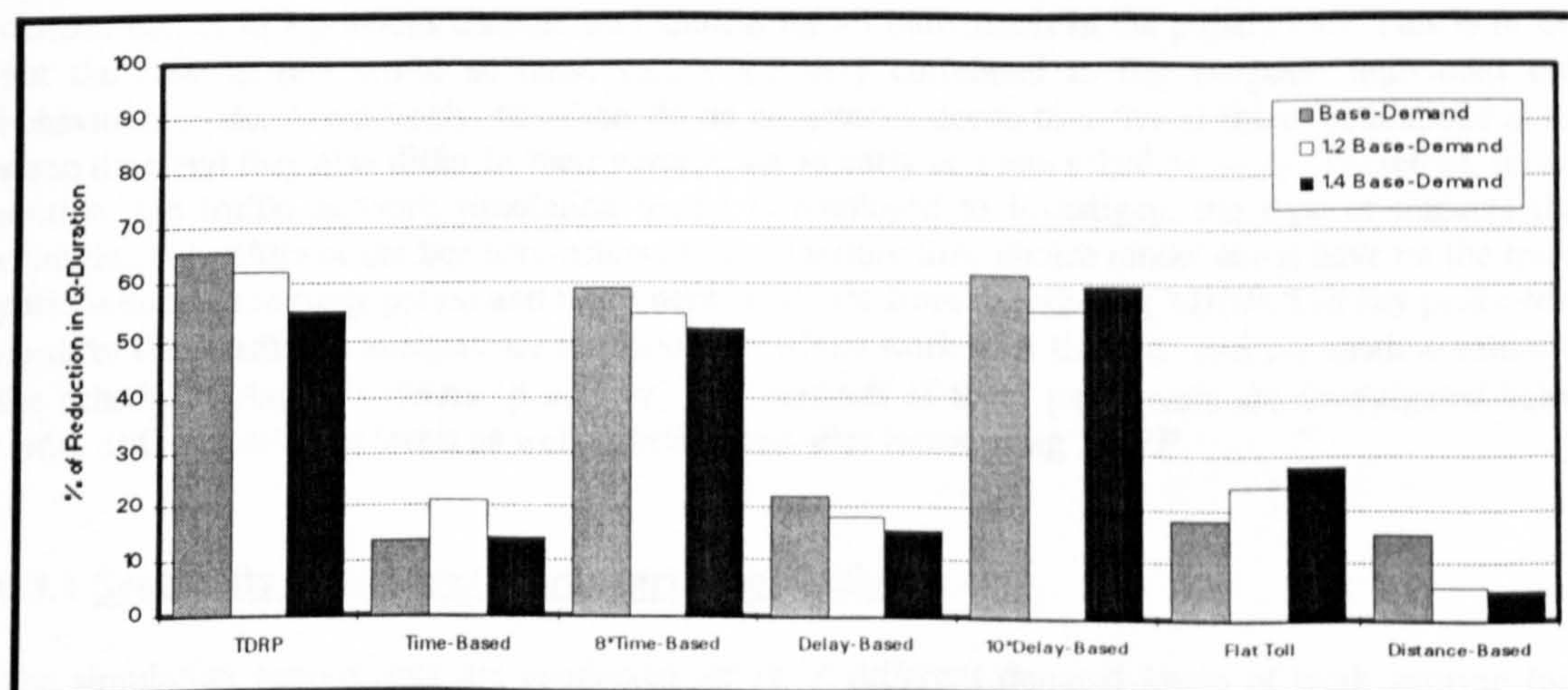


Figure (8.24): The percentage of reduction in queue-duration under different charging systems and different demand levels.

However, the shortfall of this function to achieve system optimization could be due to the following reasons:—

1. For a traffic network, while the desired arrival time for every individual at his/her destination is known, there is no information about the desired arrival time at each intermediate node along the route chosen from origin to destination. Therefore, the schedule delay changes imposed by every individual vehicle on others are not considered in calculating the value of *TDRP*.
2. In charging *TDRP*, the interactions between nodes throughout the network are ignored.
3. The original *TDRP* function, excluding the schedule delay changes imposed on one another, does not lead to convergence (i.e. user equilibrium), therefore it is approximated by the day-to-day adjustment process as explained earlier in a previous chapter. As a consequence, instead of having a toll pattern with zero charge at one end and a maximum value at the other end (i.e. a wedge-shape), it ends up with another pattern with zero charges at the two ends and a maximum value in between (i.e. a triangle-shape).
4. The starting and vanishing times of queues along links are based on some assumptions which may not be quite accurate and need to be investigated further.
5. The values of *TDRP* throughout the simulation are calculated over every individual link every 10-second time interval. However, for simplicity and in order to reduce the computer storage capacity, the average charges over every 60-second time interval for every individual link from the previous iteration/day are used to carry out the day-to-day adjustment process.

The first and the second reasons are concerned with the shortfall to achieve the perfect *TDRP* function for a traffic network, while the third reason concerns with the convergence problem under *TDRP*. The last two reasons are concerned with the computational problem. For these reasons *TDRP* does not lead to system optimization for a traffic network, although it is theoretically equivalent to the optimal toll (as it has been demonstrated for a single bottleneck in chapter (4)). Also, none of the other charging systems used for the comparative analysis lead to a system optimal, although time-based and delay-based charging systems have the potential to do so under very high charges. Hence, the optimal toll for a traffic network system as well as its users is still unknown and is perhaps worth expanding further.

8.3 Sensitivity Analysis

In modelling congestion during the morning peak period in this study the variability among individual travellers is only captured by the probabilistic nature of the departure time choice model (i.e. Logit Model) and its parameter ' η '. The model assumes that all individuals are identical in their desired arrival time as well as their values of times for late and early arrivals (i.e. B , Δ , β , and $\beta\gamma$ as defined earlier in a previous chapter are identical for all individuals in the population). This is in fact not the case in real world as these values are very correlated to trip purpose, individual user behaviour, ... etc. Accordingly, travellers do not necessarily desire to arrive at their destinations at the same time and they also differ in their perceptions to early and late schedule delay. Therefore, in this section, the traffic network simulation model is employed to investigate the type of impacts that changing the values of the key parameters of the departure time choice model could have on the travel pattern during the study period and the benefits obtained from introducing *TDRP*. The key parameters used for the sensitivity analysis are the flexibility of the work start time ' Δ ' and the shadow values of the schedule delay parameters ' β and $\beta\gamma$ '. The impacts of these parameters are investigated below under different demand levels as well as before and after introducing *TDRP*.

8.3.1 Sensitivity to different work start time flexibility

The simulation experiments are conducted for three different demand levels of work starting time flexibility ' Δ ': 0, 15 and 30 mins, which are assumed to be the same for all OD pairs. The other model parameters are assumed to take their base case values. For Δ equal to 0 mins, the travellers desired

arrival time is 9:00 a.m.; work start time flexibility equal to 15 mins and 30 mins implies that travellers desire to arrive at their destinations within the periods [8:45 a.m. to 9:15 a.m.] and [8:30 a.m. to 9:30 a.m.] respectively.

It is shown that the major effect of increasing the work start time flexibility is to reduce the level of traffic congestion. Similar results to the one presented here, were also obtained by Ben-Akiva et al. (1986b) for a situation present in a single OD pair network connected with parallel routes, and by Vythoulkas (1991) for a road network. However, the results presented below also compare between these impacts before and after introducing *TDRP*.

To demonstrate the results, the distributions of departure rates for all movements under the base demand for before and after *TDRP* are illustrated in Figure (8.25) and (8.26) respectively (the values presented in these figures represent the average of the last five iterations after 60 iterations in total). As was expected increasing the work start time flexibility ' Δ ' causes a spreading of the departure rate distribution over a longer period and reduces its peak rate, resulting in lower travel time costs and schedule delay costs as depicted in Figures (8.27) and (8.28) respectively. The highest departure rate is generated at the departure time that results in an on-time arrival. This time is denoted by the time period $[B-\Delta, B+\Delta]$, where B is the centre of the desired arrival time, i.e. 9:00 a.m. in this example. The peak occurs since the left and right derivatives of the utility function at this departure time have opposite signs. Juxtaposition of the departure rate and travel time distributions reveals that the highest departure rate occurs just before the travel time attains its maximum.

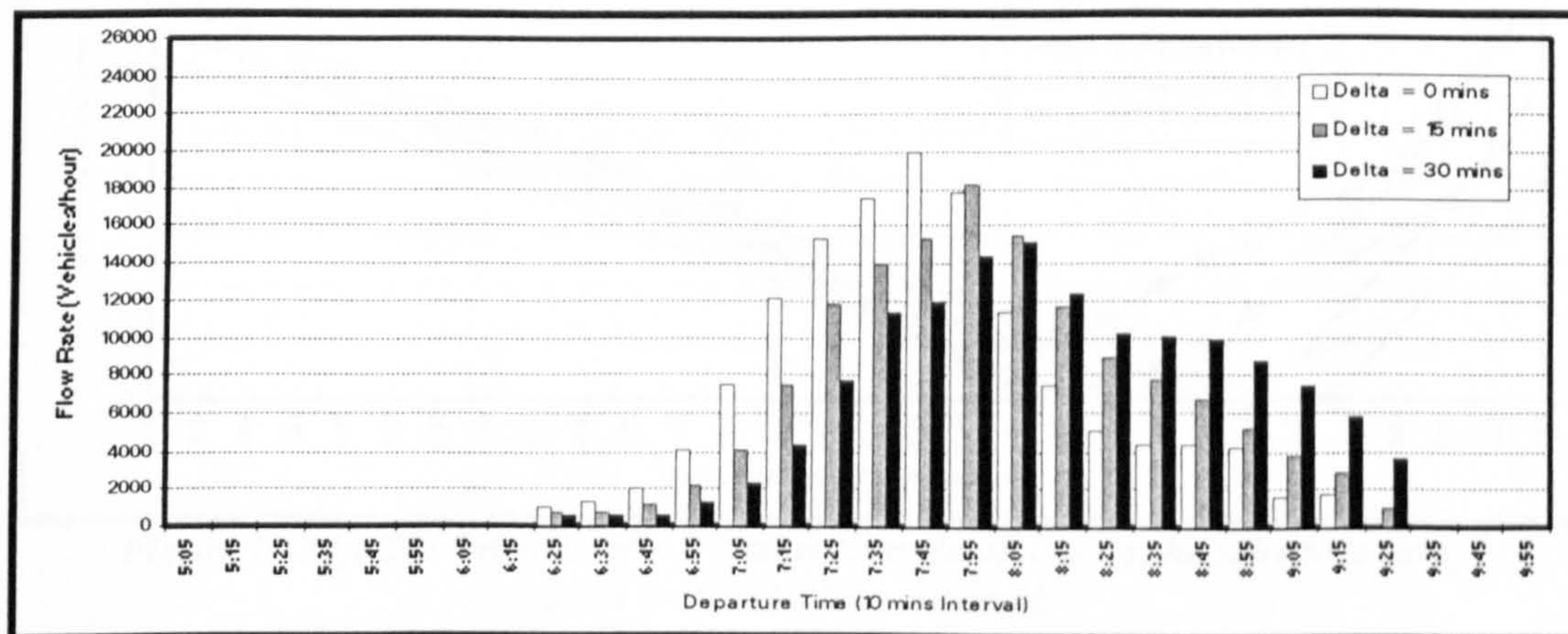


Figure (8.25): The impacts of work start time flexibility on the departure time distribution under the base demand and no road pricing.

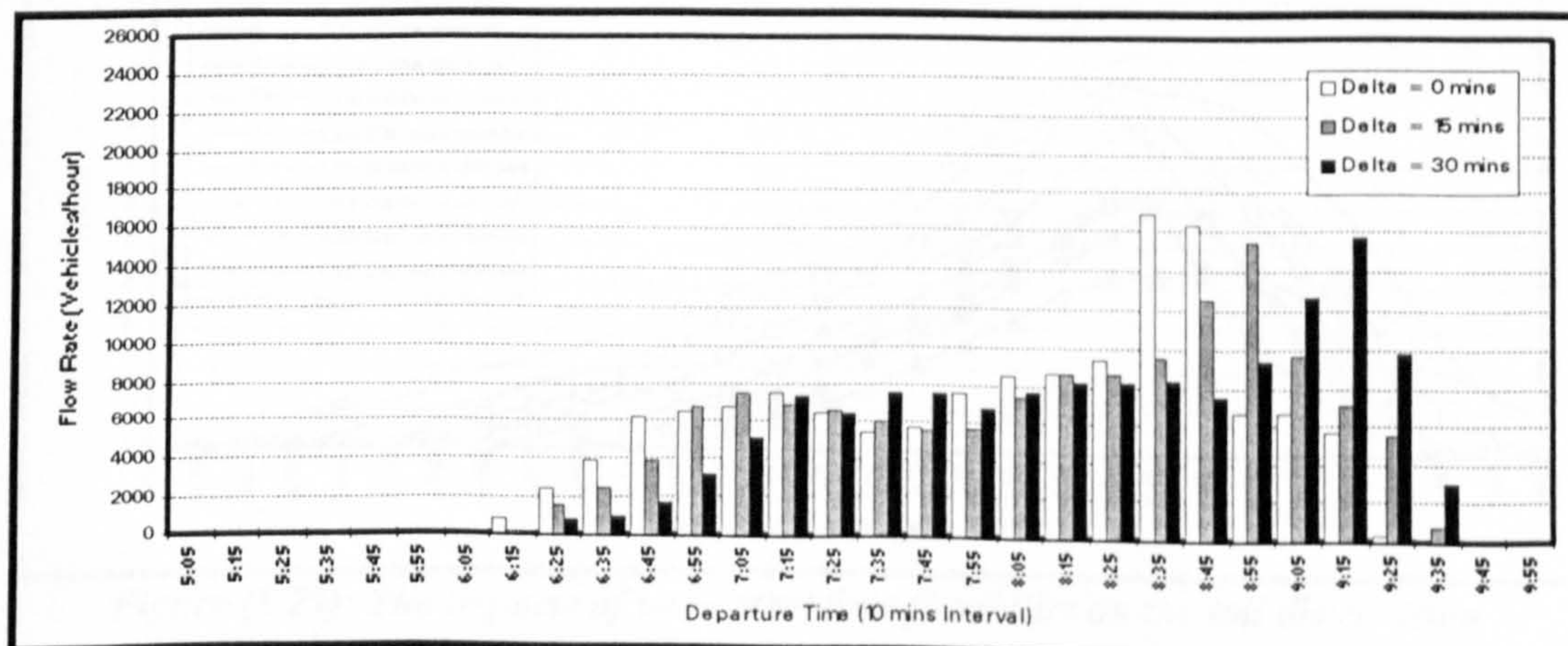


Figure (8.26): The impacts of work start time flexibility on the departure time distribution under the base demand and TDRP.

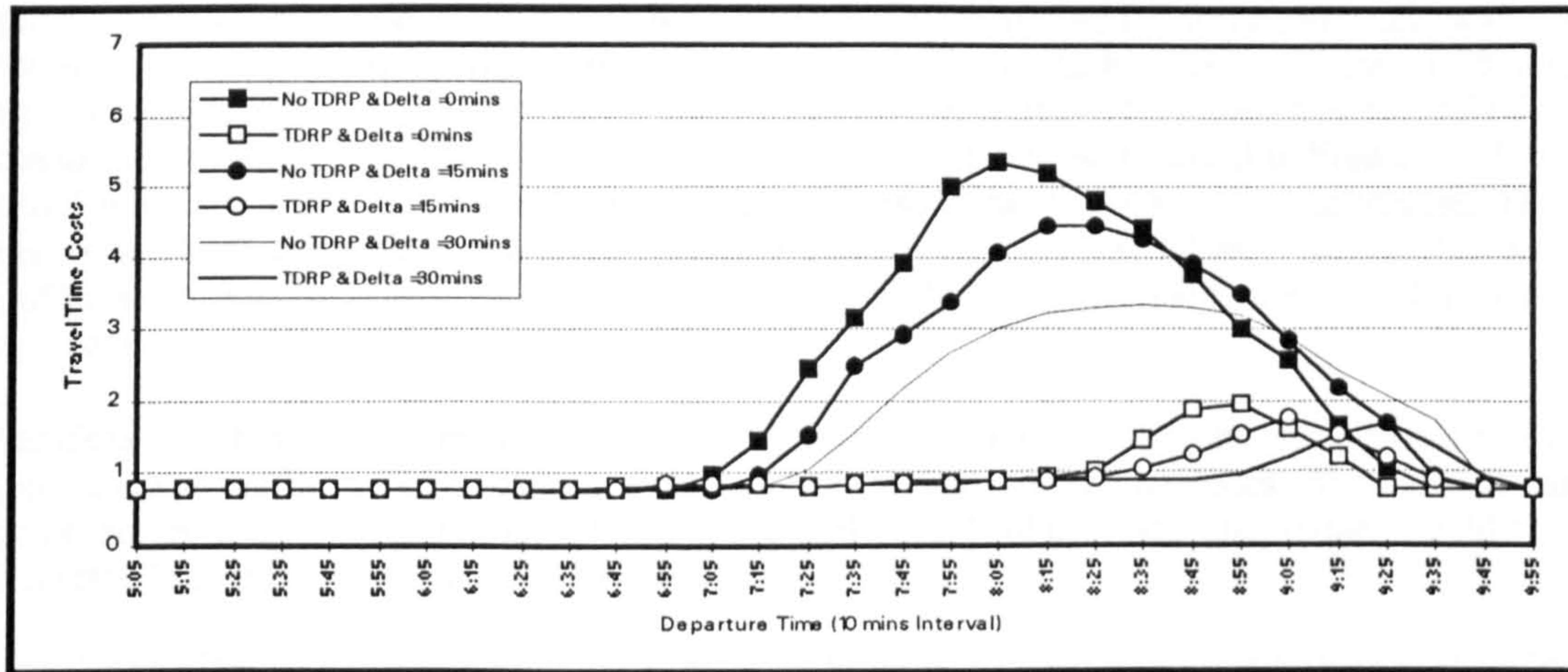


Figure (8.27): The impacts of work start time flexibility on the travel time costs under the base demand.

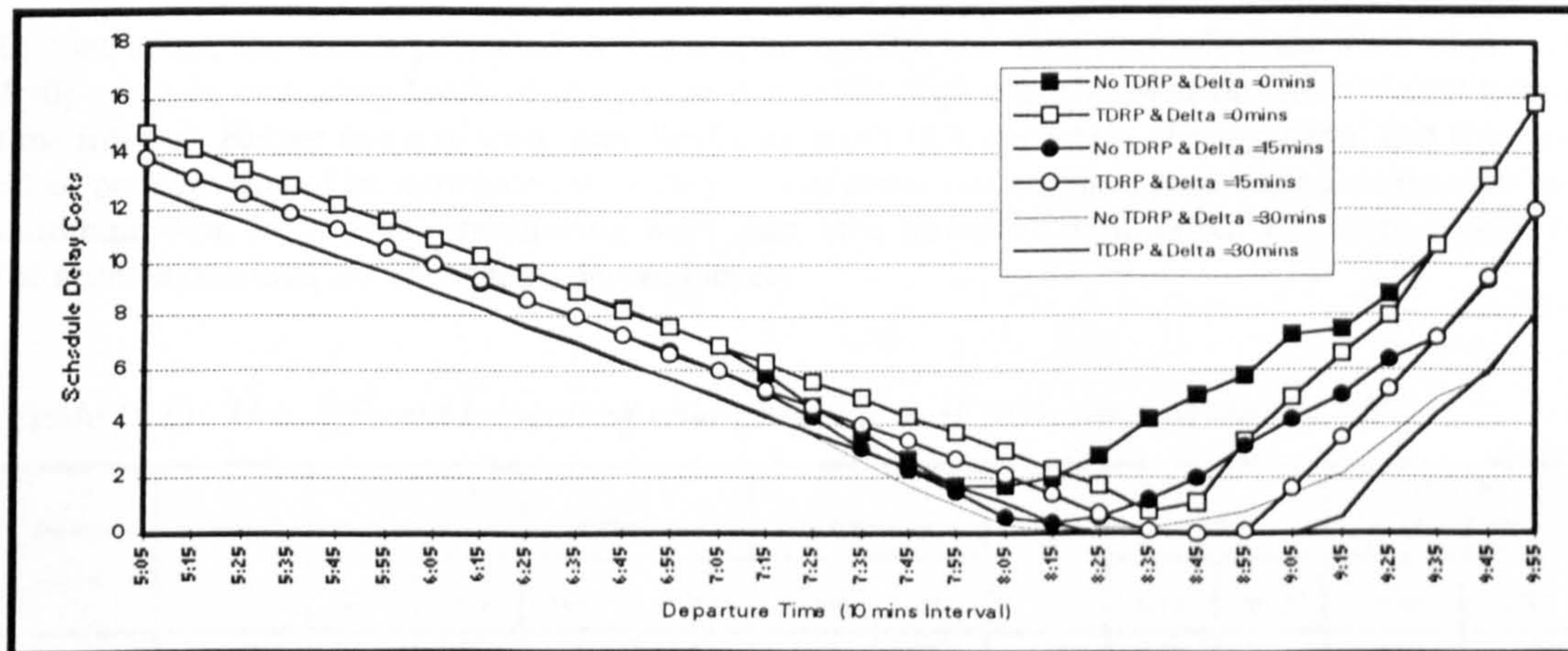


Figure (8.28): The impacts of work start time flexibility on the schedule delay costs under the base demand.

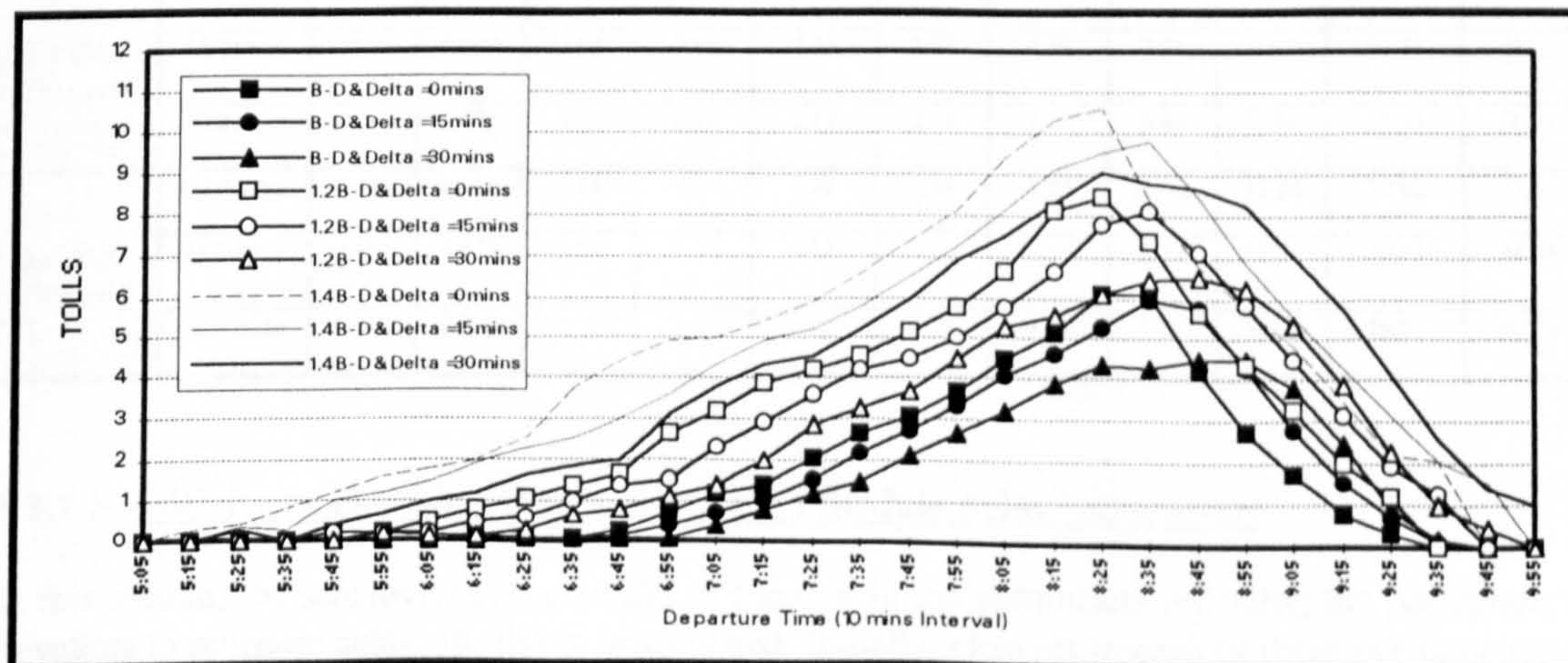


Figure (8.29): The impacts of work start time flexibility on the toll distribution under the base demand.

The departure time distributions in Figure (8.26) also demonstrate the existence of the peak-shifting phenomenon under *TDRP* with different work start time flexibility. As the work start time flexibility increases the peak departure rate is shifted more to the right hand side. Figures (8.27) and (8.28) also show the fact that the benefits from *TDRP* (in terms of savings in travel time costs and schedule delay costs) can be enhanced by increasing the work start time flexibility. As depicted in Figure (8.27), the travel time cost under *TDRP* and $\Delta=30$ mins is very close to the free-flow travel conditions. These results demonstrate the view that no single measure could offer the best solution for tackling urban traffic congestion, and the most promising results are likely to be obtained using a combination of approaches.

Paradoxically, however, increasing the work start time flexibility leads to a poor convergence since the weighted coefficient of variation of the generalized cost 'WCV' increases. These results are demonstrated under different demand levels as depicted in Table (8.6) which summarizes the different impacts of changing the work start time flexibility.

The distributions of the average toll under different levels of demand and work start time flexibility are shown in Figure (8.29). The figure shows that as the work start time flexibility increases, the toll pattern is shifted towards the right hand side and takes a trapezoidal-shape (rather than a triangle-shape) particularly under the base-demand. Also, increasing the work start time flexibility reduces the average toll under different demand levels as depicted in Table (8.6) below.

In conclusion, the results presented in this section confirm the view that inflexible work times (i.e. $\Delta=0$) result in increasing levels of congestion due to the high concentration of traffic within a short time interval. Higher levels of work time flexibility result in lower levels of congestion, and therefore when possible should be introduced since they can alleviate the morning peak period congestion at a minimum cost. In addition, introducing work start time flexibility with *TDRP* results in enhancing the benefits obtained under different demand levels.

Table (8.6): The different impacts of changing the work start time flexibility ' Δ '.

Demand Level	Value of ' Δ '	W. Coeff. of Var. (WCV)		T. Time Cost		S. Delay Cost		TDRP	WGC		Average Speed (km/h)	
		No RP	TDRP	No RP	TDRP	No RP	TDRP		No RP	TDRP	No RP	TDRP
Base Demand	00.0 mins	0.160	0.082	3.39	1.17	3.97	4.13	3.25	7.36	8.54	24.02	48.33
	15.0 mins	0.215	0.094	3.07	1.10	2.57	2.83	3.13	5.64	7.06	25.35	49.49
	30.0 mins	0.267	0.127	2.54	1.05	1.47	1.75	2.63	4.01	5.43	28.88	51.12
1.2 Base Demand	00.0 mins	0.137	0.075	4.23	1.31	4.66	5.04	4.52	8.89	10.88	20.80	45.21
	15.0 mins	0.189	0.084	3.96	1.14	3.35	3.63	4.29	7.31	9.06	21.41	48.49
	30.0 mins	0.282	0.106	3.29	1.07	2.17	2.58	3.64	5.46	7.29	24.33	50.71
1.4 Base Demand	00.0 mins	0.145	0.074	5.18	1.51	5.52	5.89	5.84	10.71	13.24	17.82	42.16
	15.0 mins	0.152	0.075	5.05	1.34	4.12	4.44	5.51	9.17	11.28	17.83	44.44
	30.0 mins	0.288	0.096	4.71	1.09	2.96	3.35	5.08	7.67	9.52	19.03	49.59

8.3.2 Sensitivity to changing the values of the schedule delay parameters

In this section, the sensitivity of the results to changes in the parameters reflecting the perception of travellers to schedule delay (β , $\beta\gamma$) is investigated. Initially, changes in each of these two parameters are investigated while the other remains fixed (i.e. equal to the base case value). Then, a combination of changes in both of them are also investigated.

Sensitivity to changes in the value of early schedule delay ' β '.

The sensitivity to changes in the parameter ' β ' which reflect travellers' perception to early schedule delay is investigated by comparing the base case results ($\beta = 3.9$) with β equal to 2.5 and 5.5. These two values are chosen such that the first one is less than the base case while the second value is greater than the base case and less than the shadow value of travel time and delay ' α ' ($\alpha = 6.4$).

Figures (8.30) and (8.31) illustrate the departure rate distributions for the three different values of ' β ' for before and after *TDRP* respectively. For $\beta = 2.5$, the early schedule delay becomes very much cheaper than travel time and delay ($\beta < \alpha$), therefore, travellers tend to depart early and avoid congestion. Thus, the departure rate distributions (before *TDRP*) exhibit lower peaks compared to the base case distribution and are spread over a longer period of time. On the other hand, for $\beta = 5.5$, the early schedule delay becomes very close to the value of travel time and delay, therefore, the trade-off between early departure and travel time results in reducing the early schedule delay and increasing the level of congestion as illustrated in Figure (8.30). Furthermore, the trade-off between *TDRP*, travel time and schedule delay results in a very high peak (with $\beta = 5.5$) as illustrated in Figure (8.31). This consequently increases the travel time cost, schedule delay cost and the generalized cost for an average journey. It also results in a decrease in the overall average speed.

The travel time distributions under the three values of ' β ' are illustrated in Figure (8.32) for the base demand as well as before and after *TDRP*. The figure shows that the travel time cost is very sensitive to changes in the value of ' β ' particularly before introducing *TDRP*, and as the value of ' β ' increases, the travel time cost increases and vice versa.

Similar results are also attained under different demand levels as presented in Table (8.7) below. The table shows that changing the value of ' β ' causes a substantial reduction in the level of convergence (i.e. increasing the WCV) compared to the base case one before introducing *TDRP*. Furthermore, as the level of demand and the value of ' β ' increase the WCV also increases. Conversely, after introducing *TDRP*, the level of convergence becomes more stable and almost insensitive to changes in the value of ' β '.

Table (8.7): The impacts of changing the shadow value of early schedule delay ' β '.

Demand Level	Value of ' β '	W. Coeff. of Var. (WCV)		T. Time Cost		S. Delay Cost		TDRP	WGC		Average Speed (km/h)	
		No RP	TDRP	No RP	TDRP	No RP	TDRP		No RP	TDRP	No RP	TDRP
Base Demand	2.50	0.181	0.112	2.23	1.03	2.71	2.98	1.98	4.95	5.99	32.19	51.75
	3.90	0.160	0.082	3.39	1.17	3.97	4.13	3.25	7.36	8.54	24.02	48.33
	5.50	0.257	0.092	4.79	1.71	5.38	5.33	4.68	10.17	11.72	19.09	39.85
1.2 Base Demand	2.50	0.189	0.098	2.74	1.05	3.28	3.61	2.77	6.02	7.42	28.39	51.59
	3.90	0.137	0.075	4.23	1.31	4.66	5.04	4.52	8.89	10.88	20.80	45.21
	5.50	0.323	0.105	6.09	2.05	6.55	6.49	6.01	12.64	14.55	16.72	36.66
1.4 Base Demand	2.50	0.255	0.080	3.16	1.12	3.87	4.15	3.83	7.03	9.10	26.81	49.89
	3.90	0.145	0.074	5.18	1.51	5.52	5.89	5.84	10.70	13.24	17.82	42.16
	5.50	0.379	0.104	7.28	2.29	7.88	7.64	7.19	15.16	17.12	14.51	35.42

* The shadow value of the late schedule delay ' $\beta\gamma$ ' = 15.20.

In summary, it is demonstrated that changing the value of the early schedule delay ' β ' has different impacts on the results before and after *TDRP*. Consequently, the relative benefits obtained from *TDRP* will be sensitive to changing the value ' β '.

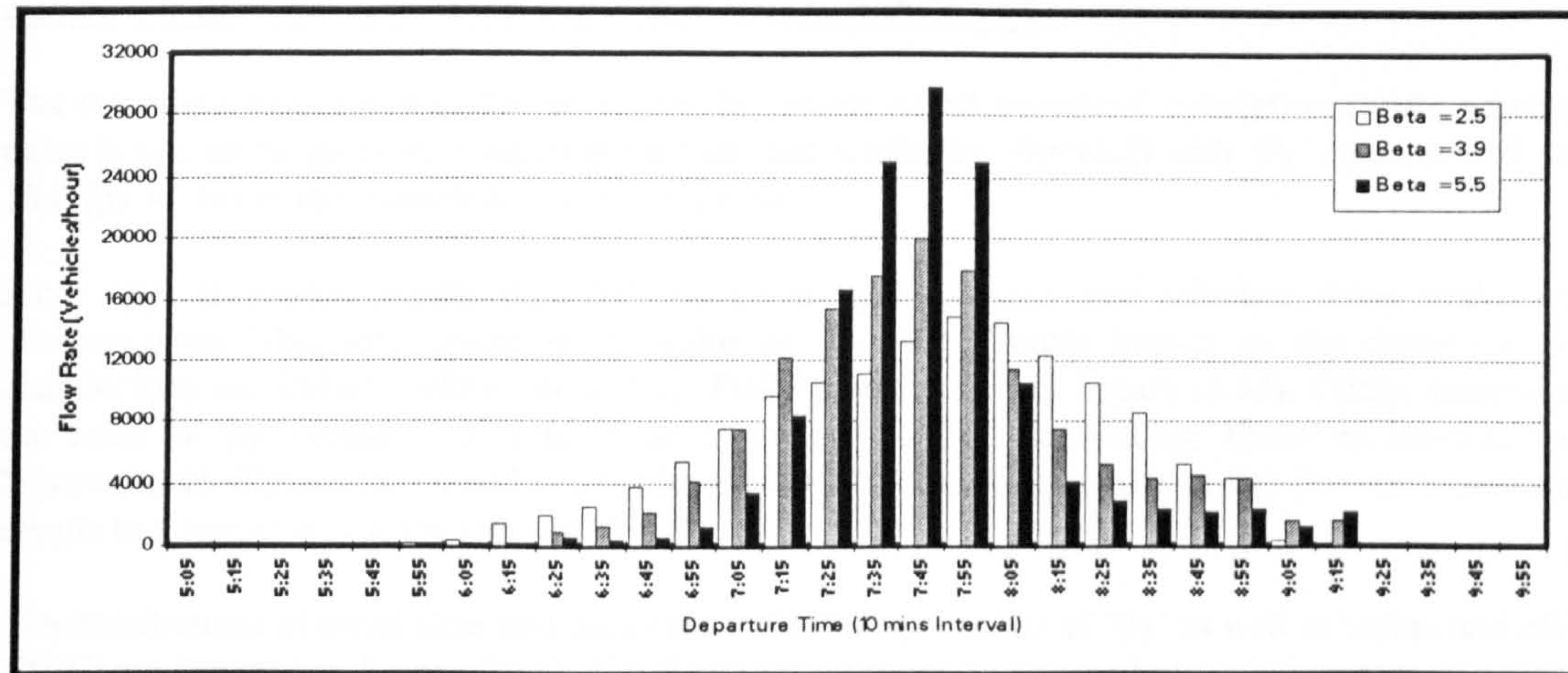


Figure (8.30): The impacts of changing the shadow value of early schedule delay ' β ' on the departure time distribution under the base demand and no RP.

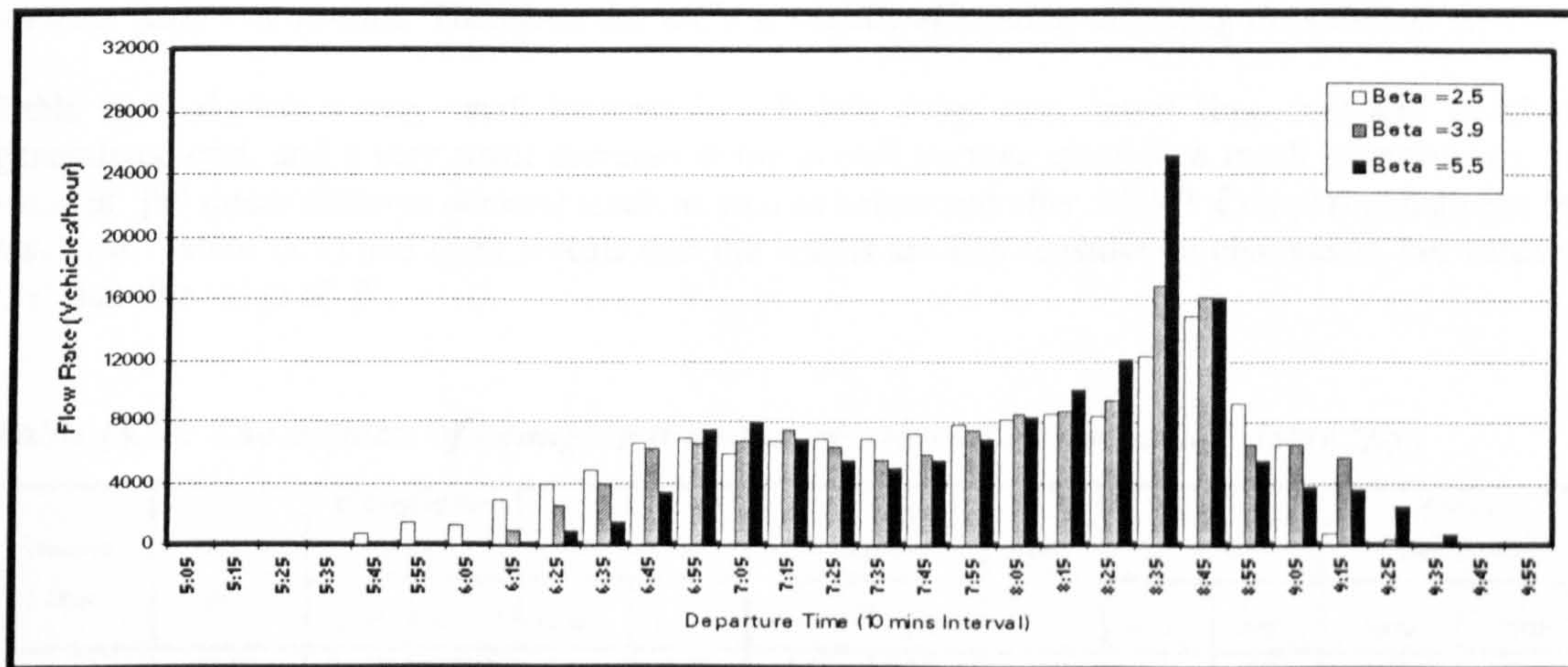


Figure (8.31): The impacts of changing the shadow value of early schedule delay ' β ' on the departure time distribution under the base demand and TDRP.

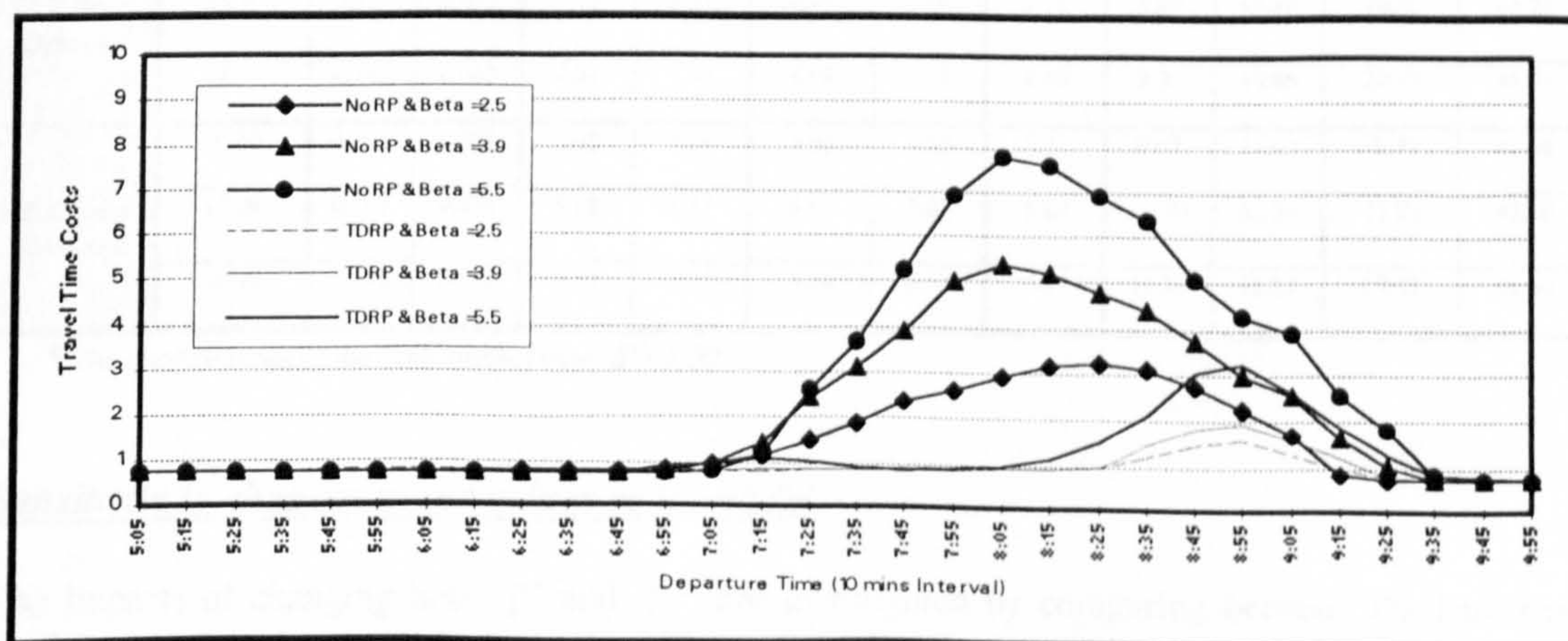


Figure (8.32): The impacts of changing the shadow value of early schedule delay ' β ' on the travel time costs under the base demand (before and after TDRP).

Sensitivity to changes in the value of late schedule delay ' $\beta\gamma$ '.

The sensitivity to changes in the parameter ' $\beta\gamma$ ' which reflect travellers' perception to late schedule delay is also investigated by comparing the base case results (i.e. $\beta\gamma=15.2$) with ' $\beta\gamma$ ' equal to 10.0 and 18.0 ($\beta\gamma$ is always more onerous than α , i.e. $\beta\gamma > \alpha$).

Since ' $\beta\gamma$ ' is always greater than ' α ', travellers tend to avoid late schedule delay under any circumstances. Therefore, changing the value of ' $\beta\gamma$ ' has a small impact on the departure rate distributions particularly before introducing *TDRP* as illustrated in Figure (8.33). Oddly, increasing the value of ' $\beta\gamma$ ' results in a higher peak departure rate after introducing *TDRP* as illustrated in Figure (8.34). Figures (8.33) and (8.34) also exhibit higher departure rates at the departure time that results in a late arrival as the value of ' $\beta\gamma$ ' decreases.

The distributions of travel time cost under the three different values of ' $\beta\gamma$ ' as well as before and after *TDRP* are depicted in Figure (8.35). The figure shows a small increase in travel time cost as a result of increasing ' $\beta\gamma$ '. The different impacts of changes in the value of ' $\beta\gamma$ ' under different demand levels before and after *TDRP* are summarized in Table (8.8) below. Unlike changes in the value of ' β ', the table exhibits a very insignificant impact on the stability of the results under the three values of ' $\beta\gamma$ ' before and after *TDRP*. This is because under the different values of ' $\beta\gamma$ ' the percentage of travellers experienced late schedule delay is expected to be very small compared with the percentage of those arriving early and on-time. Therefore, the WCV is slightly affected by changing the value of ' $\beta\gamma$ '.

Table (8.8) depicts a very small increase in schedule delay cost, travel time cost and weighted generalized cost, and a very small decrease in the overall average speed as a result of increasing the value of ' $\beta\gamma$ ' under different demand levels as well as before and after *TDRP*. Comparing between the results in Tables (8.7) and (8.8) reveals that the results are less sensitive to changes in the value of ' $\beta\gamma$ ' than the value of ' β '.

Table (8.8): The impacts of changing the shadow value of late schedule delay ' $\beta\gamma$ '.

Demand Level	Value of ' $\beta\gamma$ '	W. Coeff. of Var. (WCV)		T. Time Cost		S. Delay Cost		TDRP	WGC		Average Speed (km/h)	
		No RP	TDRP	No RP	TDRP	No RP	TDRP		No RP	TDRP	No RP	TDRP
Base Demand	10.00	0.162	0.089	3.05	1.07	3.56	3.88	2.90	6.61	7.85	25.75	50.31
	15.20	0.160	0.082	3.39	1.17	3.97	4.13	3.25	7.36	8.54	24.02	48.33
	18.00	0.161	0.082	3.52	1.20	4.06	4.24	3.32	7.58	8.76	23.39	47.82
1.2 Base Demand	10.00	0.139	0.080	3.75	1.10	4.28	4.71	3.96	8.03	9.77	22.17	49.51
	15.20	0.137	0.075	4.23	1.31	4.66	5.04	4.52	8.89	10.88	20.80	45.21
	18.00	0.185	0.082	4.37	1.35	4.93	5.11	4.40	9.30	10.86	20.33	45.01
1.4 Base Demand	10.00	0.163	0.075	4.62	1.14	5.02	5.52	5.01	9.64	11.67	19.32	48.48
	15.20	0.145	0.074	5.18	1.51	5.52	5.89	5.84	10.70	13.24	17.82	42.16
	18.00	0.172	0.072	5.59	1.55	5.74	6.13	6.19	11.33	13.87	17.05	41.60

* The shadow value of the early arrival time ' β ' = 3.90

Sensitivity to changes in the values of ' β ' and ' $\beta\gamma$ '.

The impacts of changing both ' β ' and ' $\beta\gamma$ ' are investigated by comparing between the base case ($\beta=3.9$ and $\beta\gamma=15.2$) with $\beta=5.5$ and $\beta\gamma=10.0$. The departure rate distributions for the two cases under the base demand are illustrated in Figure (8.36) for before and after *TDRP*. For $\beta=5.5$ and $\beta\gamma=10.0$, the departure rate distribution exhibits a higher peak compared to the base case one before

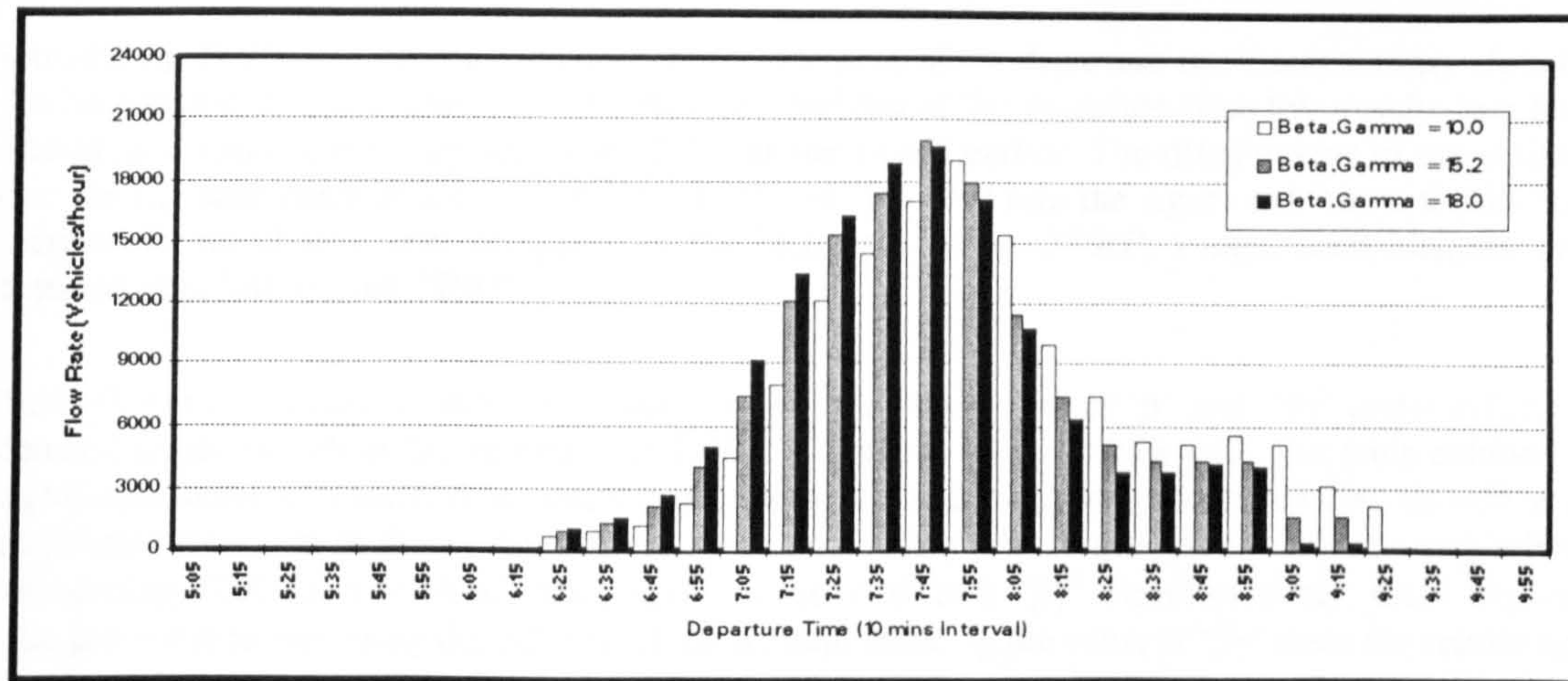


Figure (8.33): The impacts of changing the shadow value of late schedule delay ' $\beta\gamma$ ' on the departure time distribution under the base demand and no RP.

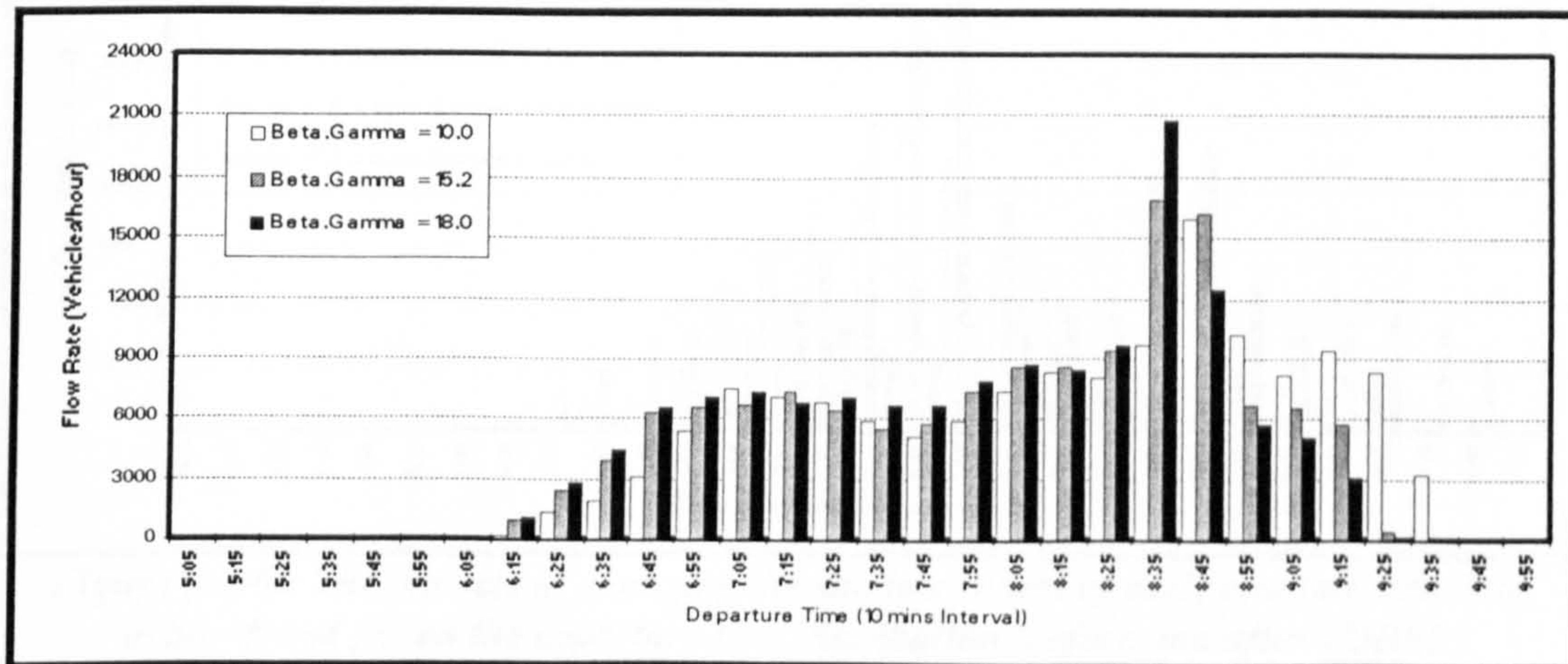


Figure (8.34): The impacts of changing the shadow value of late schedule delay ' $\beta\gamma$ ' on the departure time distribution under the base demand and TDRP.

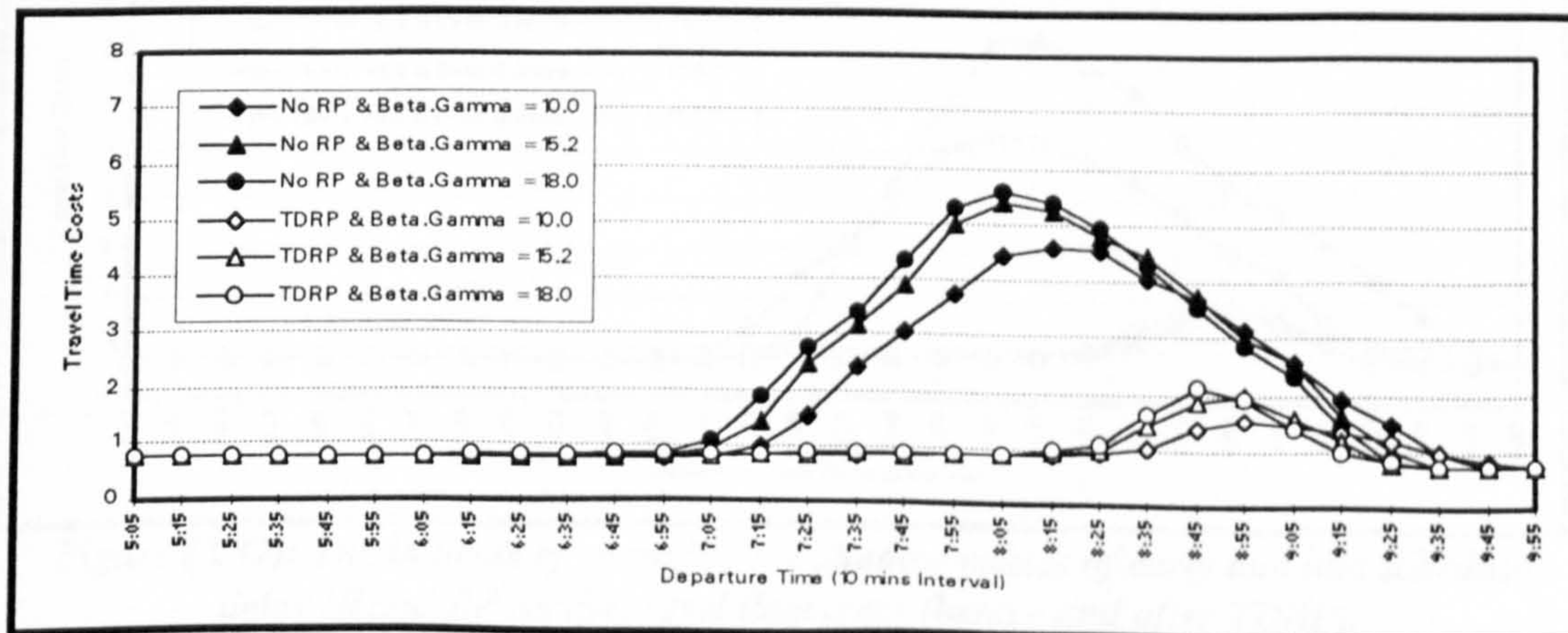


Figure (8.35): The impacts of changing the shadow value of late schedule delay ' $\beta\gamma$ ' on the travel time costs under the base demand (before and after TDRP).

introducing *TDRP*. On the contrary, after *TDRP*, the peak of the departure rate becomes very close to the base case one, and higher departure rates are depicted at the departure time that results in a late arrival as a result of reducing the value of ' $\beta\gamma$ ' as mentioned earlier. The distributions of travel time cost for the base demand are also depicted in Figure (8.37). While the figure exhibits a significant increase in travel time cost compared to the base case before *TDRP*, insignificant changes are depicted after introducing *TDRP*.

Table (8.9) summarizes the different impacts of changing the values of ' β ' and ' $\beta\gamma$ ' under different demand levels as well as before and after *TDRP*. Compared with the base case, the table exhibits a significant increase in travel time cost, schedule delay cost and weighted generalized cost, as well as a significant reduction in the overall average speed particularly with higher demand levels and before introducing *TDRP* as a result of changing the values of ' β ' and ' $\beta\gamma$ '. Needless to say, these impacts take place due to increasing the value of ' β ' rather than reducing the value of ' $\beta\gamma$ ' since the results are demonstrated to be more sensitive to changes in ' β ' than ' $\beta\gamma$ '.

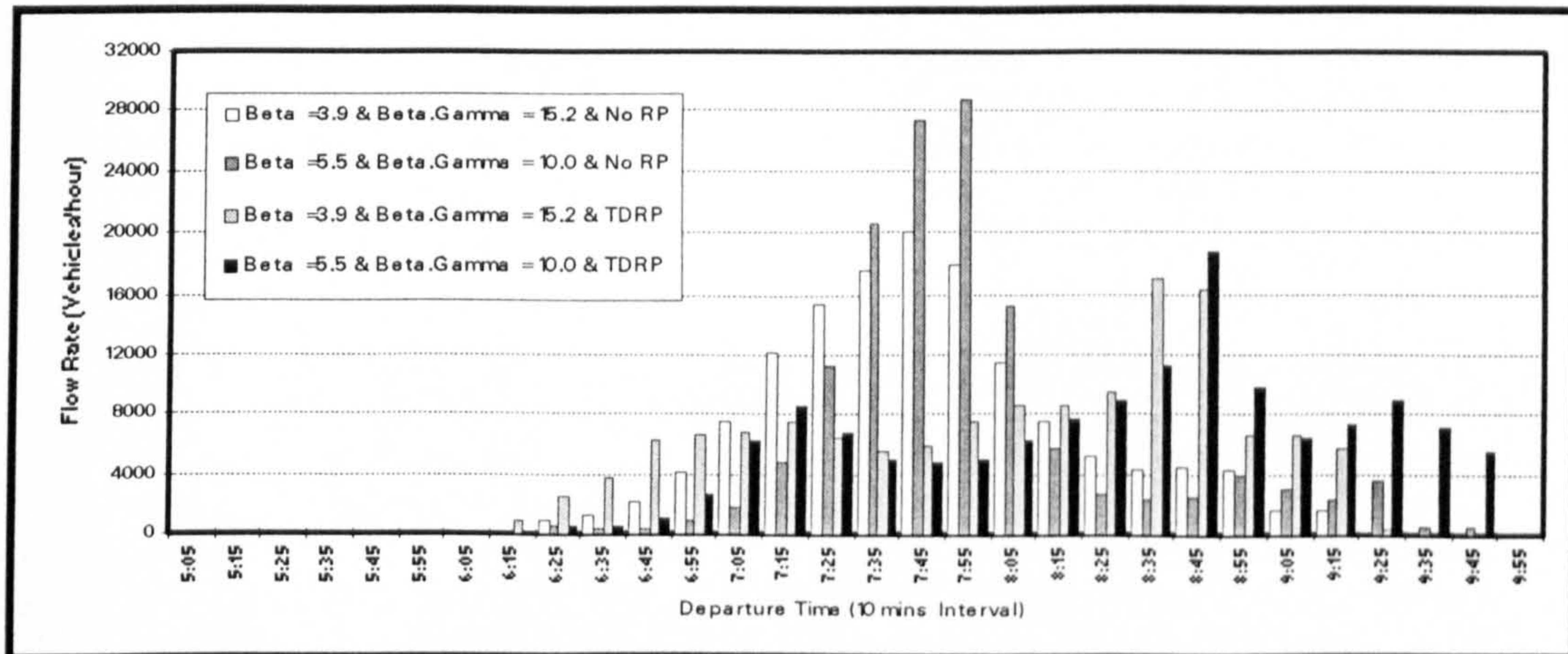


Figure (8.36): The impacts of changing the shadow values of early and late schedule delay ' β and $\beta\gamma$ ' on the departure time distribution (before and after *TDRP*).

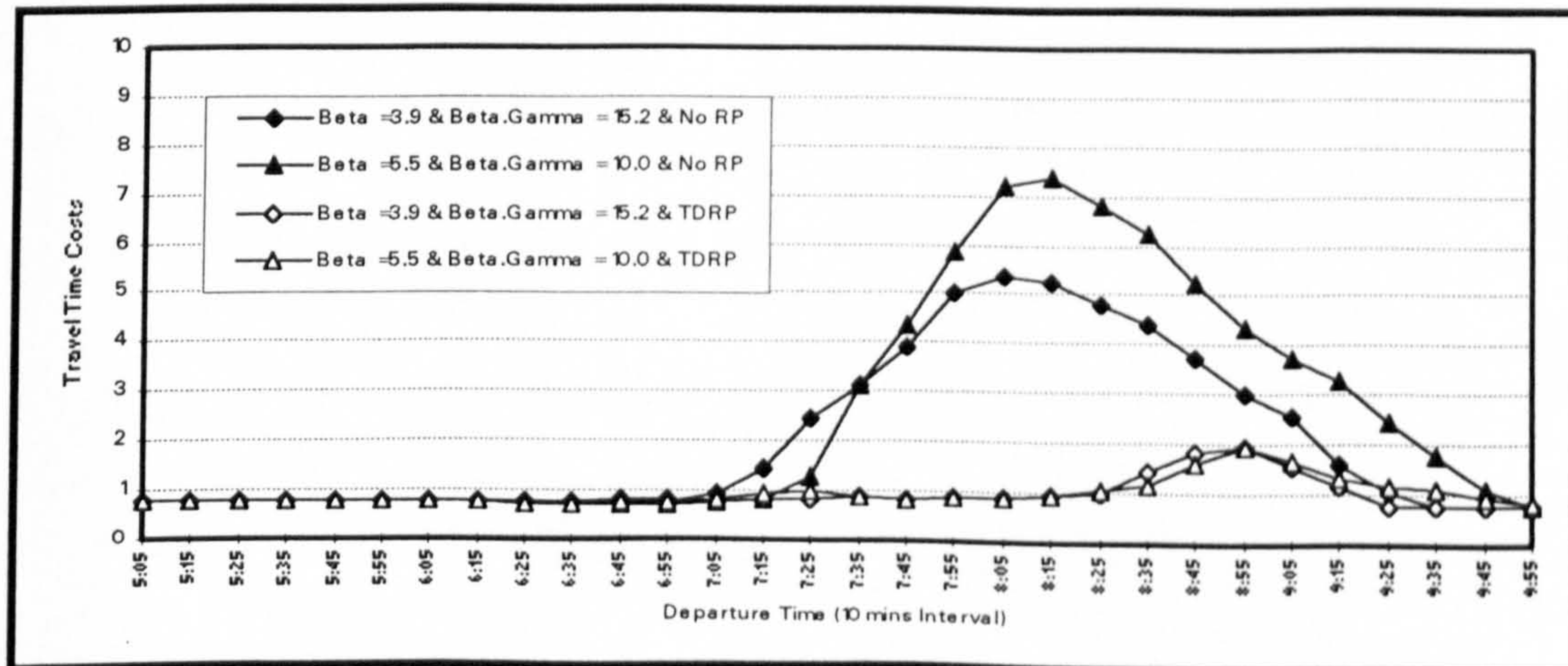


Figure (8.37): The impacts of changing the shadow values of early and late schedule delay ' β and $\beta\gamma$ ' on the travel time costs (before and after *TDRP*).

Table (8.9): The different impacts of changing the shadow values of both early and late schedule delay " β & $\beta\gamma$ ".

Demand Level	Value of ' $\beta\gamma$ '	W. Coeff. of Var. (WCV)		T. Time Cost		S. Delay Cost		TDRP	WGC		Average Speed (km/h)	
		No RP	TDRP	No RP	TDRP	No RP	TDRP		No RP	TDRP	No RP	TDRP
Base Demand	$\beta = 3.90$ $\beta\gamma = 15.2$	0.160	0.082	3.39	1.17	3.97	4.13	3.25	7.36	8.54	24.02	48.33
	$\beta = 5.50$ $\beta\gamma = 10.0$	0.211	0.077	4.49	1.24	4.72	4.84	4.02	9.21	10.10	20.19	45.99
1.2 Base Demand	$\beta = 3.90$ $\beta\gamma = 15.2$	0.137	0.075	4.23	1.31	4.66	5.04	4.52	8.89	10.88	20.80	45.21
	$\beta = 5.50$ $\beta\gamma = 10.0$	0.249	0.077	5.78	1.36	5.61	5.79	5.15	11.39	12.29	16.91	44.66
1.4 Base Demand	$\beta = 3.90$ $\beta\gamma = 15.2$	0.145	0.074	5.18	1.51	5.52	5.89	5.84	10.70	13.24	17.82	42.16
	$\beta = 5.50$ $\beta\gamma = 10.0$	0.307	0.073	7.21	1.81	6.58	6.70	6.85	13.79	15.36	14.77	37.61

8.4 Summary

In this chapter a number of simulation experiments are conducted to show the distinctive features of *TDRP* compared with other charging systems. Also, the impacts of changes in the work start time flexibility and the key parameters of the user cost function on traffic patterns and the benefits obtained from *TDRP* are investigated under different levels of congestion.

The experiments have demonstrated that *TDRP* is a superior charging system compared with others since it leads to distinctive user benefits, overall speed increase, and reduction in total queue duration under a very reasonable amount of charge. On the other hand, the benefits obtained from the flat toll equivalent to *TDRP* are remarkably less than that obtained from *TDRP* (as a time varying toll). This implies the significance of varying the charges over the time of day and its distinctive effects on the departure time choice and alleviating traffic congestion.

While distance-based charging showed the lowest user benefit, time-based and delay-based charging systems show a better performance than *TDRP* under very high charges. Thus, the *TDRP* modified by the day-to-day adjustment process does not lead to a system optimization for a traffic network, since other charging systems could show a better performance (this confirms the findings of a single bottleneck). This is because the imperfect *TDRP* function employed for a traffic network, in addition to the convergence and computational problems (as underlined in the main text). Furthermore, although time-based and delay-based charging systems are demonstrated to have the potential to lead to a system optimization under very high levels of charge, they would raise the issues of unfairness and inequity. Therefore, the optimal charging system for a traffic network is still unknown and needs further investigation.

Changes in the flexibility of the work start time are showed to have an important impact on the level of congestion developed in the network. The major effect of increasing the work start time flexibility is the spread of the departure rate distribution over a longer period, and thus to reduce the level of congestion and travel time costs. Under inflexible working time schedules, the departure rate distributions exhibit a very high peak at the departure time that results in an on-time arrival. As a result, the levels of congestion are increased due to the high concentration of traffic within a short time interval.

It is also demonstrated that introducing work start time flexibility with *TDRP* enhances the benefits obtained under different demand levels. It is thus emphasized that *TDRP* should be attached to a wide range of different strategies that could make any road pricing scheme successful in practice.

Changing the shadow value of early schedule delay ' β ' is demonstrated to have different impacts on the distribution of the departure time and the level of congestion: as ' β ' increases the level of congestion increases and vice versa. It also demonstrated that the benefits from *TDRP* are sensitive to the changes in the value of ' β '. On the contrary, the level of congestion showed less sensitivity to changes in the shadow value of late schedule delay ' $\beta\gamma$ ' under different demand level as well as before and after *TDRP*. This is because the percentage of travellers experienced late schedule delay is always very small compared with the percentage of those arriving early and on-time.

Finally, changing the values of both ' β ' and ' $\beta\gamma$ ' also showed different impacts on the departure pattern and the benefits obtained from *TDRP* under different demand levels. Since the departure pattern is more sensitive to ' β ' than ' $\beta\gamma$ ', the impacts of changing both of them are similar to that depicted under changing the value of ' β ' alone.

CHAPTER NINE

9. Conclusions and Further Research

9.1 Summary and Conclusions

9.2 Limitations and further research

9.1 Summary and Conclusions

It has been argued for many years that in order to better utilize the existing infrastructure (for road networks), a congestion toll equal to the monetary equivalent of the external diseconomies that the marginal trip maker imposes on all others should be levied on the top of the existing price (Walter 1961; Vickrey 1963; Mohring 1976; and Mills & Hamilton 1989). In other words, the optimal price should be one that reflects the full cost of making an additional trip. Evidently, the magnitude of such a congested toll varies over time and space. Although the principles have been well established for many years, much of what has been said about the practical impacts of implementing such a differential pricing scheme is based upon speculation rather than firm evidence. Also because of technical and political reasons, a full-scale pricing scheme for the use of road space has yet to be introduced.

The purpose of this research is to provide evidence of the different impacts of introducing a full-scale pricing scheme that is based on the marginal cost principle. Since the toll is differentiated by time of day, the scheme is called Time-Dependent Road Pricing '*TDRP*'. The aim of the research is to investigate the different impacts of applying this pricing scheme on a road network. To achieve this aim a list of objectives are identified in Chapter one. Having set the objectives and the overall structure of the study, the historical background of road pricing and its theoretical basis are reviewed.

The approach adopted to derive the *TDRP* function is based on the solution of time-dependent queues and delays at traffic junctions. This approach takes into account the statistical nature of vehicle arrivals and junction performance which have been ignored by the deterministic solution of queues and delays at traffic junctions. Thus, the theory of queuing and delay as well as the solution of the time-dependent queuing problem are presented to facilitate the derivation of the *TDRP* function. The *TDRP* function derived is demonstrated to have a general fixed form that is given by a linear relation with a slope of minus 45-degree (i.e. wedge-shape) and it is similar to the one derived by Fargier (1983), Small (1992), and Ghali and Smith (1993) using a deterministic approach. However, it has been demonstrated that ignoring the statistical nature of the vehicle arrivals and junction performance does not affect that shape, but it leads to an under-estimate of the toll values obtained.

In order to examine the stability of User Equilibrium '*UE*' under *TDRP*, a review of the currently available models for temporal distribution of peak period demand for a bottleneck is presented. The review is concentrated on how travellers who have particular destination target times '*DTTs*', select their departure time from the origin in a simple network under Deterministic User Equilibrium '*DUE*', Stochastic User Equilibrium '*SUE*', and System Optimal '*SO*'. The simple network used consists of one origin-destination pair connected by one major route with a bottleneck.

Examining the stability of *DUE* under *TDRP* led to the conclusion that the derived *TDRP* function leads to a very unstable equilibrium. Therefore, two different approaches are adopted to modify this function: first, by considering the schedule delay changes imposed by vehicles on one another; and second, by using the day-to-day adjustment process to smooth the *TDRP* function. The former leads to a very important conclusion: *TDRP*, modified by the schedule delay changes imposed on one another, eliminates queues and leads to the system optimal "*SO*" for a simple bottleneck.

Furthermore, a framework for the simulation solution for a single traffic bottleneck is described and used to examine the different impacts of *TDRP*, adjusted by the day-to-day adjustment process, on the results under *DUE* and *SUE*. The results demonstrated that the original pattern of *TDRP* function (with a wedge-shape and a minus 45° slope) is changed during this process to a new pattern that could trade-off the travel time cost and the schedule delay cost functions and lead to a stable equilibrium. This new pattern is a triangle-shape with an increasing fixed rate equal to the shadow value of early schedule delay ' β ' during the early arrival period and a decreasing fixed rate equal to the shadow value of travel time ' α ' afterwards. Therefore, a queue must develop during the on-time and the late arrival periods to trade-off the difference between the toll function and the late schedule delay function at equilibrium (since the shadow value of travel time ' α ' is less than the shadow value of late schedule delay ' $\beta\gamma$ '). However, although the resulting toll pattern does not eliminate the queue

(i.e. does not lead to system optimal), it results in a very substantial reduction in queuing delay and travel time.

The simulation solution is also used to demonstrate that the *TDRP* modified by including the schedule delay changes imposed on one another, leads to the system optimal for a simple bottleneck.

It is also concluded that the distribution of the departure rate under *TDRP* has a typical pattern distributed over the same period as before introducing the charge (i.e. the congested period does not spread over a longer period as a result of *TDRP*). This resulting pattern starts with a departure rate equal to (or fluctuating around) the bottleneck capacity during the period of early arrival, then the departure rate increases dramatically for a short period before it goes below the bottleneck capacity sometimes before the end of the congestion period. Therefore, the peak of this typical pattern is always shifted to the right hand side (i.e. later in time). The interpretation of building up such departure pattern is discussed in some detail.

In practice, of course, most bottleneck situations are not as simple or clear cut as the case of a simple bottleneck. Also a traffic bottleneck can not be considered in isolation from others in traffic networks. Therefore, a road network traffic assignment model is developed to investigate the different impacts of *TDRP* on a road network. This model embraces the route choice, departure time choice and *TDRP* function. It also has the ability to evaluate the road network under different charging systems.

As a result of proceeding the traffic flow from one node to another, the queues developed at nodes along the same route are demonstrated to be influenced by each other. Therefore, in applying *TDRP* for a traffic network, the interaction between nodes (i.e. between queues) emerged to have an important impact on influencing the amount of charge at each individual node along the route taken by individual vehicles from origin to destination. Moreover, it is demonstrated that the interaction between nodes along the route is not restricted to every two successive nodes, but any pair of nodes along that route could be interrelated and could influence the value of *TDRP* paid by every individual vehicle passing through them. Therefore, the charges paid by individual vehicles at a certain node throughout the network is concluded to be dependent on the route chosen by the vehicles from origin to destination.

In an attempt to solve the problem of the interaction between nodes, an algorithm is suggested for a general network under specific conditions (i.e. fixed merging and diverging flows throughout the study period). However, the solution for this problem under all traffic conditions poses difficult analytical issues and needs further attention and additional investigation.

Since the interaction between nodes is left unresolved under general traffic conditions, three different scenarios of *TDRP* are adopted to examine the different impacts on the results and to explore the significance of charging some nodes and exempting the others. These three scenarios are: charging all nodes, charging only the entrances to the network, and charging only the most congested nodes.

Numerical simulation experiments using the assignment model are conducted to analyse the different impacts of the three scenarios of *TDRP* on a typical traffic network during the morning peak period. The main conclusions derived from these experimental analysis are summarized below:—

- The resulting distribution of the departure rate under *TDRP* for any individual OD pair, all movements from any individual origin, and for the system as a whole showed a similar pattern to the one depicted for a single bottleneck.
- Although *TDRP*, under any of the three scenarios, does not eliminate the queues completely, it leads to a very substantial saving in travel time and queuing duration. Therefore, it could be concluded that *TDRP* would have the potential to achieve an efficient use of the network facilities throughout the peak period under different levels of demand.
- Although exempting some nodes from the charge would lead to a substantial travel time saving, this amount of saving is very much less than that obtained from charging all the congested nodes.

Another set of simulation experiments are conducted to: (i) show the distinctive features of *TDRP* compared with other charging systems; and (ii) investigate the impacts of changes in the work start time flexibility and the key parameters of the user cost function on traffic pattern and the benefits obtained from *TDRP*. The main conclusions derived from this experimental analysis are summarized below:—

- As a result of varying the charges over time, *TDRP* is demonstrated to be a superior charging system compared with other charging systems. It leads to distinctive user benefits, overall speed increase and reduction in total queue duration under different levels of congestion.
- Under very high levels of charge, time-based and delay-based charging systems showed a better performance than *TDRP*, and they also could lead to a system optimization. Thus, the *TDRP* modified by the day-to-day adjustment process does not lead to a system optimization for a traffic network, since other charging systems could show a better performance (this confirms the findings of a single bottleneck). However, deciding the system of charging in practice is not as simple as that, but there are some other factors that should be considered, e.g. the complexity of implementing and running the system, travellers' preference to the system, availability of information, ... etc. Therefore, the optimal toll for a road network is unknown and still needs to be investigated further.
- The level of work start time flexibility has a major impact on the development of the congestion patterns during the morning peak. Flexible work start time allows travellers to arrive at their destinations within a specific period of time during the morning peak and thus, the longer this period is the wider the spread of the demand over time will be. The major effect of increasing the work start time flexibility is therefore the spreading of the departure rate distribution over a longer time period which results in lower levels of congestion and lower travel times. This confirmed the view that inflexible working times result in higher levels of congestion due to the high concentration of traffic within a short time interval while higher levels of work start time flexibility result in lower travel times. It is also demonstrated that introducing work start time flexibility with *TDRP* enhances the benefits obtained under different levels of congestion. This implies that *TDRP* could be attached to a wide range of different strategies that could make any road pricing scheme successful in practice.
- Changing the shadow value of early schedule delay ' β ' is demonstrated to have different impacts on the distribution of the departure time and the level of congestion: as ' β ' increases the level of congestion increases and vice versa. It is also demonstrated that the benefits from *TDRP* are sensitive to the changes in the value of ' β '. On the contrary, the level of congestion showed less sensitivity to changes in the shadow value of late schedule delay ' $\beta\gamma$ ' under different demand level as well as before and after *TDRP*. This is because the percentage of travellers experienced late schedule delay is always very small compared with the percentage of those arriving early and on-time. Furthermore, changing the values of both ' β ' and ' $\beta\gamma$ ' also shows different impacts on the departure pattern and the benefits obtained from *TDRP* under different demand levels. Since the departure pattern is more sensitive to ' β ' than ' $\beta\gamma$ ', the impacts of changing both of them are similar to that depicted under changing the value of ' β ' alone.

The major achievements of the study are summarized as follows:—

1. Deriving a mathematical model for estimating the amount of charge that is equivalent to the queuing delay imposed by every individual vehicle on others behind. This model is based on the solution of the time-dependent queuing problem.
2. Demonstrating that the derived *TDRP* model (without any modification) does not lead to a user equilibrium. In other words, it leads to a very unstable user equilibrium.

3. The optimal toll for a simple bottleneck is equivalent to the schedule delay imposed by every vehicle on others. This toll eliminates the queue completely and leaves the system running under free flow conditions.
4. Development of a dynamic network traffic assignment model which defines travellers' departure time and route choice and the resulting time varying traffic pattern during the morning peak. The model also has the ability to evaluate a wide range of different charging systems including *TDRP*.
5. Raising the importance of the phenomenon of interaction between nodes and its influence on the amount of charge in applying *TDRP* system for a traffic network. Also, an algorithm is suggested for solving this problem under very specific traffic conditions (i.e. fixed merging and diverging flows).
6. Although *TDRP* modified by the day-to-day adjustment process does not lead to a system optimization for either a single traffic bottleneck or a traffic network, it has a very distinctive feature in reducing traffic congestion and travel time. Thus, it would have the potential to achieve an efficient use of the network facilities throughout the peak period under different demand levels.

Finally, a full evaluation of *TDRP* system shows that there are four major aspects that influence the decision of implementing such pricing scheme in practice. These four aspects are discussed below:—

- (i) The first aspect implies the availability of data and information required for calculating the charges, e.g. traffic pattern and the starting and vanishing times of queues over every individual link throughout the network. It is also necessary to have enough information about the stochastic nature of the drivers' behaviour and their likely response to the time-varying charge, and the average values of the parameters representing these behaviours for the whole population. These behaviours are represented in this study by the probabilistic nature of the departure time choice model and its parameter η and the reviewing factor $R1$, as well as the parameters of the cost function (B , Δ , α , β , and $\beta\gamma$) and the day-to-day toll adjustment factor $R2$.
- (ii) The success of such pricing scheme will depend crucially on the behavioural response of drivers to the system which in turn depends on the provision of information providing the drivers with a larger spatial and temporal choice set from which they may select a suitable plan of action. Thus, drivers should be provided with improved up-to-the-minute information on delays and tolls using local radio stations and/or electronic road signs.
- (iii) Since the original *TDRP* function does not lead to convergence, the final (equilibrium) value of toll adjusted by the day-to-day adjustment process can not be calculated mathematically. Therefore, the simulation solution should be adopted under all circumstances. Thus, this will raise the issue of the complexity of simulating a bigger network and estimating the charges for every individual link.
- (iv) Lastly, but most important, is the economic evaluation of the scheme using cost/benefit analysis. In considering the cost/benefit analysis, the costs of the above three aspects (i), (ii) and (iii) should be included in addition to the costs of installing and operating the system. The sum of these costs is compared with the users and non-users' benefits from the system. The users benefit is equal to the travel cost saving for all users (i.e. the sum of travel time and delay saving and schedule delay saving for all users). Furthermore, in addition to the alternative pricing schemes used for the cost/benefit analysis, different scenarios of *TDRP* should also be considered based on the number of nodes (links) to be charged. However, it is worth mentioning that the economic evaluation depends on the traffic network, traffic conditions and the drivers' behaviour and their response to the system. Therefore, the results of this evaluation will be very specific to the case under consideration and can not be extended further under different conditions.

9.2 Limitations and Further Research

As an initial effort to examine a full-scale time differentiated toll scheme for a road network, a number of aspects are not examined in depth and more research is still needed. Below, the limitations of this study that need to be considered for further improvements and research are discussed.

- In estimating the value of *TDRP*, no considerations are given to either the costs imposed on non-users (e.g. noise, visual, atmospheric pollution, accident, delays to bus passengers, pedestrian, commercial vehicles and emergency services, etc.) or to the fixed cost of providing the road "by the supplier". Therefore, the *TDRP* considered in this study does not necessarily represent the full marginal social cost pricing as it refers only to the travel market in isolation and not to the impacts of this travel upon the rest of the society.
- The main source of weakness in calculating the value of *TDRP* relates to the valuation of travel time particularly non working times. This is because individuals vary in values they assign to time spent in various places at various times during the peak period. However, in considering the large numbers involved on the congested road network, it is apparent that an average figure may legitimately be used. The problem is how to obtain average figures for the valuation of journey time.
- The study also assumes that the parameters of the departure time model (η , R_1 , B , Δ , α , β , and $\beta\gamma$) are the same for all individuals in the population. A more realistic model must explicitly account for the differences in these parameters by assuming, for example, that the population can be divided into N homogenous segments and each segment has its own parameters.
- The dynamic adjustment process for the departure time choice model assumes that the trip decisions on any day (ω) are based on the distribution of travel times on day ($\omega-1$). Thus, road users are assumed to be perfectly informed about road conditions during the previous day. Therefore, it would be useful to explore situations in which individuals are not perfectly informed, differ in their level of information and ability to form expectations, and base their decisions on their past experience. Further research is also required to develop more realistic procedures which will model the day-to-day adjustment of traveller's choice. Moreover, empirical work is required in order to identify the nature of information acquired by travellers, and the mechanisms used by them to process this information in conjunction with their own experiences.
- An area that requires further research involves the formulation of the link performance functions and their use in the dynamic assignment procedures. A queuing theory based travel time model would provide a more realistic representation, since it has the ability to represent the evolution of queues and the delays associated with them. However in such a procedure, the travel time needed to traverse a link should not be constant, but flow-dependent.
- The network traffic simulation model is primarily developed for assigning traffic in urban express traffic networks, where all intersections are grade-separated. Therefore, it can only handle one type of traffic junctions (i.e. un-controlled intersections or priority junctions) and no other methods of traffic control are considered. The simulation model has also another major shortcoming in handling more than two links merging at a single node. Thus, there still remains a need for further improvements in the simulation model which concentrate upon including a traffic-light system which is the most common traffic control system in urban areas, and allowing more than two links to merge at any single node.
- The assignment model assumes that the number of trips between every OD pair is fixed. The possible responses to the charge in the demand side are to change departure time or divert to alternative routes or to do both of them at the same time. In other words, the model does not make any allowance for the elasticity of car travel; and nor does it consider modal split. Thus, a more complete model should allow for the options that may be available, such as alternative modes (or car pooling) and destinations, and the option of not making the trip.

- An important area of interest involves the issue of interaction between nodes and its impacts on the value of *TDRP* at every individual node throughout the network. This issue is left unresolved and needs to be investigated further under different types of traffic junctions (i.e. traffic control) and different traffic conditions.
- Since individuals are assumed to have a pre-defined desired arrival time at their end destinations rather than at every individual node throughout the network, the changes in schedule delay imposed by vehicles on one another at every node can not be considered in applying *TDRP* for a traffic network. This is because, in principle, the value of *TDRP* is calculated on a link (or node) basis. However, including these costs to the original *TDRP* function is demonstrated to eliminate the queues completely and lead (by their own) to the optimal toll for a single bottleneck. Thus, *TDRP* function applied for a road network is imperfect and more work is needed in order to include these costs to the original function of *TDRP* applied for a traffic network.
- Throughout the study, the simulation experiments are performed to demonstrate the possible changes in the pattern of peak-period congestion when the capacity of a bottleneck (i.e. traffic junction) is unchanged. However, a more realistic situation where the capacity of the bottleneck changes (such as using different phases for the cycle of traffic signals during the peak period) needs to be investigated.
- Another area of interest involves the rate of convergence to a stationary state that is dependent on the value of the reviewing rate $R1$ (the percentage of motorists changing their departure time from one day to another for every OD movement) and the toll day-to-day adjustment factor $R2$. For higher values, a convergence to a stationary state is not guaranteed. Simulation experiments consistently show that with higher initial values for these two factors and then decreasing them gradually during iterations convergence occurs towards a unique stationary state. However, in almost all simulation experiments, the simulation terminates when the last specified day is reached (i.e. after 60 days) and the specified convergence ratio is not achieved (weighted coefficient of variation of the generalized cost 'WCV' is less than 5%). It is therefore necessary to investigate how to improve the convergence and explore the other factors affecting its speed. Further work should also be given to the properties of the final state (i.e. stationary state) of the system and the stability of its dynamic evolution.
- All of the above are technical issues of modelling. The introduction of *TDRP* (and road pricing in general) will always fall short of the optimal theoretical solution, which is still unknown for a traffic network and needs further investigation. Also, there is a need to understand the benefit loss due to tolling (charging) only a restricted set of links and/or a restricted set of vehicles over time periods. This, however, involves a trade-off between the benefits from introducing the system and the costs of installation and operation. A greater understanding of this trade-off is needed.

References

Abkowitz M., (1980), "The Impact of Service Reliability on Work Travel Behaviour", PhD Thesis, Department of Civil Engineering, Massachusetts Institute of Technology.

Abkowitz M., (1981), "Understanding the effect of transit service reliability on work travel behaviour", Transportation Research Record 794, pp. 33-41.

Alfa A. S. and Minh D. L., (1979), "A stochastic model for the temporal distribution of traffic demand- the peak hour problem.", Transportation Science, No. 13, pp 315-324.

Alfa A. S., (1986), "A review of models for the temporal distribution of peak traffic demand", Transportation Research B, No. 6, pp 491-499.

Armstrong-Wright, A. T. (1986), "Road Pricing and User Restraint: Opportunities and Constraints in Developing Countries", Transportation Research A, vol. 20A, No. 2, March 1986, pp. 123-127.

Beckmann, M., McGuire C. B. and Winston C. B. (1956), "Studies in the Economics of Transportation.", Yale University Press, New Haven.

Beesley, M. E. (1969), "The technical possibility of special taxation in relation to congestion caused by private users", ECMT, Second International Symposium on Theory and Practice in Transport Economics, Rome.

Ben-Akiva M. and de Palma A. (1983), "Modelling and Analysis of Dynamic Residential Location Choice", Working Paper No. 83-19. Department of Economics, McMaster University, Hamilton, Ontario.

Ben-Akiva M. and Lerman, S. R., (1985), "Discrete Choice Analysis: Theory and Application to Travel Demand", The MIT Press, Cambridge, Massachusetts. London, England.

Ben-Akiva M. and Watanatada T., (1981), "Applications of a continuous spatial choice logit model", In Structural Analysis of Discrete Data with Economic Applications (Edited by C. Manski and D. McFadden). MIT Press, Cambridge, Mass.

Ben-Akiva M., Cyna M. and de Palma A. (1984), "Dynamic model of peak period congestion", Transportation Research , vol. 18B, No. 4/5, pp. 339-355, 1986.

Ben-Akiva M., de Palma A. and Kanaroglou P. (1986a), "Effects of Capacity Constraints on Peak-Period Traffic Congestion", Transportation Research Record 1085, pp 16-26.

Ben-Akiva M., de Palma A. and Kanaroglou P. (1986b), "Dynamic Model of Peak-Period Traffic Congestion with Elastic Arrival Rates", Transportation Science, Vol. 20, No. 2, August 1986, pp 164-181.

Bhatt, K. (1974), "Road Pricing Technologies: A Survey.", The Urban Institute, Washington, DC.

Boardman A. E. and Lave L. B. (1977), "Highway Congestion and Congestion Tolls.", Journal of Urban Econ., No. 4, pp. 304-359.

Borins, S. (1984), "The Economic Effects of Non-optimal Pricing and Investment Policies for Substitutable Transport Facilities.", Canadian, J. Econ. No. 17, pp. 80-99.

Borins, S. (1988), "Electronic Road Pricing: An idea whose time may never come.", Transpn. Res., Vol. 22A, No. 1, pp. 37-44, 1988.

Buchanan, C. and Partners (1986), "Traffic Limitation Scheme: An Appraisal for the commune de Milano.", London, 1986.

Buchanan, James M., (1952), "The Pricing of Highway Services", National Tax Journal, vol. (V), No. 2, June 1952.

Button, K. J. (1983), "Road Pricing: Some Practical and Economic Considerations for its Introduction in Britain", Logistics and Transportation Review, Vol. 18, No. 4, pp. 353-371.

Button, K. J. and Pearman, A. D. (1983), "Road Pricing— some of the more neglected theoretical and policy implications.", Transportation Planning and Technology, 1983, No. 8, pp. 15-28.

Catling, I. and Harbord, B. J. (1985), "Electronic road pricing in Hong Kong: 2. The Technology", Traffic Engineering and Control, Vol. 26, No. 12, pp. 608-615, Decemder 1985.

Churchill, A. with Huber K., Meldau E., and Walter A., (1972), "Road User Charges in Central America.", World Bank Staff Occasional Paper No. 15, Washington, DC.

Cosslett S., (1977), "The trip timing decision for travel to work by automobile: Demand Model Estimation and Validation", The Urban Travel Demand Forecasting Project Phase I, Final Report, Vol. V, Institute of Transportation Studies, University of California, Berkeley, CA.

Coventry P. S. and Dickinson K. (1977), "Sheffield and Rotherham L.U.T.S.: Urban Traffic Control for Sheffield", Traffic Engineering and Control, No. 18, pp. 342-346.

Daganzo C. F. and Sheffi Y., (1977), "On stochastic models of traffic assignment", Transportation Science, No. 11, pp 253-274.

Daganzo, C. (1985), "The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck", Transp. Sci. 19, pp. 29-37.

Dawson, J. A. L. and Catling, I., (1986), "Electronic Road Pricing in Hong Kong.", Transp. Res. A, vol. 20A, No. 2, pp. 129-134, March 1986.

Dawson, R. F. F. (1961), "An Analysis of the cost of Road Improvement Schemes.", Road Research Technical Paper No. 50, HMSO, 1961.

de Palma A. and Ben-Akiva M., (1981), "An interactive dynamic model of residential location choice", Paper prepared for presentation at the Structural Economic Analysis and Planning in Time and Space Conference, Umea, Sweden.

de Palma A. and Lefever C., (1983), "Individual decision making in dynamic collective systems", J. Math. Socio. No. 9, pp 103-124.

de Palma A., Ben-Akiva M., Lefevre C. and Litinas N., (1983), "Stochastic Equilibrium Model of Peak Period of Traffic Congestion", Transportation Science, No. 17, pp 430-453.

De Vany A. and Saving T. R. (1980), "Competition and highway pricing for stochastic traffic", J. Business 53, pp. 45-60.

Deen, Thomas B. (1963), "Fiscal Policies and Transportation Planning", Traffic Quarterly, January 1963.

Department of Environment, (1976a), "A Study of Some Methods of Traffic Restraint", Research Report 14, (1976), Report prepared by a DoE Working Group.

- Department of the Environment, (1976b), "A linked system of traffic signal with priority access to the route for buses", Bus Demonstration Project Summary Report No. 8, DoE, Southampton, London.
- Dial R. B., (1971), "A Probabilistic Multi-path Traffic Assignment Algorithm which Obviates Path Enumeration", *Transportation Research* Vol. 5, No. 2, pp. 83-111.
- Dijkstra, E. (1959), "A note on two problems in connexion with graphs", *Numerical Mathematics* 1, pp. 269-271.
- Dupuit, J. (1844), "On the measurement of the utility of public works", *Annales des ponts et Chausees*, 2nd series, vol. 18 (1844), translated by R.H. Barback, *International Economic Papers* No. 2, London, 1952.
- Dupuit, J. (1849), "On tolls and transportation charges", Reprinted in *Institute of Economic papers* No. 11, London, 1962.
- Ellet, C. (1840), "A popular Exposition of the Incorrectness of the Tariffs on Tolls in Use on the Public Investments of the United States", J. Franklin Institute, 1840.
- Elliott, W. (1986), "Fumbling Toward the End of History", *California's Quest for a Road Pricing Experiment*. Transportation Research, 20A, Oxford, 1986.
- Elliott, W. E. (1976), "Road Use Charges and Jitneys: Some Thoughts on How to Introduce them to Los Angeles.", Rose Institute, Claremont, CA.
- Else, P. K. (1984), "Road Pricing: some Further Comments", in *Transportation Planning and Technology*, 1984, No. 8, pp. 295-300.
- Evans, A. W. (1992), "Road congestion pricing: when is it a good policy?", *Journal of Transport Economics and Policy*, 26, pp 213-243.
- Everall, P. (1972), "Urban Freeway Surveillance and Control: The State of the Art". FHWA, Washington.
- Fargier, Paul-Henri, (1983), "Effects of the Choice of Departure Time on Road Traffic Congestion", in *Proceedings of the eight International symposium on Transportation and Traffic Theory*, ed. by V. F. Hurdle, E. Hauer, and G. N. Steuart. Toronto: University of Toronto Press (1983), pp. 223-263.
- Farid Amara, (1994), "Smart solution for traffic managment", *Traffic Technology International*, Winter 1994, pp. 30-32.
- Fitch, Lyle C. and Associates (1963), "Urban Transportation and Public Policy.", Chandler Publishing Co., San Francisco, 1963.
- Florian M. and Fox B., (1976), "On the Probabilistic Origin of Dial's Multi-path Traffic Assignment Model", *Transportation Research*, Vol. 10, pp. 339-341.
- Fong, P. K. W. (1985), "Issues of the electronic road pricing system in Hong Kong", *Transportation Planning and Technology*, 1985, No. 10, pp. 29-41.
- Gaver L. R., (1967), "Headstart strategies for combating congestion", *Transportation Science*, No. 2, pp 172-181.
- Gerlough, D. L. and M. J. Huber (1975), "Traffic flow theory.", *Special Report 165*. Transportation Research Board, Washington DC.

- Ghali, M. O. and Smith, M. J., (1993), "Traffic Assignment, Traffic Control and Road Pricing.", Proc., 12th Int. Symp. On Transportation and Traffic Theory, pp 147-169.
- Goodwin, P. B. (1989), "The 'Rule of Three': A possible solution to the political problem of competing objectives for road pricing." Traffic Engineering and Control, 30, pp 495-497.
- Goodwin, P. B. and P. M. Jones (1989), "Road Pricing: The political and strategic possibilities", Report prepared for Round Table 80, "Systems of Infrastructure Cost Coverage (Road Pricing, Principles and Application)", European Conference of Ministers of Transport, Paris, February, 1989.
- Gross, Donald and Harris, Carl M., (1985), "Fundamentals of queuing theory", Second Edition, Wiley series in probability and mathematical statistics.
- Haight F. A., (1963), "Mathematical Theories of Traffic Flows", Academic Press Inc. New York, 1963.
- Henderson J. V., (1974), "Road Congestion: a reconsideration of pricing theory", J. Urban Econ. No. 1, pp 346-365.
- Henderson J. V., (1977), "Economic Theory and the Cities", Chap. 8. Academic, New York (1977).
- Henderson J. V., (1981), "The economics of staggered work hours", J. Urban Econ. No. 9, pp 349-364.
- Hendrickson C. and Kocur G., (1984), "Schedule delay and departure time decisions in a deterministic model", Transportation Science, No. 15, pp 62-77.
- Hendrickson C. and Plank E., (1984), "The flexibility of departure times for work trips", Transportation Research A, Vol. 18, pp. 25-36.
- Hendrickson, C. and G. Kocur (1981), "Schedule delay and departure time decisions in a deterministic model", Transp. Sci. 15, pp. 62-77.
- Hewitt, Patrica, (1989), "A Clear, Faster London: Road Pricing, Transport Policy and the Environment" Green Paper No. 1.
- Highway Research Record (1964), "Traffic Congestion as a Factor in Road-User Taxation", Report No. 47 (6 reports), Presented at the 43rd Annual Meeting, January 13-17, 1964.
- Hollis, Erica M. and Kimber, R. M. (1977), "Queues and delays at road junctions with time-varying traffic demand and capacity", TRRL Working Paper TSN 41, 1977 (unpublished- available on direct personal request only).
- Hong Kong Government (1985a), "Electronic Road Pricing Pilot Scheme.", Government Printer, Hong Kong.
- Hong Kong Government, (1985b), "A Fair Way to Go.", Hong Kong Government Publications, Hong Kong.
- Horowitz J. L., (1984), "The Stability of Stochastic Equilibrium in a Two-Link Transportation Network", Transportation Research B, vol. 18, No. 1, pp 13-28.
- Hurdle V. F., (1981), "Equilibrium Flows on Urban Freeways", Transportation Science, No. 15, pp 255-293.

- Hurdle V., (1974), "The effects of queuing on traffic assignment in a simple road network", Proceedings of the 6th Int. Symp. on Transp. and Traffic Theory, University of New South Wales, Australia, 1974.
- Inman, R. P. (1978), "A generalised congestion function for highway travel.", Journal of Urban Econ. No. 5, pp. 21-34.
- Kimber, R. M. (1976), "The capacity of some major/minor priority junction", Department of the Environment, Department of Transport, TRRL Report LR 735, Crowthorne, Transport and Road Research Laboratory.
- Kimber, R. M. and Hollis, Erica M. (1979), "Traffic queues and delays at road junctions.", Department of the Environment, Department of Transport, TRRL Report LR 909, Crowthorne, Transport and Road Research Laboratory.
- Kimber, R. M., (1978), "Traffic delays and queues in conditions of time-varying demand: The co-ordinate transformation method", TRRL Working Paper TSN 47, 1978 (unpublished- available on direct personal request only).
- Kimber, R. M., Summersgill, I. and Burrow, I. J., (1986), "Delayed Processes at un-signalised junctions: the inter-relation between geometric and queuing delay", Submitted to Transportation Research 1986.
- Knight, F. H. (1924), "Some facilities in the interpretation of social cost.", Quarterly Journal of Economic, vol. 38, pp. 582-606.
- Kraus M. (1982), "Highway pricing and capacity choice under uncertain demand.", J. Urban Econ. 12, pp. 122-128.
- Kulash, Damian J. (1974), "Income-distributional consequences of road pricing", The Urban Institute, Washington, D.C., July 1974.
- Lane, R. and Hodgkinson D. M., (1976), "A Permit System for Traffic Restraint". Traffic Engineering and Control No. 17, pp. 94-97.
- Larsen, O. I. (1988), "The Toll Ring in Bergen, Norway - The First year of Operation.", Traffic Engineering and Control, London, 1988.
- Larsen, Odd. I. and Ramjerdi, Farideh, (1990), "The Toll Rings in Norway in the Prospective of Road Pricing.", Institute of Transport Economic- Oslo, Norway.
- Leonard D. R. and Gower P., (1982), "User guide to CONTRAM Version 4", Department of the Environment, Department of Transport TRRL Report, SR735, Transport and Road Research Laboratory, Crowthorne.
- Leonard D. R., Gower P. and Taylor N. B., (1989), "CONTRAM: Structure of the Model", Department of Transport TRRL Report, RR178, Transport and Road Research Laboratory, Crowthorne.
- Leonard, D. R. and Gower, P., (1982), "User guide to CONTRAM Version 4", Department of the environment Department of Transport, TRRL Supplementary Report 735, Crowthorne, Transport and Road Research Laboratory.
- Lighthill M. J. and Whitham G. B., (1955), "On Kinematics Waves; II. A Theory of Traffic Flow on Long Crowded Roads", Proc. Roy. Soc. (London), ser. A, Vol. 229, No. 1178.

Litinas N. and Ben-Akiva M., (1979), "Behavioural modelling of continuous spatial distributions of trips, residential locations and workplaces", Unpublished paper, Massachusetts Institute of Technology, Cambridge, Mass.

Mahmassani H. and Herman R., (1983), 'Dynamic user equilibrium departure time and route choice on idealised traffic arterials', Transportation Science, No. 18, pp 362-384.

Mahmassani H. S. and Chang G.-L., (1985), 'Dynamic Aspects of Departure-Time Choice Behaviour in a Commuting System: Theoretical Framework and Experimental Analysis', Transportation Research Record No. 1037, pp 88-101.

Mahmassani H. S. and Chang G.-L., (1986), 'Experiments with departure time choice dynamics of urban commuters', Transportation Research B, No. 4, pp 297-320.

Mahmassani H. S., Chang G.-L. and Herman R., (1984), 'Individual Decisions and Collective Effects in a Simulation Traffic System', University of Texas at Austin 1984.

Martin Wohl and Brian V. Martin (1967), "Traffic System Analysis for engineers and Planners".

Martin, W. (1981), "Pricing Urban Transportation: a critique of current policy", APA Journal, 47 (3), pp. 243-251, July 1981.

May, A. D. (1978), 'Supplementary Licensing: An Evaluation.', Transportation Research Board, SR 181, pp. 30-38.

May, A. D., (1983), 'Making Progress with Traffic Restraint: The Role of Research', Transp. Res. Rec. 906, pp. 10-17.

May, A. D., (1986), "Traffic Restraint: A Review of the Alternatives". Transp. Res. A, vol. 20A, No. 2, pp. 109-121, 1986.

May, A. D., D. S. Milne, D. Van Vliet, M. J. Smith and M. Wisten (1994), "The network effects of alternative road pricing systems.", Proceedings of the 3rd IEE International Conference on Traffic Monitoring and Control, Institute of Civil Engineers, London.

Maycock, G., (1974), "Capacity, safety and delay at new types of roundabout with offside priority", PIARC 12th International Study Week: Traffic Engineering and Safety, Theme IX.

Meland, S. (1994), "Generalised and Advanced Urban Debiting Innovations: The GAUDI Project, 3. The Trondheim Toll Ring", Traffic Engineering + Control, Vol. 36, No. 3, pp. 150-155, March 1995.

Meyer, J. R. and Straszheim, M. R. (1971), 'Pricing and Project Evaluation.', The Brookings Institution.

Mills, E. S. and Hamilton, B. W., (1989), Urban Economics. Glenview, Illinois: Scott, Foresman and Co.

Milne, D. (1993), "The network effects of alternative road user charging systems." Proceeding of Seminar F, PTRC Annual Summer Meeting, pp 91-102.

Milne, D. S., A. D. May and D. Van Vliet (1994), 'Modelling the network effects of road user charging.', Proceedings of the International Conference on Advanced Technologies in Transportation and Traffic Management, Centre for Transport Studied, Nanyang Technological University, Singapore, pp 113-120.

- Minh D. L., (1977), 'Discrete time, single server queue from a finite population', *Management Sci.* No. 23, pp 765-767.
- Mohring, H., (1976), *Transportation Economics*. Cambridge, Mass.: Ballinger.
- Morrison, S. A. (1986), "A survey of road pricing", *Transportation Research*, Vol. 20A, No. 2, pp. 85-97, March 1986.
- Murdoch, J. (1978), "Queuing Theory: Worked examples and problems."
- MVA consultant and Buchanan C. & Partners (1989), *Strategic Planning Advice for London: Desk Studies of Transportation Topics Report for London Planning Advisory Committee*, March 1989.
- Newell, G. F., (1971), "Applications of queuing theory", First published 1971, Printed in Great Britain by Willmer Brothers Limited Birkenhead.
- OECD, (1973), "Techniques of Improving Urban Conditions by Restraint of Road Traffic".
- Oldridge B., (1990), 'Electronic Road Pricing: an answer to traffic congestion?', *Proc. IT and Traffic Management*. London, NEDO.
- Pigou A., (1932), "The Economic of Welfare", MacMillan, 4th edition, London.
- Pigou, A. C. (1920), "wealth and welfare.", Macmillan London.
- Pretty, Robert L., (1988), 'Road Pricing: A Solution for Hong Kong?', *Transportation Research A*, vol. 22A No. 5, pp. 319-327.
- Ramerjdi, F. (1989), 'Evaluation of the Combination Area Licensing Scheme and Different Public Transport Studies as Policy Measures in the Stockholm County.', *Proceedings of the International Conference on Urban Environmental Improvement and Economic Development*, Berlin., 1989.
- Reynolds, D.J. (1961), 'Planning Transport and Economic Forces', *Journal of the town planning Institute*. vol. XLVII, 1961.
- Roth, G. (1967), 'Paying for Roads: The Economics of Traffic Congestion', Baltimore, Penguin Books, 1967.
- Roth, G. J. (1966), "A self-financing road system.", Published by the Institute of Economic Affairs, 1966.
- Roth, G. J. and Thomson, J. M., (1963), 'Road Pricing- a cure for congestion?', *Aspect*, April 1963.
- Roy, Thomas (1991), "Traffic Assignment Techniques", Avebury Technical.
- RRL, (1965), "The Costs Arising from the Use of Roads by Goods Vehicles.", *Road Research Laboratory Note No. LN/751/JMT*, January 1965 (unpublished).
- Satty, T. L., (1961), "Elements of Queuing Theory with Applications.", 1961.
- Small K., (1982), "The scheduling of consumer activities: work trips.", *American Economic Review*, No. 72 (3), pp 467-479.
- Small, A. Kenneth, Clifford Winston and Evans, Carol A. (1989), *Road Work. "A New Highway Pricing Investment Policy"*, The Brookings Institute, Washington, DC.

- Small, K. A., (1992), "Urban Transportation Economics," *Fundamental of Pure and Applied Economics* 51, 1992.
- Smeed, R. J. (1964), "Road Pricing: The Economic and Technical Possibilities.", Ministry of Transport, 1964.
- Smith M. J., M. B. Wiston, and M. O. Ghali (1994b), "Road-User Charging Systems Applied To The Cambridge Road Network", York Network Control Group; Working Paper Number 1. April 1994.
- Smith, M. J. (1983), "The existence and calculation of traffic equilibria.", *Transport. Res.* 17B, pp. 291-303.
- Smith, M. J. and M. O. Ghali, (1992), "Two new models for assessing urban traffic control and road pricing strategies." *Traffic Engineering and control*, 33, pp 245-249.
- Smith, M. J., A. D. May, M. B. Wisten, D. S. Milne and M. O. Ghali (1994a), "A comparison of the network effects of four road-user charging systems.", *Traffic Engineering and Control*, 35, pp 311-315.
- Smith, M. J., M. O. Ghali, and M. B. Wiston, A. D. May and D. Milne (1994c), "Road Pricing and Traffic Control With Application to York and Cambridge.", the 22nd European Transport Forum 1994, PTRC, Transportation Planning. Methods: Volume I, 12-16 September 1994.
- Snack, R. B. (1973), "The Restraint of Traffic: An area of possible compromise between the necessity for urban activity and the need to protect the environment.", Department of Town Planning, Oxford Polytechnic, April 1973.
- Starkie, David (1986), "Efficient and Politic Congestion Tolls", *Transpn. Res.*, Vol. 20A, No. 2, pp. 169-173, 1986.
- The Chartered Institute of Transport, (1990), "Paying for Progress", A report on congestion and road use charges, March 1990.
- Thomas Higgins (1981), "Road Pricing: a clash of analysis and politics", in *Policy Analysis*, No. 1 (Spring 1981), pp. 71-89.
- Vickrey W. S., (1969), "Congestion theory and transport investment", *American Economic Review*, No. 59, pp 251-261.
- Vickrey, W. (1973), "The pricing of Urban Transportation: Economic Desiderata, Technological Possibilities, and Political Constraints.", Presented at the 1973 Inter-Society Conference on Transportation, Denver.
- Vickrey, W. S. (1963), "Pricing in Urban and Suburban Transport", *Amer. Econ. Rev.* 53, pp. 452 - 465.
- Vythoulkas P. C. (1991), "Modelling Dynamic Stochastic User Equilibrium for Urban Road Networks", PhD Thesis, Centre for Logistics and Transportation, Cranfield University, 1991.
- Waersted, K. (1992), "Automatic Toll Ring no-stop electronic payment in Norway: Systems layout and full-scale experiences.", *Proc. 6th Int. Conf. on Road Traffic Monitoring and Control*, Conference Publication 355, IEE, London 1992.
- Walters, A. A. (1961), "The theory and measurement of private and social cost of highway congestion". *Econometrica* 29 (4), pp. 676-699.

- Walters, A. A. (1968), "Economics of Road User Charges", Johns Hopkins Press, Baltimore.
- Wardrop J. G. (1952), "Some theoretical aspects of road traffic research", Proceeding of the Institute of Civil Engineers, Part II, Vol. 1, pp 325-378.
- Watson P. L. and Holland E. P., (1978), "Relieving Traffic Congestion: The Singapore Area License Scheme", SWP281. The World Bank, Washington, DC.
- Whol, Martin (1970), "Congestion Toll Pricing for Public Transport Facilities.", Highway Research Record, No. 314, 1970.
- Wigan, M. R. and Bamford, T. J. G. (1973), "The effects of network structure on the benefits derivable from road pricing.", Department of Environment, TRRL Report, LR 557. Transport and road Research Laboratory, Crowthorne, England.
- World Bank (1975), Urban Transport Sector Policy Paper, World Bank, Washington, DC.
- Yoshii T. and Kuwahara M., (1993), "A Traffic Simulation Model for Oversaturated Traffic Flow on Urban Area", Institute of Industrial Science, University of Tokyo, Japan, "Unpublished" paper written in Japanese and translated by F. Katakura, Institute of Transportation Statistic, Tokyo, Japan.

Appendix (A)

Analytical Approximation for TDRP Model

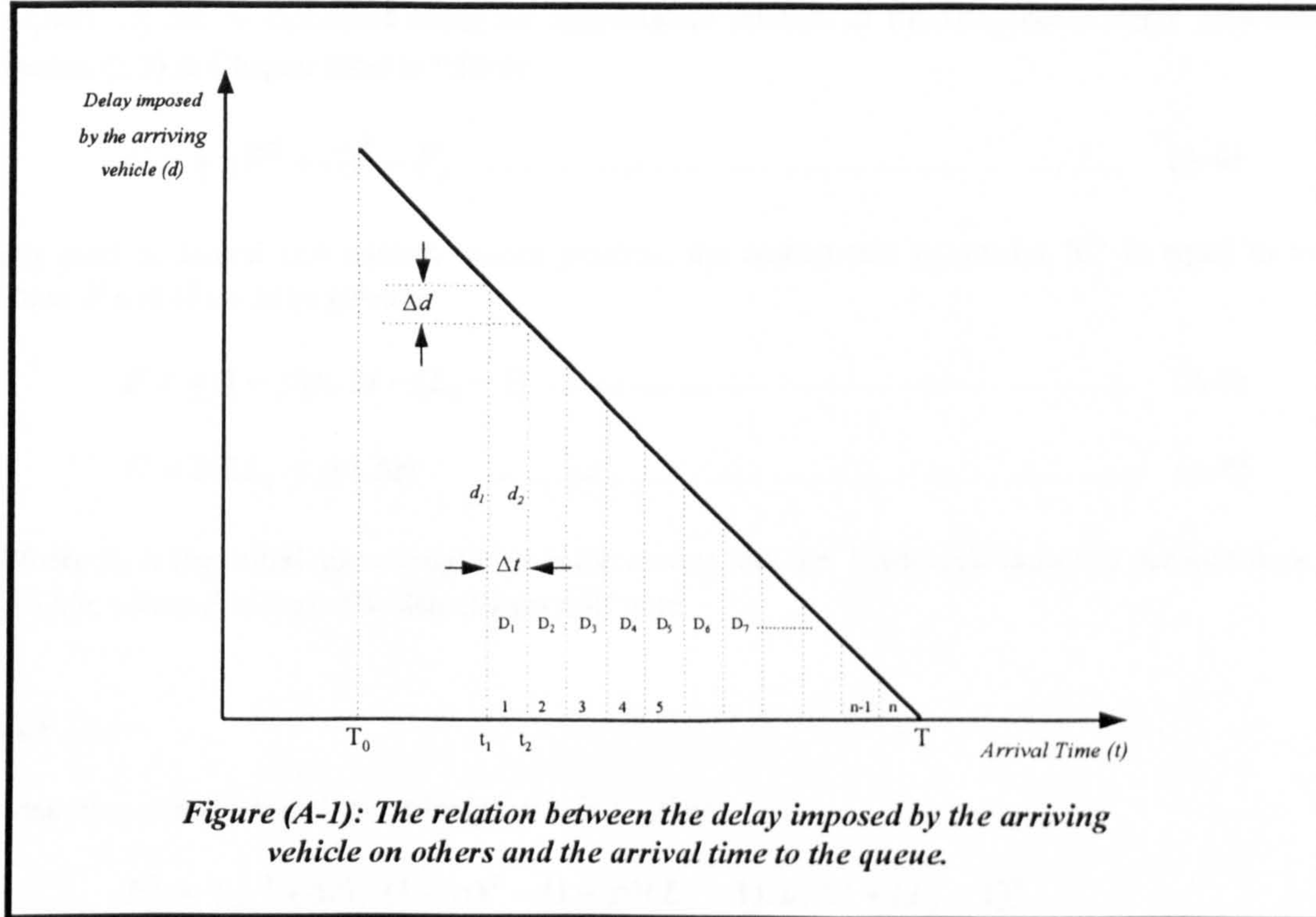


Figure (A-1): The relation between the delay imposed by the arriving vehicle on others and the arrival time to the queue.

Assuming that d_1 and d_2 are the delays imposed on other vehicles by two vehicles arriving at the tail of the queue at time t_1 and t_2 respectively, and Δd is the difference between d_1 and d_2 .

$$\Delta d = d_1 - d_2 \quad \dots\dots\dots (A-1)$$

d_1 and d_2 can be represented as follows:

$$d_1 = \Delta t \cdot [(D'_1 + D'_2 + D'_3 + \dots + D'_n) - (D_1 + D_2 + D_3 + \dots + D_n)] \quad \dots (A-2)$$

$$d_2 = \Delta t \cdot [(D''_2 + D''_3 + D''_4 + \dots + D''_n) - (D_2 + D_3 + D_4 + \dots + D_n)] \quad \dots (A-3)$$

where

$D_1, D_2, D_3, \dots, D_n$ are the total delay per unit time during intervals 1, 2, 3, ..., n.

$D'_1, D'_2, D'_3, \dots, D'_n$ are the total delay per unit time during intervals 1, 2, 3, ..., n after adding one more vehicle to the queue at time t_1 .

$D''_2, D''_3, D''_4, \dots, D''_n$ as above but for adding a vehicle at time t_2 .

Substituting in (A-1) by (A-2) and (A-3), yields

$$\Delta d = \Delta t \cdot [(D'_1 - D_1) + (D'_2 - D_2) + (D'_3 - D_3) + \dots + (D'_n - D_n)] \quad \dots (A-4)$$

Because Δd is very small, then adding one vehicle at t_1 will consequently increase the initial queue length at t_2 by one vehicle. Thus, $D'_2 = D''_2$, $D'_3 = D''_3$,, $D'_n = D''_n$. Hence equation (A-4) could be reduced to

$$\Delta d = \Delta t.(D'_1 - D_1) \dots\dots\dots (A-5)$$

D_1 and D'_1 can be calculated using the approximate solution of time-dependent delay presented in section (3.5) in Chapter three as follows:-

$$D = \frac{1}{2}((F^2 + G)^{\frac{1}{2}} - F) \dots\dots\dots (A-6)$$

For random arrival and random service patterns, the randomness parameter 'C' is equal to unity. Thus, F and G could be given by

$$F = \frac{1}{2}(1 - \rho)\mu.\Delta t - (L_0 - 1) \dots\dots\dots (A-7)$$

$$G = 2(2L_0 + \rho\mu.\Delta t) \dots\dots\dots (A-8)$$

Where L_0 is the initial queue length, μ is the service rate (i.e. bottleneck capacity) per unit time and $\rho = \lambda/\mu$, where λ is the traffic demand per unit time.

For D_1 :-

Assuming that the initial queue length at t_1 is L_0 , then

$$F_1^2 = \frac{1}{4}\mu^2.(\Delta t)^2.(1 - \rho)^2 - (1 - \rho)(L_0 - 1).\mu.\Delta t + (L_0 - 1)^2$$

As (Δt) and $(1 - \rho)$ are very small, then $(\Delta t)^2 \rightarrow 0$ and $\Delta t.(1 - \rho) \rightarrow 0$. Thus

$$F_1^2 \cong (L_0 - 1)^2$$

$$F_1 \cong (L_0 - 1) \dots\dots\dots (A-9)$$

Similarly

$$G_1 = 2(2L_0 + \rho\mu.\Delta t)$$

As Δt is very small, it could be assumed that $2L_0 \gg \rho\mu.\Delta t$. Thus, $(2L_0 + \rho\mu.\Delta t) \approx 2L_0$. Hence

$$G_1 \cong 4L_0 \dots\dots\dots (A-10)$$

Substituting by (A-9) and (A-10) in (A-6) yields

$$\begin{aligned} D_1 &\cong \frac{1}{2}(((L_0 - 1)^2 + 4L_0)^{\frac{1}{2}} - 1 + L_0) \\ &\cong \frac{1}{2}((L_0^2 + 2L_0 + 1)^{\frac{1}{2}} - 1 + L_0) \\ &\cong \frac{1}{2}((L_0 + 1) - 1 + L_0) \\ &\cong L_0 \dots\dots\dots (A-11) \end{aligned}$$

For D'_1 :-

The initial queue length in this case is $(L_0 + 1)$. Similar to the above approximations, F'_1 and G'_1 could be given as follows:

$$F'_1 = \frac{1}{2}(1 - \rho) \cdot \mu \cdot \Delta t - L_0$$

$$F'_1 \cong -L_0 \quad \dots\dots\dots (A-12)$$

$$G'_1 = 2(2L_0 + \rho\mu \cdot \Delta t)$$

$$G'_1 \cong 4(L_0 + 1) \quad \dots\dots\dots (A-13)$$

Then

$$D'_1 = \frac{1}{2}((F'_1)^2 + G'_1) - F'_1$$

$$D'_1 \cong \frac{1}{2}((L_0^2 + 4L_0 + 4)^{\frac{1}{2}} + L_0)$$

$$\cong \frac{1}{2}((L_0 + 2) + L_0)$$

$$\cong (L_0 + 1) \quad \dots\dots\dots (A-14)$$

Substituting by (A-11) and (A-14) in (A-5) gives

$$\Delta d = \Delta t \cdot (L_0 + 1 - L_0) = \Delta t \quad \dots\dots\dots (A-15)$$

Equation (A-15) implies that the difference between the delay imposed by the two vehicles arriving at t_1 and t_2 on other vehicles is only dependent on the time difference between t_1 and t_2 (i.e. Δt). In other words, the relation between the arrival time and the delay imposed by the arriving vehicles on others is given by a linear relation with a slope of -45° . This could be interpreted mathematically as in Equation (A-16) below.

$$d = \begin{cases} T - t & \dots\dots\dots t \in [T_0, T] \\ 0 & \dots\dots\dots otherwise \end{cases} \quad \dots\dots\dots (A-16)$$

Where d is the delay imposed on other vehicles by the vehicle arriving at time t , including the delay experienced by this vehicle. T_0 and T are the starting and vanishing times of the queue respectively.

Appendix (B)Network Traffic Data

64	NL = NUMBER OF LINKS
43	NN = NUMBER OF NODES
6	
38	
16	
13	
14	
15	
7	
31	
32	
29	
30	
8	
9	
10	
11	
12	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
33	
34	
35	
36	
37	
39	
1	
40	
2	
41	
3	
42	
4	
43	
5	NORDER(I)=NODE SEQUENCE
5	NORI = # OF ORIGIN NODES
5	NSEGDES = SEGMENT OF THE START OF DESTINATION NODES
3	NDES = # OF DESTINATION NODES
0	NSEGBUN = SEGMENT OF THE START OF BUNRYU NODES
10	INTASN = ASSIGNMENT INTERVAL (SEC)
10	IVEH = VEHICLES IN A PACKET
1	ISCAN = SCAN INTERVAL (SEC)
25200	ITT = TOTAL SIMULATION TIME (SEC)
600	IODCHG = TIME INTERVAL FOR OD DEMAND CHANGE (SEC)
-60 32	NODE ZAHYOU(IXN(I),IYN(I))
60 32	
-63 58	
0 110	
-35 110	
50 34	
-52 30	
20 55	
0 55	
0 37	
-3 32	
3 32	

32	34	1000.0	2000.0	0.50
31	7	1500.0	2000.0	1.00
30	27	1200.0	1600.0	0.50
24	29	1200.0	1600.0	0.40
32	11	1500.0	2200.0	0.60
12	32	1500.0	1600.0	0.50
10	13	1000.0	1400.0	0.60
13	12	1000.0	1600.0	0.65
11	9	350.0	2200.0	0.40
9	10	350.0	1400.0	0.35
9	8	1000.0	3200.0	1.00
9	20	550.0	1600.0	0.50
18	9	550.0	2200.0	0.60
19	17	1000.0	1600.0	0.40
37	18	1000.0	1600.0	0.35
22	19	500.0	2200.0	0.60
20	21	500.0	1400.0	0.30
21	24	1500.0	1600.0	0.35
25	23	1500.0	2000.0	0.50