MATHMATICAL ANALYSIS OF RESPONSIVE
PRIORIT FOR BUSES AT TRAFFIC SIGNALS

by

Benjamin Gorringe Heydecker

A thesis submitted to the University of London
for the degree of Doctor of Philosophy

Transport Studies Group
University College London

July 1983
For Betty and
the loving memory of
Walter
ABSTRACT

The purpose of this thesis is to investigate the consequences for all traffic of providing priority for buses. The range of priority methods which are currently available is reviewed and the method of responsive priority is selected for detailed analysis. The roles, both as factors in evaluation and as performance measures, of the traffic capacity of a junction and the mean delay incurred by each kind of vehicle in passing through it are identified.

A versatile model of a signal-controlled road junction is extended to incorporate a representation of responsive priority. This model is analysed to determine the capacity of each stream of traffic at a junction where responsive priority is provided for buses. An investigation is presented of the problem of finding signal-settings which, when implemented with responsive priority, give rise to a similar capacity for each stream to that arising when some given ones are implemented in the absence of priority. A formula is derived which is appropriate for estimating the mean delay incurred by non-priority vehicles. Separate methods are developed to estimate the mean delay incurred by priority vehicles. These methods are applied to the most usual forms of responsive priority so that the likely effects of their implementation can be investigated. Numerical results for two example junctions are given and discussed.
ACKNOWLEDGEMENTS

This thesis was produced with the benefit of the patient supervision of Professor Richard Allsop. Without his clear insights, many of the aspects considered here would have remained latent.

The author's colleagues, friends and relatives are to be thanked for their moral support and encouragement throughout the production of this thesis. In particular, thanks are due to David Holden, John Murchland, Jon Nicholl and Steve Gallivan for their helpful discussions. John Golias and Tasos Danas generated a most fruitful working atmosphere in the office which we shared. Steve Schwartz and Isabelle Henderson showed kindness and hospitality without which the final stages of production would have been considerably more difficult. Saskia Fry typed the bulk of the text and formulae skilfully and Fereidoon Ghahri-Saremi helped with the production of the graphs in chapter 6.
"It is a profoundly erroneous truism, repeated by all copybooks and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilisation advances by extending the number of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in a battle - they are strictly limited in number, they require fresh horses, and they must only be made at decisive moments."

A.N. Whitehead (1911,p41).
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>5</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>7</td>
</tr>
<tr>
<td>Table of contents</td>
<td>11</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>19</td>
</tr>
<tr>
<td>1.1 Traffic management schemes</td>
<td>19</td>
</tr>
<tr>
<td>1.2 Analytic methods in traffic management</td>
<td>20</td>
</tr>
<tr>
<td>1.3 Definition of terms</td>
<td>23</td>
</tr>
<tr>
<td>1.4 Discussion of the model</td>
<td>25</td>
</tr>
<tr>
<td>1.5 Conventions of notation and terminology</td>
<td>29</td>
</tr>
<tr>
<td>2. Priority for buses</td>
<td>33</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>33</td>
</tr>
<tr>
<td>2.2 Evaluation of priority schemes</td>
<td>34</td>
</tr>
<tr>
<td>2.3 Bus priority methods</td>
<td>43</td>
</tr>
<tr>
<td>2.3.1 Introduction</td>
<td>43</td>
</tr>
<tr>
<td>2.3.2 Passive priority methods</td>
<td>44</td>
</tr>
<tr>
<td>2.3.3 Active priority method</td>
<td>47</td>
</tr>
<tr>
<td>3. Responsive priority for buses</td>
<td>51</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>51</td>
</tr>
<tr>
<td>3.2 Rules to give priority when buses have right of way</td>
<td>51</td>
</tr>
<tr>
<td>3.3 Rules to give priority when buses do not have right of way</td>
<td>53</td>
</tr>
<tr>
<td>3.4 Rules to suspend priority</td>
<td>56</td>
</tr>
<tr>
<td>3.5 Rules to mitigate the adverse effects of priority</td>
<td>58</td>
</tr>
<tr>
<td>3.6 Scope of the study</td>
<td>60</td>
</tr>
</tbody>
</table>
4. Capacity

4.1 Introduction

4.2 The capacity of a stream of traffic
4.2.1 Introduction
4.2.2 Discussion
4.2.3 Traffic capacity in relation to signal timings
4.2.4 Discussion

4.3 Maximisation of capacity at a junction
4.3.1 Introduction
4.3.2 Allsop's formulation
4.3.3 Alternative formulations

4.4 Capacity at a junction where there is priority for buses
4.4.1 Introduction and preliminaries
4.4.2 Priority by extension
4.4.3 Priority by hurry-call
4.4.4 Priority by recall
4.4.5 The inhibition rule

4.5 Signal-settings at a junction where there is priority for buses
4.5.1 Introduction
4.5.2 A mathematical statement of the problem
4.5.3 Priority by extension
4.5.4 Priority by hurry-call
4.5.5 Priority by recall
4.5.6 The inhibition rule
5. Delay

5.1 Introduction

5.2 A delay formula for non-priority vehicles
   5.2.1 Introduction and preliminaries
   5.2.2 Review
   5.2.3 Mean delay
   5.2.4 Mean overflow
   5.2.5 Covariation of the overflow

5.3 Statistics of service
   5.3.1 Introduction
   5.3.2 Priority by extension and recall
   5.3.3 Priority by extension and recall with inhibition

5.4 Delay to priority vehicles
   5.4.1 Introduction
   5.4.2 A method to estimate delay in the absence of priority
   5.4.3 Priority by extension
   5.4.4 Priority by extension and recall
   5.4.5 Priority by extension and recall with inhibition

5.5 Practical considerations for the estimation of delay to non-priority vehicles
   5.5.1 Introduction
   5.5.2 Procedures to estimate delay
   5.5.3 Some possible simplifications
6. Worked examples
   6.1 Introduction

6.2 The junction at Derby
   6.2.1 Introduction
   6.2.2 Capacity-equivalent signal-settings
   6.2.3 Estimates of delay

6.3 The Chapel Hill junction
   6.3.1 Introduction
   6.3.2 Capacity-equivalent signal-settings
   6.3.3 Estimates of delay

6.4 Discussion
   6.4.1 Introduction
   6.4.2 Capacity-equivalent signal-settings
   6.4.3 Estimates of delay

7. Suggestions for further investigation

References

Appendix 1: Implicit functions
   A1.1 Introduction
   A1.2 Existence results
   A1.3 The Gauss-Seidel solution method

Appendix 2: An unexpected property of capacity-equivalent signal-settings when priority is provided by recall
   A2.1 Introduction
   A2.2 Analysis
   A2.3 An example
Appendix 3: Index of major notation

A3.1 Introduction 363
A3.2 Lists of notation 363
   A3.2.1 Introduction 363
   A3.2.2 Upper case Roman letters 364
   A3.2.3 Lower case Roman letters 366
   A3.2.4 Greek letters 369
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Data for the Derby junction</td>
<td>270</td>
</tr>
<tr>
<td>6.2</td>
<td>Priority by extension and recall: results for the Derby junction</td>
<td>284</td>
</tr>
<tr>
<td>6.3</td>
<td>Priority by extension and recall with inhibition: results for the Derby junction</td>
<td>286</td>
</tr>
<tr>
<td>6.4</td>
<td>Results for the Derby junction as the mean bus arrival rate tends to zero</td>
<td>288</td>
</tr>
<tr>
<td>6.5</td>
<td>Data for the Chapel Hill junction</td>
<td>302</td>
</tr>
<tr>
<td>6.6</td>
<td>Priority by extension and recall: results for the Chapel Hill junction</td>
<td>312</td>
</tr>
<tr>
<td>6.7</td>
<td>Priority by extension and recall with inhibition: results for the Chapel Hill junction</td>
<td>314</td>
</tr>
<tr>
<td>6.8</td>
<td>Results for the Chapel Hill junction as the mean bus arrival rate tends to zero</td>
<td>325</td>
</tr>
<tr>
<td>6.9</td>
<td>Values of the coefficients in Equation (6.1) for the Chapel Hill junction</td>
<td>326</td>
</tr>
<tr>
<td>6.10</td>
<td>Estimates of the mean delay incurred by priority vehicles when the signal-settings are not compensated</td>
<td>332</td>
</tr>
<tr>
<td>A2.1</td>
<td>Signal-settings used for the example in section A2.3</td>
<td>362</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>4.1</td>
<td>Stage transitions together with their associated durations</td>
<td>76</td>
</tr>
<tr>
<td>5.1</td>
<td>Flow chart for estimation of delay when there is priority for buses</td>
<td>258</td>
</tr>
<tr>
<td>5.2</td>
<td>Flow chart for calculation of statistics (5.84)</td>
<td>259</td>
</tr>
<tr>
<td>5.3</td>
<td>Possible classes of streams depending on the number of stages in the cycle</td>
<td>261</td>
</tr>
<tr>
<td>6.1</td>
<td>The Derby junction</td>
<td>268</td>
</tr>
<tr>
<td>6.2</td>
<td>Priority by extension: mean extension to stage ( k )</td>
<td>273</td>
</tr>
<tr>
<td>6.3</td>
<td>Priority by extension and hurry-call: capacity-equivalent signal-settings for the Derby junction</td>
<td>275</td>
</tr>
<tr>
<td>6.4</td>
<td>Priority by extension and recall: capacity-equivalent signal-settings for the Derby junction</td>
<td>277</td>
</tr>
<tr>
<td>6.5</td>
<td>Priority by extension and hurry-call with inhibition: capacity-equivalent signal-settings for the Derby junction</td>
<td>279</td>
</tr>
<tr>
<td>6.6</td>
<td>Priority by extension and recall with inhibition: capacity-equivalent signal-settings for the Derby junction</td>
<td>281</td>
</tr>
<tr>
<td>6.7</td>
<td>Priority by extension and recall: Delay estimates for the Derby junction with fixed signal-settings</td>
<td>289</td>
</tr>
<tr>
<td>6.8</td>
<td>Priority by extension and recall with inhibition: Delay estimates for the Derby junction with fixed signal-settings</td>
<td>291</td>
</tr>
<tr>
<td>6.9</td>
<td>Priority by extension and recall: Delay estimates for the Derby junction with capacity-equivalent signal-settings</td>
<td>293</td>
</tr>
<tr>
<td>6.10</td>
<td>Priority by extension and recall with inhibition: Delay estimates for the Derby junction with capacity-equivalent signal-settings</td>
<td>295</td>
</tr>
<tr>
<td>6.11</td>
<td>Minimum ratio of priority to non priority vehicle occupancy to give a net reduction in passenger delay. Derby junction</td>
<td>297</td>
</tr>
<tr>
<td>6.12</td>
<td>The Chapel Hill junction</td>
<td>300</td>
</tr>
<tr>
<td>6.13</td>
<td>Priority by extension and hurry-call: capacity-equivalent signal-settings for the Chapel Hill junction</td>
<td>303</td>
</tr>
<tr>
<td>6.14</td>
<td>Priority by extension and recall: capacity-equivalent signal-settings for the Chapel Hill junction</td>
<td>305</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.15</td>
<td>Priority by extension and hurry-call with inhibition: capacity-equivalent signal-settings for the Chapel Hill junction</td>
<td>307</td>
</tr>
<tr>
<td>6.16</td>
<td>Priority by extension and recall with inhibition: capacity-equivalent signal-settings for the Chapel Hill junction</td>
<td>309</td>
</tr>
<tr>
<td>6.17</td>
<td>Priority by extension and recall: delay estimates for the Chapel Hill junction with fixed signal-settings</td>
<td>315</td>
</tr>
<tr>
<td>6.18</td>
<td>Priority by extension and recall with inhibition: delay estimates for the Chapel Hill junction with fixed signal-settings</td>
<td>317</td>
</tr>
<tr>
<td>6.19</td>
<td>Priority by extension and recall: Delay estimates for the Chapel Hill junction with capacity-equivalent signal-settings</td>
<td>319</td>
</tr>
<tr>
<td>6.20</td>
<td>Priority by extension and recall with inhibition: Delay estimates for the Chapel Hill junction with capacity-equivalent signal-settings</td>
<td>321</td>
</tr>
<tr>
<td>6.21</td>
<td>Minimum ratio of priority to non priority vehicle occupancy to give a net reduction in passenger delay. Chapel Hill junction</td>
<td>323</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

1.1 Traffic management schemes

Most human activities depend to some extent on the movement of people or goods from place to place. Indeed, transport facilities provide personal access for many purposes, including work, education, shopping, leisure and personal business. Thus the extent and quality of the transport facilities available has a direct effect on the economic, social and physical conditions of people's lives. Further, because of environmental pollution, both chemical and acoustic, accident risks and competition for limited resources such as capital, space in urban areas and petroleum stocks, many indirect effects of the construction, use and maintenance of transport facilities accrue to the whole community. The object of traffic management schemes is then to make good use of those facilities which are available and to allocate equitably costs and benefits, in a general sense, between travellers and non-travellers and between different categories of travellers.

In accordance with this, in past years the emphasis of traffic management schemes has changed from assisting vehicles to considering directly the effects on all members of the community, including pedestrians and others who are not using vehicular transport (Webster, 1972; Turner, 1973). In the particular case of buses, while each bus makes a greater demand on resources than does a single car, a bus normally carries a sufficient number of passengers for the demand per passenger to be less than for car passengers (Smeed and Wardrop, 1964; Webster, 1968; Danas, 1981). Bus priority schemes aim to exploit this to the benefit of the community as a whole.
1.2 Analytical methods in traffic management

While there is general agreement that traffic management schemes can be of benefit to the community in many circumstances, the problem of identifying the most appropriate scheme in each instance is not trivial. There are three aspects to this problem. Firstly, there may be many candidate schemes, including the option of making no change, between which a choice must be made. Secondly, there is no single evaluation criterion which is appropriate to all circumstances: some possible choices are discussed in Chapter 2. Finally, each of the components of any evaluation criterion to be used must be quantified. Once an evaluation criterion has been selected, if methods are available to quantify the effects of each candidate scheme, then a rational choice can be made between them.

Several approaches to the investigation of the consequences of traffic management schemes are available. Certainly the most direct of these is to install the scheme, or at least a small-scale pilot scheme, and then to monitor the effects. This monitoring process is an important part of the implementation of any scheme, especially those of which there is little practical experience. Indeed, the ultimate test of any scheme is its performance when implemented: a scheme which appears to perform badly is unlikely to be considered satisfactory. However, because of the costs of installation and observation, this method is not normally suitable for choosing between alternative schemes, especially when more than a few are available. Furthermore, in cases where variables are at the disposal of the traffic manager, any attempt to determine the most suitable values for these variables by implementing different ones and then comparing the consequent results is unlikely to succeed because of the large number of experiments which may be required and the possibility of spurious fluctuations dominating the results (Nicholl, 1981).
Simulation methods which incorporate representations of individual vehicles can provide powerful tools for these investigations: examples of flexible models for this are due to Needham (1970), Gipps (1976) and Logie (1979). A major advantage of these microscopic simulation methods is that once a model has been postulated for each of the components of the system under investigation, no further analysis is necessary: thus the models incorporated can be very sophisticated. However, some preliminary analysis is usually possible and indeed is desirable in order to reduce the computational burden of the simulation. A further advantage is that because of the way in which they mimic real processes, a considerable amount of information can be gleaned from microscopic simulations in a manner analogous to that of field observations but without the same risk of corruption by extraneous fluctuations.

However, these simulation methods are not without drawbacks. Although they can yield adequate numerical results, they give little information about the mechanism of the dependence of these results on the independent variables. Thus they are little better suited to the determination of the most appropriate values for variables than are field observations. Some particular problems arise in the interpretation of results. In general, there will be an initial period in each simulation run during which the model cannot be considered to represent accurately the system under study. Results from this period must be disregarded in the analysis of the scheme. The durations of this initial period and of the entire simulation run as well as the number of replications to be performed all require careful selection because of the inherent trade-off between the accuracy of the results and the computational effort required. These problems have been discussed by several authors (see, for example, Gordon, 1969).

By contrast with these methods, the principal advantage of analytical ones is the scope for further investigation. A simple but nevertheless
important example of this is the determination of the dependence of a quantity of interest on the independent variables. In cases where a single evaluation criterion has been selected, the candidate schemes should be compared after this has been optimised in each scheme over all the variables which are at the disposal of the traffic manager. If analytic estimates are available for all the components of the evaluation criterion, then powerful optimisation techniques can be invoked to reduce the computational effort required (see, for example, Vajda, 1961; Whittle, 1971). In many of these cases, further analysis can determine the sensitivity of the optimal value of the evaluation criterion to inaccuracies in the data used and variations in the values taken by parameters. Examples of this are discussed in Chapters 4 and 5.

The chief disadvantage of analytical methods is that in many cases an exact treatment can be intractable. In these cases, if either approximate methods are used or an exact solution is obtained to a similar but more straightforward problem, then the approximate solutions can be compared to results from a microscopic simulation or numerical solutions to the original problem. These comparisons can be used in the absence of analytical error bounds to show variously that simplifications to a problem (e.g. Ohno, 1978), approximations made within an analysis (e.g. Golias, 1981), or functional approximations to numerical solutions (e.g. Miller, 1969) introduce only small errors within the range of variables tested. Alternatively, formulae which are significantly in error due to their heuristic derivation can be modified to give better results (e.g. Webster, 1958).

Methods from each of these three primary categories of field observations, microscopic simulation and mathematical analysis can be used in conjunction and to complement each other in a satisfactory manner. Each approach has its own characteristics and none on its own is sufficient for
all purposes. In this thesis, literature on the application of each of these approaches to responsive priority for buses is reviewed. The literature on mathematical analyses of capacity and delay at signal-controlled junctions is also reviewed and established methods are developed to apply to junctions where responsive priority is implemented. Several optimisation problems relating to traffic signals are discussed and the problem of calculating signal-settings which are suitable for use at a junction where responsive priority is provided for buses is investigated in detail. Thus the contribution of this thesis lies in the last of the three categories of methods introduced above.

1.3 Definition of terms

There are no universally accepted definitions for terms to describe the features of a signal-controlled road junction — indeed, some terms are used by different authors in mutually incompatible ways. In this section a model of signal-controlled junctions is introduced and definitions are given for the terms which are used to describe it. These terms are used throughout this thesis in the sense given here.

As drivers approach a junction, some of them may have a free choice between lanes which are adjacent and are controlled in an identical manner. Each driver who has such a choice is supposed to select the lane which permits him to depart from the junction as soon as possible. In such cases vehicles in more than one lane will, in effect, form a single queue, although they do not necessarily all leave the junction in the same direction. The traffic using a set of lanes between which some drivers divide their number in this manner is called a stream. Each lane that does not belong to one of these sets is used by drivers performing one or more manoeuvres and, either necessarily or by choice, queueing separately from those using other lanes. All the traffic using such a lane is also called a stream.
At signal-controlled junctions, each stream of traffic is subject to alternating periods when vehicles in it may enter the junction and periods when they may not. A stream has right of way when vehicles in it may enter the junction; at signal-controlled junctions this occurs only when that stream receives a green indication. Two streams are compatible if they may safely have right of way simultaneously.

At the beginning of a period during which a stream has right of way, the rate at which vehicles depart rises to a value which, when allowance is made for differences between the size and performance of vehicles, is supposed to have a constant mean while some vehicles are still queueing. This is called the saturation departure rate and is expressed in terms of passenger car units (pcu's), each vehicle being considered equivalent, in terms of the time it requires to enter a junction, to some number of average passenger cars. The first few vehicles depart at less than the saturation rate because they are still accelerating as they enter the junction: to compensate for this, vehicles are supposed to depart at the saturation departure rate, starting from some instant after the green indication is first received. Vehicles may continue to enter the junction for a short time after the green indication is removed. To compensate for this, each stream is supposed to retain right of way and queueing vehicles to depart at the saturation departure rate for an interval after the green indication is removed. The period during which the stream is supposed to have right of way is called the effective green period. The period between two successive effective green periods is called the effective red period.

At signal-controlled junctions, collections of mutually compatible streams are granted right of way in turn. The period during which all streams in such a collection receive a green indication is called a stage.
The duration of each stage is subject to a non-negative minimum constraint, called the minimum green time, which often arises for reasons of safety. Between two successive stages occurs a transition period, during which the signal indications change. A stage ends, and a transition period starts, when the next stage is called. Once a stage has been called, it must start and run for its minimum green time.

At junctions where stages occur cyclically, the collection of stages, each numbered according to the order of occurrence starting from some arbitrary stage, is called the sequence. The period between the start of two successive sequences is called a cycle. If the duration of each stage in a sequence is fixed, then the traffic signals are fixed-time. If the durations are varied in a manner which depends on the detection of vehicles in various streams at the junction, then the traffic signals are vehicle-actuated.

Three terms are now defined which are used to denote quantities which are the subject of a large part of the investigations presented in this thesis. The capacity of a stream is the greatest mean rate, measured in pov's, at which traffic can enter a junction from that stream. The delay caused to a vehicle is the difference between its travel time and the time it would have taken if no other vehicles were present and the stream of traffic in which the vehicle travels had constant right of way. The overflow in a stream is the number of vehicles remaining in the queue at the end of an effective green period.

1.4 Discussion of the model

The model of a signal-controlled junction described in Section 1.3 is flexible, but is neither the only one available nor universally applicable. Furthermore, some details have not been specified. A thorough review of
alternative models which are found in the literature has been given by Allsop (1972a), together with accounts of their development by various authors. Literature on these topics which is relevant to this thesis is reviewed in Chapters 4 and 5. In this section a brief discussion is given of the restrictions imposed implicitly by the definitions given in Section 1.3 and the possibilities which are available for the remaining details.

The definition of a stream of traffic depends on the rather simplistic behavioural assumption that each driver who has a choice between lanes identifies and chooses the one which minimises his own delay at the junction. This assumption ensures that the traffic in each stream forms a single queue.

The departure model, which stems from observations by Greenshields, Schapiro and Ericksen (1947), depends heavily on the concept of the saturation departure rate. Methods for estimating this, the relationship between the displayed and the effective green times and the passenger car equivalents of vehicles at a signal-controlled junction from observations made there have been described by the Road Research Laboratory (1963) and by Branston and Van Zuylen (1978). Some observations (see, for example, Teply, 1981) show that the mean rate at which vehicles enter a junction during a long effective green period in which the queue persists increases to a maximum value and then decreases gradually. In these cases, the difference between the displayed and the effective green times may depend on the duration of the displayed green.

Several variable factors have been observed to affect the saturation departure rate of each stream. These include the time of day, the weather conditions (Branston, 1979; Teply, 1981) and the proportion of vehicles in the stream which perform various manoeuvres at the junction (Webster and Cobbe, 1966). An important example of the last of these factors is that of
vehicles making opposed turns which are permitted to proceed through gaps which occur in the opposing traffic. In many such cases these vehicles depart at different rates during different stages or parts of each stage, so several saturation departure rates are required for a complete representation, each to be used at different times (Yagar, 1974; Allsop, 1977a).

The way in which vehicles depart from the queue during an effective green period has not been specified fully. A commonly used model (Clayton, 1940; Haight, 1959; Miller, 1963; McNeil, 1968) is adopted here which supposes that while the queue persists, vehicles depart at intervals which are determined by the saturation departure rate and the type of vehicle. If the queue persists throughout the effective green period and the end of this period is not coincident with the end of the departure time of a vehicle, then one further vehicle is supposed to depart in the last part of this period with a probability proportional to its duration. In the long run, the same number of pcv's will depart from the queue with this model as if the queue is diminished by that part of a vehicle which corresponds to the probability that the whole vehicle departs. A model in which part of a vehicle departs in every such case has been investigated by Newell (1965). This represents the queue as a continuous quantity which is diminished at the saturation departure rate throughout the effective green period.

There remains the question of the behaviour of vehicles which arrive after the queue has dissipated but before the end of the effective green period. Haight (1959), Darroch (1964) and Ohno (1978) have developed a model in which any vehicles that arrive during this interval are not delayed, irrespective of how frequently they arrive. With this model an unlimited number of vehicles can enter the junction from a stream in an effective green period during which the queue dissipates. In order that the capacity of a stream is not unlimited because of this, in this thesis
vehicular arrivals are supposed not to be strongly correlated with the state of the traffic signals.

In some circumstances a significant advantage can be gained by supposing that the mean rate at which vehicles discharge from the queue during the effective green period cannot exceed the saturation departure rate. This assumption has been used by Newell (1960), McNeil (1968) and Griffiths (1981) amongst others in the estimation of the mean overflow. For this purpose, the departure model with this assumption can be likened to a queue where service occurs instantaneously at the end of the effective green period: the capacity of the service is determined from the duration of the effective green period and is compared to the number of pcu's in the queue at that time. Because of this likeness, this will be referred to as the bulk-service model.

The definition given for stages ensures that a stage starts when the last stream or streams gain their green indication and ends when the first stream or streams lose theirs. Thus if the starts and ends of the green indications for different streams do not all occur together but are staggered, then some streams receive green indications for part of the appropriate transition period. Furthermore, if a stream receives a green indication in two consecutive stages, it will normally also receive a green indication throughout the transition period between them.

The duration of each transition period is normally stipulated by the safety requirement of allowing adequate time for vehicles in each stream which loses right of way to clear the junction before right of way is gained by any stream with which it is not compatible. In this thesis the duration of each transition period is supposed to be fixed. By contrast, the durations of the stages are supposed to be at the disposal of the traffic engineer, subject to various constraints.
Tillotson (1981) has discussed the problem of finding a satisfactory definition of delay and of relating it to quantities which can be measured. The definition used here is essentially the same as that of Webster (1958), who compared the actual journey time between a notional point on each side of the junction, beyond which the influence of the junction is supposed not to extend, with the unimpeded journey time between the same two points. As Tillotson remarked, delay is an artefact and cannot be related directly to the welfare of travellers.

1.5 Conventions of notation and terminology

This is a convenient point at which to introduce several conventions of notation and terminology. These conventions are used throughout this thesis without further notice: their usage is explained here for the sake of clarity.

This thesis contains 7 chapters, each of which is divided into sections. Some sections are divided further into sub-sections. Each section is numbered according to the chapter which contains it and the position it occupies within that chapter. Thus Section 4.2 is the second section of Chapter 4. This system of numbering is extended further for sub-sections so that sub-section 4.2.3 is the third sub-section of Section 4.2.

Each result (lemma, theorem or corollary) is numbered according to the chapter which contains it and the position it occupies within that chapter. Thus lemma 4.26 preceeds corollary 4.27 immediately and is the twenty-sixth result given in Chapter 4. The end of the proof of a result is indicated by the symbol \[\]
: in the few cases where no proof is offered here, the same symbol is used to indicate the end of the statement of the result. Marginal numbers are assigned to various expressions (equations, inequalities or formal statements of problems) when reference is made to them later on in
the thesis. A similar numbering system is used for these as is used for results. The two numbering systems are independent, so equation 4.26 is not necessarily associated with lemma 4.26.

Various Greek and Roman letters are used to represent particular quantities. These are introduced where they are first used in the text; a general list of the notation used is given in Appendix 3. Greek letters are used for dimensionless quantities: where these involve measures of time, the unit used is the duration of a cycle of the signals.

In some cases the value taken by one quantity will vary according to the value taken by one or more others. When this occurs, the function which relates the variables involved is denoted by the symbol for the former variable together with the tilde followed by parentheses enclosing the variables upon which the function depends. Thus \( \epsilon \) represents a quantity whereas \( \tilde{\epsilon}(\beta, \gamma) \) represents the function which maps the values of \( \beta \) and \( \gamma \) to the value of \( \epsilon \). Tildes are used more widely to indicate that a symbol represents a function of the quantities in the parentheses which follow. Some functions and operators are written without tildes where this is unlikely to cause confusion: principal examples of this are the probability generating functions \( \Psi(z) \), \( \phi(z) \) and \( \Phi(z,t) \), the minimum and maximum functions (abbreviated to Min and Max respectively) and the probability operators discussed next.

The probability notation adopted here is standard. The symbol \( P(A) \) is used to denote the probability that the event \( A \) occurs and \( P(A|B) \) to denote the conditional probability that event \( A \) occurs given that the event \( B \) does. The mean (or expected) value of a random variable \( X \) is denoted by \( E(X) \) and the variance by \( \text{Var}(X) \) where \( \text{Var}(X) = E(X^2) - [E(X)]^2 \). The covariance of two random variables \( X \) and \( Y \) is written as \( \text{Cov}(X,Y) \) and is given by \( \text{Cov}(X,Y) = E(XY) - E(X)E(Y) \).
The quantity $E(X^n)$ is referred to as the $n$th moment of the distribution of $X$. The conditional moments of a random variable $X$ given that an event $B$ occurs are written as $E(X^n|B)$.

The set of all variables which differ only in the values of their subscripts is denoted by a typical element enclosed in braces. Thus \{\lambda_1,\lambda_2,\ldots,\lambda_m\} is abbreviated to $\{\lambda_j\}$ where this results in no loss of clarity. The symbol $\Sigma$ is used to denote summation over the range which is indicated for an indexing variable. If this range is null, then the value of the sum is taken to be identically equal to 0. Thus, for any $\{\lambda_i\}$, if $n_1 > n_2$, then $\sum_{i=n_1}^{n_2} \lambda_i = 0$. This convention is used extensively and avoids the need to consider cases such as $n_1 > n_2$ and $n_1 \leq n_2$ separately in the above example.

Some set-theoretic notation is used as follows. The form $\{\lambda_i | 1 \leq i \leq n\}$ is used for the set of all $\lambda_i$ for which the value of the sub-script $i$ lies in the range 1 to $n$. The set-inclusion symbol $\subset$ is used: thus if $A \subset B$ then every element of the set $A$ is also an element of the set $B$. Finally, the elemental symbol $\epsilon$ is used: thus if $a \in A$ then $a$ is an element of the set $A$.

Mathematical programming problems are stated in forms such as

Minimise $\tilde{F}(\mu)$
\[ \mu, \lambda \]

Subject to
\[ \mu - \lambda \geq 0 \]
and $\lambda \geq T/c$

This particular example can be read as "find the minimum value attained by the function $\tilde{F}(\mu)$ in the set of all values of $\mu$ and $\lambda$ for which $\mu - \lambda$ is positive and $\lambda$ is at least as large as $T/c$.\]
CHAPTER 2
PRIORITY FOR BUSES

2.1 Introduction

The objectives of providing priority for buses are to improve the quality and image of the service provided on the grounds that this can benefit the community as a whole. The range of benefits that can arise from such improvements has been summarised by Finnamore and Jackson (1978).

"Passengers can benefit from greater reliability. Crews can benefit from a greater likelihood of reaching relief points on time, a more even spread of the workload, and greater job satisfaction resulting from the provision of a better service. Finally the bus undertaking itself can benefit from increased patronage attracted by a better service and better use of resources leading to more economical working."

Methods to provide numerical estimates of some of these benefits are mentioned in the next section together with a brief discussion of their interpretation.

Against any benefits must be weighed the cost of implementing and maintaining the priority system and any disbenefits caused to other members of the community. Methods to provide numerical estimates of some of these quantities are also mentioned. However, a full economic evaluation of bus priority schemes is beyond the scope of this thesis.

Many methods are available to the traffic manager to provide priority for buses. These can be used individually or in combination and in
isolation or as part of wider traffic management schemes. Comprehensive surveys of these methods have been published by the Transport and Road Research Laboratory (Department of the Environment, 1973a; Webster and Bly, 1976) and guidelines on the practical aspects of their implementation have been given by the Department of Transport (1976). A survey is given in Section 2.3 of priority methods which are appropriate to signal-controlled junctions together with indications of how the consequences of their implementation can be investigated. A detailed description of the responsive priority method which is the main subject of this thesis is given in Chapter 3.

Urban bus services can be adversely affected by traffic congestion, particularly during peak periods. The scope for improvement by implementing priority at signal-controlled junctions is indicated by the results of two studies: Buchanan and Coombe (1973) and Shields (1976) found that during busy periods over 20 per cent of the total journey time of buses on urban routes was stopped time at traffic signals. Further analysis of results from the latter of these studies, reported by the Greater London Council and London Transport Executive (1976), showed that there is considerable variation in this stopped time according to the time of day. Some consequences of these variations for the quality and costs of bus services are discussed in the next section.

2.2 Evaluation of priority schemes

Many different criteria are available to determine whether or not a priority scheme can be considered to be beneficial. In this section, some possible components of these criteria are introduced and discussed. However, no attempt is made to identify a single criterion to represent a measure of total benefit. Rather, emphasis is placed on the range of criteria available. Chapman (1976), in a review of 16 different reliability
indices for bus services, noted that some priority schemes may appear to be beneficial or not depending on which evaluation is criterion adopted.

A major consideration in any assessment of the consequences of a priority scheme is the change in the mean delay caused to buses (see, for example, Michalopoulos, 1976; MacGowan and Fullerton, 1979; Department of Transport, 1980), while any reduction in the standard deviation of this delay is often considered to be of similar importance (Department of the Environment, 1973b; Richardson and Ogden, 1979; Cooper, Vincent and Wood, 1980; Cottinet, Amy de la Breteque, Henry and Gabard, 1980).

Changes in the mean delay to buses can be compared to changes in mean delay to other vehicles and, by making allowance for the numbers and occupancy of vehicles in each class, the total delay caused to travellers at a junction can be calculated. This has been used as a measure of community disbenefit by Evans and Skiles (1970), Tarnoff (1975), Seward and Taube (1977), and Gallivan, Young and Peirce (1980). Further allowance can be made for the differing costs of delay to different kinds of travellers and the differing operating costs of vehicles, and thus a monetary value can be associated with any changes in delay (Jenkins and Moseley, 1981).

These methods are incomplete as evaluation criteria in several respects. The effects of changes in variability of delay on travellers are ignored. Further, these methods assume that the cost associated with an incremental delay is independent of the delay already experienced and that the values of any time savings are additive – for example, that a saving of 1 second to each of 10 passengers is equivalent to a saving of 10 seconds to 1 passenger.

Variability in journey times can be of considerable importance to a traveller who wishes to complete his journey before some deadline. Because of travellers who allocate extra time for a journey for this reason, the
total time attributable to travelling will exceed the mean journey time. Richardson and Ogden (1979) suggested that a suitable estimate for the time budgeted for a journey is the sum of the mean and the standard deviation of the journey time. Similar but more detailed analyses are given by Buchanan and Coombe (1973) and Chapman (1976). Knight (1974) reported a more general investigation of the consequences of allocating different values of time to the periods before the start of a journey, after the journey is completed but before the deadline and during any part of the journey which overruns the deadline. This contrasts with methods which suppose that an early arrival is of no benefit to the traveller.

Shields (1976) and Horowitz (1981) have shown that passengers' subjective value for time spent waiting at a bus stop is roughly twice as high as for the same amount of time spent travelling in a bus. Shields (1976), Coe and Jackson (1977) and Gault and Doherty (1979) also found that disproportionately large values were associated with unusually long waits at a bus stop. Richardson and Ogden (1979) and Koppelman (1981) argued that the utility of a change in journey time should be taken as a more general function of the size of the change than a linear multiple. Richardson and Ogden proposed a utility function which is quadratic for small time changes and linear for larger ones. Koppelman found that travellers choose between modes of transport as if their utility function for time is the elapsed time raised to the power 2.40. Both of these utility functions associate a greater value with a single time change to one traveller than with several changes which sum to this but accrue to different travellers.

Richardson (1980) showed that because they normally save large amounts of time for bus passengers at the expense of smaller increases to comparatively many other travellers, bus priority schemes are favoured by the use of utility functions like the ones described above. Furthermore, if the savings to bus passengers due to several priority implementations are
added, then their benefits are enhanced considerably by this method of evaluation. Frith (1981) has argued that the time losses caused to other travellers will tend to accumulate in a similar manner so the net benefit of a priority scheme is less than it would be if each such traveller experienced no more than one additional delay. A more general discussion of the merits of various methods to evaluate any time changes which arise from traffic management schemes has been given by Hensher (1976).

The variability in delay experienced by buses at each junction is of some interest in itself. However, since any deviations from schedule tend to be magnified as buses pass along a route, the consequent effects on both headways and running times of buses are of great interest. Welding (1957) discussed the inherent instability of a bus service with regular headways. Newell and Potts (1964) analysed the way in which departures from a pattern of regular headways are magnified along a bus route and found that a quantity of prime importance in this process is the ratio of the mean passenger arrival rate to the mean passenger boarding rate. As this ratio increases, so does the degree to which deviations from schedule of a bus are amplified between successive stops. This process was shown to cause buses to form pairs as they proceed along a route. Potts and Tamlin (1964) found that bunches of buses can form in a manner which is consistent with this model. Other observations by Bly and Jackson (1974), Chapman, Gault and Jenkins (1976) and Shields (1976) have shown how the variability of bus headways grows along a route. Some consequences of this process and its possible remedy by the introduction of timing points along the route have been discussed by Huddart (1973). Golshani and Thomas (1981) found from simulation studies that the regularity of buses is likely to be improved most by using as many timing points as possible and allocating more slack time to those near the start of a route than to those near the end.
In view of the relative importance to passengers of the length of time spent waiting for a bus, changes in this may form a large part of the effects of a priority scheme. Coe and Jackson (1977) discussed the difficulty of finding reliable methods to relate the waiting times of passengers to the distribution of bus headways. A formula which can be used to estimate the mean waiting time under fairly weak hypotheses has been given by Cox and Smith (1954). Provided that
(a) arrivals of passengers are independent of arrivals of buses;
and
(b) each passenger boards the next bus to leave the stop after his arrival,
then
\[
\bar{w} = \left(1 + \frac{s^2}{\bar{h}^2}\right)\frac{\bar{h}}{2} \quad (2.1)
\]

where
\[
\bar{w} \quad \text{is the mean waiting time of passengers},
\]
\[
\bar{h} \quad \text{is the mean bus headway},
\]
and \(s\) is the standard deviation of bus headways.

Welding (1957) supposed that a sample of headways \(\{h_i\}_{i=1}^n\) was available, and gave the corresponding formula

\[
\bar{w} = \frac{\sum_{i=1}^n h_i^2}{2 \sum_{i=1}^n h_i} \quad (2.2)
\]

According to these formulae, for any given mean headway the mean waiting time of passengers takes its minimum value of \(\bar{h}/2\) when \(s=0\), in
other words when bus arrivals are regular. The mean waiting time increases linearly with the variance of the bus headways and, as would be expected, is equal to \( \bar{h}^2 \) if the bus arrivals form a Poisson process since in that case \( \bar{s}^2 = \bar{h}^2 \).

Hutchinson and Nicholl (1978) discussed errors which arise in the normal methods used to estimate \( \bar{w} \) from observations of a bus service. In particular, they cast some doubt on the accuracy of the estimates made by previous authors, particularly in cases where \( \bar{h} \) is large and the period of observation is not correspondingly long.

Hollroyd and Scrags (1966) made observations of bus headways in Central London in an attempt to find a simple relationship between \( \bar{w} \) and \( \bar{h} \). Noting that \( \bar{s}^2 = \bar{h}^2 \) when \( \bar{h} \leq 2 \) minutes, they used the empirical function

\[
\bar{s}(\bar{h}) = \left( \frac{\bar{h}^2}{A + \bar{h}^2} \right)^{1/2}
\]

(2.3)

to estimate \( \bar{s}^2 \). This has the property that \( \bar{s} = \bar{h} \) for \( \bar{h} \ll A \) and \( \bar{s}^2 \approx A \) for \( \bar{h} \gg A \). Use of the value \( A = 35 \) (minutes) in (2.3) was found to give good agreement between (2.1) and (2.2) when applied to the data they collected.

O'Flaherty and Mangan (1970) made similar observations during off-peak periods in Leeds and Harrogate and found that appropriate values for \( A \) to be 15 and 20 (minutes) respectively. Nicholl (1977) repeated the survey of Hollroyd and Scrags in central London and found that the value of 158 (minutes) for \( A \) gave the best fit, indicating an increase in variability of bus headways and thus a decrease in the quality of the service offered. However, none of these authors compared the estimates of \( \bar{s}^2 \) given by (2.3) with their observations.
Seddon and Day (1974) showed that for mean bus headways of over 10 minutes, the arrivals of passengers at bus stops in Manchester did not appear to form a Poisson process. Mean waiting times for passengers boarding buses with mean headways of less than 10 minutes were similar to, but generally about 10 per cent less than those given by (2.2). For services with longer mean headways, the mean waiting times were considerably less than those given by (2.2).

Danas (1980) reported observations made at two bus stops in London during peak periods. He found that in cases where the mean bus headway was less than 7 minutes, the estimates given by (2.2) were between 10 and 20 per cent higher than the mean time between the arrival of a passenger and the departure of the next bus. The mean waiting time during peak periods in Central Leeds for buses with mean headways of more than 6 minutes were found by O'Flaherty and Mangan (1970) to be less than half the mean headway. These results suggest that assumption (a) that passengers arrive independently of buses is not born out in practice, particularly when they use infrequent services during peak periods.

Jolliffe and Hutchinson (1975) proposed a model of passenger arrivals which incorporates two distinct methods by which mean waiting times can be reduced. Firstly, some passengers arrive at the same time as the bus because they run to the stop or curtail some interim activity when they see the bus approaching. This behaviour has been observed in practice by Bly and Jackson (1974) and Danas (1980) and can have a considerable effect on the mean waiting time since these passengers avoid comparatively large delays by their action. Secondly, noting from their observations of 10 bus stops in London that passenger arrivals tend to increase at times when the expected waiting time is small, Jolliffe and Hutchinson supposed that some passengers who were familiar with the service use their experience to choose times at which to arrive in order to reduce their expected waiting
times. The remainder of the passengers were supposed to arrive independently of the buses and thus have mean waiting times given by (2.2).

Bowman and Turnquist (1981) developed a model which includes a representation of the way in which passengers choose their time of arrival at a bus stop. Any improvement in the adherence of buses to their scheduled departure times will increase the incentive for passengers to plan their journeys with reference to the bus schedule. According to Bowman's and Turnquist's model, this will be reflected in the times at which passengers arrive at the bus stop. Thus the estimated mean waiting time arising from this model is rather more sensitive to improvements in the adherence of buses to their schedule than are the estimates given by other methods.

The assumption that each passenger is able to board the first bus to depart does not always hold and this gives rise to under-estimates of waiting times. Shields (1976) reported that as many as 25 per cent of buses were full at some stops in London during morning peak periods. Danas (1980) found that this was responsible for an increase of as much as 49 per cent in waiting times over what would be the case if all passengers could board the first bus.

There are then three reservations concerning the interpretation of \( \bar{W} \), given by (2.1) as the mean waiting time of passengers at a bus stop and its use in an economic evaluation. Firstly, the methods used to estimate \( \bar{h} \) and \( \bar{a} \) require attention; secondly, passengers appear not to arrive at a bus stop independently of buses, and finally, passengers can not always board the first bus to leave the bus stop. Notwithstanding these reservations, the quantity \( \bar{w} \) does give an indication of some aspects of the quality of a bus service (Chapman, 1976; Shields, 1976).

Stops made by vehicles passing through a junction can have several different effects. Michalopoulos (1976) discussed the consequences for the
comfort of passengers of changing the number of times a vehicle stops, but this effect cannot be quantified readily. Robertson, Lucas and Baker (1980) investigated the effect of stopping and restarting a vehicle on the amount of fuel it consumes. By associating a rate of fuel usage with delay, they arrived at an expression for fuel usage of the form

$$F = aB + bD_L + cH$$  \hspace{1cm} (2.4)

where

- $F$ is the total rate of fuel consumption,
- $B$ is the total distance travelled per unit time,
- $D_L$ is the total rate of delay,
- $H$ is the total number of stops per unit time made,

and $a$, $b$ and $c$ are constant coefficients.

Robertson, Lucas and Baker reported the results of an experiment in which traffic signals were set so as to minimise $F$ rather than $D_L$, as is common practice, and significant reductions in fuel usage were recorded. Akçelik (1981) also considered other variables such as chemical pollution, accident risks and delay and arrived at a composite evaluation measure in which these quantities are traded off against each other. Richardson (1980) noted that including these quantities in the evaluation of bus priority schemes could have important effects on the conclusions.

The benefits to bus operators of bus priority schemes can be considerable. Rescheduling may be possible if a reduction can be made in the time allocated for each bus to traverse a route. Because this period is normally long enough for a large majority of buses to complete one run and be available for the next, any reduction in the variability of the time taken will be of great importance (Skinner, 1980). The consequences can extend beyond these marginal reductions in operating costs. Cooper, Vincent
and Wood (1980) reported that in Swansea the implementation of priority schemes enabled central bus routes to be extended without necessitating any changes in timetables. Seward and Taube (1977) showed how in an example which they considered, if the proposed priority scheme was introduced, then the number of buses in service during peak periods could be reduced without reducing the service offered.

2.3 Bus priority methods

2.3.1 Introduction.

The survey of bus priority methods presented in this section is restricted to those which are applicable to or have some direct effects on signal-controlled junctions. While there is considerable scope (Greater London Council and London Transport Executive, 1976) for improving urban bus services by controlling the departures of buses from termini (Osuna and Newell, 1971; Finnamore and Jackson, 1978) and from intermediate timing points (Huddart, 1973; Barnett, 1974; Finnamore and Jackson, 1978), a discussion of these methods is beyond the scope of this survey.

The methods discussed here fall naturally into two distinct categories. The majority of them require only summary information concerning flows of buses and of other traffic and are passive by nature. Because they do not react to the arrivals of individual vehicles, the effects of these priority methods can be estimated by methods of analysis which are applicable to normal signal-controlled junctions. The second category comprises two methods which do respond to the arrivals of individual vehicles. If, as is normally the case, the arrivals of vehicles are not determined in advance, then these priority methods introduce stochastic variations to the control of the traffic signals. Because of these variations, the effects of these priority methods are comparatively
difficult to analyse. Priority methods in each of these groups are now
discussed together with indications of how normal methods of analysis can
be adapted to apply to junctions where they are implemented.

2.3.2 **Passive priority methods.**

These priority methods are generally effected by considering the
particular requirements and performance characteristics of buses separately
from other traffic. The way in which priority is provided may be apparent,
as is the case when buses are permitted to use roads, parts of roads or to
perform manoeuvres which are prohibited to other traffic. Other methods are
possible, such as biasing signal-settings, which favour the movement of
buses by taking into account their direction and speed of travel, but which
operate without any clearly visible signs. While these methods require
summary statistics concerning the running of bus services, they do not
require details such as the arrival times of individual vehicles.

Exclusive bus lanes are used to reserve road space for buses in two
distinct ways. Firstly, lanes for buses travelling in the same direction as
the adjacent traffic (usually known as *with-flow* bus lanes) can be used to
allow buses to avoid long queues, particularly those forming at junctions
(Oldfield, Bly and Webster, 1977). In some circumstances, *with-flow* bus
lanes are only required during peak periods; these can be designated as
being bus lanes during only part of the day and available to all traffic at
other times (Department of Transport, 1976).

If a *with-flow* bus lane extends to the stop-line at a signal-
controlled junction, then it can be treated as a separate stream of
traffic. The chief disadvantage of *with-flow* bus lanes that extend this far
is that they cause reduction in the road space available at the stop-line
for the adjacent streams of traffic: this can be of considerable
consequence at busy junctions. To mitigate this, with-flow bus lanes can be terminated short of the stop-line. The distance from the end of the bus lane to the stop-line is called the set-back. If the set-back is sufficiently long for more vehicles to be able to queue between the end of the bus lane and the stop-line than can depart during a single green period, then the junction can be analysed without reference to the bus lane. Otherwise the rate at which vehicles cross the stop-line in a period of effective green time for these streams throughout which the queue persists will decrease after an initial period. This can be represented by supposing that different departure rates are appropriate to these streams at the beginning and end of the effective green period. An analysis of the capacity of streams which behave like this has been given by Yagar (1974) and of the delay experienced by vehicles in these streams by Allsop (1977a). An alternative method of analysis is to retain a single departure rate for each stream during the effective green period and to allow for the initial period during which vehicles depart at a higher rate by a notional period of extra effective green time (Allsop, 1977b). This method is adequate for the estimation of the capacity of a stream but may give rise to errors in the estimation of delays since in the latter case the time at which vehicles depart is material. Webster (1972) investigated the problem of determining the best set-back under a variety of criteria.

The second kind of exclusive bus lane enables buses to travel along one-way streets in the opposite direction to other traffic. These are called contra-flow bus lanes and can be used to reduce the distances travelled by buses in networks of one-way streets by enabling them to bypass circuitous sections of routes (Department of the Environment, 1973c, 1973d). Because buses travel in contra-flow bus lanes in the opposite direction to all other traffic on the street, when they enter a junction they form a stream on their own. As a consequence of this, normal analyses
of signal-controlled road junctions are applicable to junctions at which these bus lanes discharge. Bly and Webster (1979) have discussed some of the effects of providing contra-flow bus lanes. Sections of road which are closed to all traffic except buses have similar effects to those of contra-flow bus lanes and can be analysed in an identical manner.

Sufficiently many drivers may wish to perform a turning manoeuvre at some junctions that they would interfere with other traffic if they were permitted to do so. In these circumstances, the manoeuvre can be prohibited and alternative routes provided. If, as is often the case, smaller volumes of traffic performing the manoeuvre could be accommodated, then buses can be exempted from the prohibition with an effect similar to that of providing a contra-flow bus lane. The analysis of junctions where this form of priority is implemented can proceed like that of an ordinary signal-controlled junction with the turning buses forming part of a normal stream.

The remaining priority methods discussed here are effected by adjusting the timings on the signal-controller in a manner which is favourable to buses. This can be done by selecting a measure of performance which takes proper account of the particular characteristics of buses and then attempting to optimise it over the available range of signal-settings.

The choice of these performance measures has been discussed in Section 2.2. This method has been extended by Robertson and Vincent (1975) who incorporated a representation of the way buses travel between junctions into a standard optimisation procedure for networks controlled by traffic signals (Robertson, 1969; Vincent, Mitchell and Robertson, 1980). Since these methods treat the evaluation measure explicitly as an objective function, estimates of their effects arise intrinsically.
2.3.3 Active priority methods

These priority methods use information relating to the positions of vehicles in a road network to modify the way in which the traffic signals operate in favour of buses. Two methods in this category are discussed here: one which meters non-priority traffic into an area which is liable to congestion while allowing buses ready access and one which detects buses individually as they approach a signal-controlled junction and responds accordingly.

Successful implementation of the first method depends on a combination of circumstances. An instance of this occurred in Southampton (Department of the Environment, 1976) where serious congestion occurred during peak periods on a main radial route where it passed through Bitterne. To reduce journey times on this main road, traffic signals were used to control the rate at which traffic could join it according to conditions measured at critical sites. This control strategy, which is known as gating, had the effect of reducing congestion on the main route at the expense of causing additional queues on side roads at signal-controlled junctions. This resulted in a reduction in the total rate of delay and was found not to cause any increase in the journey times of private vehicles travelling to the city centre from within the city boundary. Buses were granted priority access to the main road by means of bus-only streets and exemption from prohibition on turns and thus were able to benefit from the reduction in congestion on the main road without having to queue with other traffic for access to it.

The second method enables signal controllers to detect the imminent arrival of buses and to use this advance information to modify the way in which the traffic signals change. When buses approach a junction in a stream of mixed traffic, they can be distinguished from other traffic by the signal controller in a variety of ways: Dow (1977a, b) has described the method in current use in Britain. Priority which is granted in this way
is denoted variously as responsive priority or priority by selective vehicle detection.

Responsive priority has been implemented in many countries throughout the world with a variety of different rules to determine the action to be taken by the controller when a bus is detected. A review of the rules available is given in Chapter 3: in the remainder of this sub-section some examples of the application of this priority method are discussed.

The first applications of responsive priority were to single junctions in urban areas where the traffic signals operate independently of those at other junctions. Richbell and Van Averbeke (1972), the Department of the Environment (1973b), Richardson and Ogden (1979) and Cottinet, Amy de la Breteque, Henry and Gabard (1980) all found that in these circumstances, buses could benefit substantially from these priority methods by reductions in the mean and variability of the delays they experienced. Cooper, Vincent and Wood (1980) and the Department of Transport (1980) found that these benefits were approximately additive when responsive priority was implemented at several neighbouring junctions which are controlled independently of each other.

In urban areas, adjacent junctions are often controlled together to make use of the way in which groups of vehicles, known as platoons, form and progress through the road network (see, for example, Gartner, Little and Gabbay, 1976; Lieberman and Woo, 1976; Vincent, Mitchell and Robertson, 1980). Responsive priority has been introduced successfully into such areas on various scales. Evans and Skiles (1970) investigated the effects of this at two adjacent junctions on a two-way road in Los Angeles. Cottinet et al. (1980) examined five successive junctions on a one-way road in Nice and Gallivan, Young and Peirce (1980) examined a group of 7 junctions connected by one and two-way streets in Glasgow. The latter two studies compared the
effects of adopting different priority strategies, and all three studies showed that substantial benefits could be achieved by responsive priority methods within networks of coordinated traffic signals.

Responsive priority methods can be used in conjunction with passive ones in a variety of ways. Signal settings can be adjusted while taking into account the effects of the priority system. The Department of Transport (1980) showed that this can play an important part in ensuring a reasonable balance between different classes of road users; this process is discussed further in Chapter 4. At busy junctions, the queue of vehicles may extend beyond the point at which buses can first be detected by the responsive priority system, so buses may be delayed before they can be granted priority. If a with-flow bus lane is provided together with responsive priority, then buses can avoid any such interference from long queues. Michalopoulos (1976) and Cottinet et al. (1980) found that this combination of priority methods gave considerable benefits to buses, including large reductions in the variability of delay. Delays due to long queues of other traffic, which are especially important during peak periods, were obviated. Finally, responsive priority methods can be used to provide special facilities at junctions where some manoeuvres are performed exclusively by buses. If, for example, buses are exempted from prohibition on an opposed turning movement, then a special stage may be provided when required so that turning buses can depart safely. If buses enter a junction from a contra-flow bus lane, then responsive priority can be granted to them to good effect (Cooper and Layfield, 1977).

In the remainder of this thesis, only responsive priority methods are considered. The analyses presented here can be extended to incorporate the effects of other priority methods used in conjunction with responsive priority. Before proceeding with the detailed analyses, a description is given in Chapter 3 of the priority rules which are investigated in later chapters.
CHAPTER 3
RESPONSIVE PRIORITY FOR BUSES

3.1 Introduction

A considerable number of different rules have been used in implementations of responsive priority to determine what action, if any, is to be taken by the signal controller when a bus is detected. In this chapter, details of these rules and qualitative discussions of some of their effects are given. The scope of the analyses presented in Chapters 4 and 5 is then discussed in terms of these rules and the various kinds of signal-controlled junctions introduced in Chapter 1.

The rules described here are of three kinds: those which determine how the signal controller is to respond to give priority to buses, those which determine when rules of the first kind should not be invoked, and those which aim to mitigate any adverse effects of granting priority. Rules of the first kind can be categorised according to the part of the cycle during which they are effective. In particular, different categories of rules are appropriate depending on whether or not the traffic signals which control the stream in which buses travel are displaying green when the bus is detected. Possibilities for each of these kinds of rules are presented in this chapter.

3.2 Rules to give priority when buses have right of way

If a bus is detected when the traffic signals which control the stream in which it travels are displaying green but would otherwise change before the bus could cross the stop line, then the green period can be extended for a sufficient time to allow the bus to pass. This is normally achieved
by allowing the last stage during which buses have right of way to terminate after its normal running time only when no buses have been detected in an immediately preceding period of fixed duration. This is called priority by extension.

Several detectors can be used in a single stream, each calling an extension sufficiently long in normal circumstances to allow a bus to reach the next detector or the stop-line. The use of more than one detector in this way can result in better matching of the total extension period to the time taken by individual buses to travel from the outermost detector to the stop-line and is therefore recommended by the Department of Transport (1977), particularly for use at junctions which are heavily loaded.

Since any bus which is granted priority by extension would otherwise have been delayed for a large part of a normal effective red period, it will benefit by a large reduction in delay. Furthermore, since this priority method affects only the longest delays, it will tend to reduce the variability of the delays incurred by buses. Because this rule is effective only during the last part of the effective green period for buses, only a small proportion of buses will benefit from it.

The effects of granting an extension on the operation of traffic signals are not marked. Any stream that has right of way during the last stage in which buses do will receive additional green time and any stream that does not have right of way during this stage will not be given a green indication until later than would otherwise be the case. However, when priority by extension is granted, the order in which stages occur is not altered and stages are not normally truncated. An exception to this latter remark occurs when stages that occur after the one which is being extended are required to end according to some schedule regardless of any priority extensions that may have been granted. This can happen when traffic signals
at adjacent junctions are coordinated to enhance the progression of platoons of traffic (Evans and Skiles, 1970; Cottinet et al, 1980; Gallivan et al, 1980). In these cases, not only is the last stage during which buses receive green extended, but also some later stage is reduced by a corresponding amount.

Simulation studies of a simple two-stage junction conducted by Vincent, Cooper and Wood (1978) have shown that priority by extension can be used in a wide range of conditions to reduce the total delay incurred by passengers. Because of the benign nature of the effects of this priority method, it is recommended for use at all responsive priority installations (Department of Transport, 1977). It can be implemented alone or with the other priority rules described below.

3.3 Rules to give priority when buses do not have right of way

There is considerable scope for variety in rules to provide priority for buses which are detected at times when the stream in which they travel does not have right of way. If, as normally happens, that stream would gain right of way in any case, then the action of these rules on the detection of a bus is to advance the time at which this occurs. When this is done, any queue which has formed before the detection of the bus can discharge and the bus can pass with reduced delay.

There are several reasons why an instantaneous return to a green indication for buses might be undesirable. The most important of these is that of safety: the minimum green time for any stage that has been called must elapse before another stage can be called and adequate clearance times must be allowed between green indications for streams which are not compatible. Some streams may receive a significant part of their green time during the stages in which buses do not have right of way. These streams
may be sufficiently important to warrant the inclusion of at least a
minimal time for them in every cycle. Finally, some of the transitions from
stages in which the buses do not receive green to the first stage in which
they do may necessitate the use of such long transition periods that they
are undesirable. These transitions can be avoided by including other stages
in order to make better use of the time between the detection of a bus and
the provision of a green indication for it.

Two examples of rules in this category are analysed in this thesis.
The first rule causes the controller to make a transition to the first
stage in which buses have right of way as soon after the detection of a bus
as is practicable. It can be stated as follows. If a bus is detected when
it does not have right of way, then the most recently called stage is
terminated after the minimum green time has elapsed. A transition is then
made directly to the first stage in which buses have right of way. This is
called priority by recall. The second rule, which does not omit any stage
from the sequence, can be stated as follows. If a bus is detected when it
does not have right of way, then the most recently called stage is
terminated after the minimum green time has elapsed and all remaining
stages during which buses do not have right of way are run only for their
minimum permissible durations. This is called priority by hurry-call and
differs from priority by recall only at junctions where there are at least
two stages in which buses do not have right of way. Similarly to the
case of the extension rule, at junctions where the traffic signals are
coordinated with those at adjacent junctions, the truncation or omission of
some stages in order to advance the start of green for buses can cause the
extension of some later stages.

Although other variants are possible, these two rules are the only
ones to have been used in practice. Indeed, nearly all implementations and
investigations of responsive priority have used priority by recall; the
exceptions to this being analytical investigations by Allsop (1977b) and Richardson and Ogden (1979) and practical trials reported by Wood (1976) and Gallivan et al (1980) of priority by hurry-call. Hybrid rules could be devised where some stages may be omitted only if some previous stage has not already been called when a bus is detected; these rules can be analysed in a similar manner to recall and hurry-call. A final possibility is that priority by recall could cause the controller to make a transition to some stage after the first one in which buses have right of way; this is applicable only at junctions where there is more than one such stage.

Since these rules are effective during all of the time in which buses do not have right of way, a comparatively large proportion of buses will benefit from them to some extent. Each such bus will gain less and, because of the truncation or omission of stages from the sequence, will cause more disruption to other streams than those which are granted priority by extension. Indeed, in some circumstances where responsive priority is implemented by recall, some streams may have to wait for a considerable length of time before gaining right of way because of frequent demands made by buses.

In cases where the stream containing buses receives right of way only when a bus is detected, a special stage giving it right of way is introduced into the sequence when appropriate. This form of responsive priority can be used at junctions where buses enter a junction from a bus-only street or a contra-flow bus lane or perform an exclusive manoeuvre which requires some special provision.

Depending on the requirements at a particular installation, a special stage can be introduced in a variety of different ways. It can be introduced immediately after the last stage to be called before the detection of the bus or between two particular stages. In the former case,
the normal sequence can be resumed either with some particular stage or
with the stage that would have been called next but for the detection of
the bus. If the special stage is always inserted between the same two
stages, then any stages remaining before the special stage at the time when
a bus is detected can be shortened so that the bus gains right of way
quickly. A special stage of this kind could be used to allow buses to make
an exclusive opposed turning movement in safety and might just consist of
an extended transition period.

If the special stage is always followed immediately by a particular
stage, then the behaviour of the traffic signals is similar to that where
priority is implemented by recall or hurry-call. From the point of view of
non-priority streams, the extra stage is similar to a long transition
period and some of them may safely receive green indications during this
period.

3.4 Rules to suspend priority

The previous two sections have dealt with different ways in which
priority can be given to buses. In some circumstances an application of
these rules might cause excessive disruption to non-priority traffic. A
number of different rules are available to identify circumstances in which
priority should be suspended to prevent this.

The simplest rules suspend responsive priority during periods when the
prevailing conditions are considered to be unsuitable for it. Tarnoff
(1975) and MacGowan and Fullerton (1979) described a rule which suspends
priority whenever the flow in any of the streams at a junction exceeds some
pre-set level. This rule has the consequence that priority is denied to
buses at exactly those times when they could benefit most from it; namely
during busy periods.
A more usual form of rule is based on the principle that after each time priority is granted the complete sequence of stages should be allowed to run uninterrupted. The rule recommended by the Department of Transport (1977) for use in Britain can be stated as follows. After a recall or hurry-call has been granted, no further recalls or hurry-calls are to be granted for the duration of one complete cycle. This is called the inhibition rule. A more restrictive rule has been considered by Allsop (1977b) and Cottinet et al (1980) but has not found any practical application. This rule suspends all priority for the duration of one complete cycle after any form of priority has been granted. According to this rule, if a bus is granted priority by extension, recall or hurry-call, then no further priority is to be granted before the stream containing buses next gains right of way in the normal manner.

Evans and Skiles (1970) used a different rule which resulted from consideration of the duration of the effective green period for streams at a two-stage junction that are not compatible with the stream containing buses. Because the traffic signals at the junction which they considered were coordinated with those at adjacent junctions, any extension granted for buses resulted in a reduction in the duration of the next stage. To prevent this from being reduced further, a rule was introduced which allowed priority to be granted by extension whenever it was required but by recall only if the previous stage had not been extended.

The total duration of the last stage in which the stream containing buses has right of way may be subject to a maximum constraint, in which case priority by extension may not be granted if it would violate this constraint. This results in priority by extension being available for a period that stops short of the last instant at which the stream containing buses can have right of way. Gallivan et al (1980) reported the use of two methods to calculate the value of the maximum constraint which were also
used to calculate suitable minimum green times for the stages. One of these methods used estimates of the size of the queue in each stream derived from knowledge of previous stage durations to determine these values in each cycle.

3.5 Rules to mitigate the adverse effects of priority

Responsive priority has the effect of extending stages in which buses have right of way and of truncating or omitting others. Thus unless remedial action is taken, the proportion of all time for which the stages in which buses have right of way will be increased to the detriment of others. That this could aggravate the problem of providing adequate capacity for all streams of traffic during busy periods has been noted by Huddart and Allen (1972) among others. Three different rules have been used in practice to mitigate this effect and are discussed here.

The first method has been used successfully in the USA (Tarnoff, 1975; MacGowan and Fullerton, 1979) and operates by suspending all priority for the duration of periods when measured traffic flows exceed some preset level, and has been discussed in Section 3.4. This certainly prevents responsive priority from causing any reduction in the capacity of streams at times when the consequences of this would be most serious. However, it suffers from the inherent disadvantage discussed above.

The second method is that of active compensation, where each stage that is truncated or omitted in one cycle is prolonged when it next occurs. This method can be used to restore the proportion of all time for which each stage runs (Richardson and Ogden, 1979) and to correct the starting times of each stage in relation to those at adjacent junctions (Cottinet et al, 1980). It has been used successfully in practice (see, for example, Department of the Environment, 1973b) and is recommended by the Department
of Transport (1977) for use with priority by recall. However, simulation studies conducted by El-Reedy and Ashworth (1978) showed that even in comparatively light traffic, this compensation method can upset the smooth progression of platoons between junctions on arterial roads.

Finally, a passive form of compensation is possible where the times allocated for each stage during normal operation are biased towards the stages in which buses do not have right of way. The aim of this method is to distribute over all occurrences of those stages which can be truncated or omitted the extra green time necessary to restore some desired balance between the durations of the stages. This has been used to good effect at one of the junctions in the Swansea scheme where it was found by the Department of Transport (1980) to give a considerable reduction in delay to non-priority traffic over the active compensation method previously used there (Cooper, Vincent and Wood, 1980).

Vincent, Cooper and Wood (1978) found that the method of passive compensation extended the range of conditions which their simulation model indicated to be suitable for the application of responsive priority. However, their method of calculating suitable adjustments to the stage times was approximate and applicable only to one priority method at junctions with 2 stages in the sequence. The same method was used to calculate the adjustments recommended by the Department of Transport (1977). Allsop (1977b) considered this problem in greater detail for several different priority methods applied to junctions with an arbitrary number of stages. The adjustments were expressed as a power series in the mean bus arrival rate, thus permitting the calculation of the desired signal settings when this rate is sufficiently small.
3.6 Scope of the study

Two aspects of the operation of responsive priority are investigated in this thesis. In Chapter 4, formulae are derived to estimate the capacity of each stream at a junction where responsive priority is implemented and Allsop's (1977b) analysis of the problem of finding suitable signal settings is extended. In Chapter 5, formulae to estimate the delay incurred by buses and non-priority vehicles are derived and applied to the most usual forms of responsive priority. Before proceeding with these analyses, their scope is described here in some detail.

Although mention is made only of providing priority for buses, responsive priority and these analyses of its effects are applicable equally to other kinds of vehicles. Examples of this are provision for other public service vehicles, as reported by Vincent and Hoppe (1970) and Mertens (1973) and for emergency vehicles, as reported by Griffin (1978) and Griffin and Johnson (1980).

In both analyses, buses are supposed to arrive according to the Poisson law. Thus if the mean bus arrival rate is $\beta$ buses per cycle, then the probability (denoted by $\tilde{p}_0(\xi)$) that no buses arrive in a period of time corresponding to a proportion $\xi$ of a cycle is given by

$$\tilde{p}_0(\xi) = e^{-\beta \xi} \quad (3.1)$$

Furthermore, the probability density function $\tilde{p}(\xi)$ of the bus headway distribution is given by

$$\tilde{p}(\xi) = \beta e^{-\beta \xi} \quad (3.2)$$
The standard deviation of this headway distribution is equal to the mean value of $1/\beta$; this is in general agreement with observations of individual bus routes in urban areas by Chapman, Gault and Jenkins (1976) and the Greater London Council and London Transport Executive (1976). This pattern corresponds to a uniform probability of a bus arrival at all times and will model arrivals quite well, especially in cases where buses on several independent routes enter a junction in the same stream of traffic (Cox, 1962, p 77). Allsop (1977b) considered this and a more regular arrival pattern which allowed for no more than one bus arrival between successive calls of the first stage in which buses have right of way. Any differences between the calculated signal-settings were shown to be of second or higher order in the mean bus arrival rate, indicating that these calculations are fairly insensitive to the pattern of bus arrivals when the average number of arrivals per cycle is appreciably less than 1.

In the first instance, all the buses for which priority is provided are supposed to enter the junction in the same stream. Furthermore, each such bus is supposed to be detected at just one location as it approaches the junction. There is no difficulty in applying the analysis to junctions where responsive priority is provided for buses which enter the junction in several streams provided that the signal controller responds to the detection of a bus independently of which stream the bus is in. While it is possible to provide responsive priority for buses arriving at a junction in streams which are not compatible (see, for example, Wood, 1976; Cottinet et al, 1980), a great number of different combinations of priority rules is available, so a general analysis is impractical. However, the methods presented here could be extended and applied to particular instances of this as required.

In Chapter 4, both priority by recall and hurry-call are considered together with extension and inhibition. In Chapter 5, explicit formulae are
given only for the combinations of priority by extension and recall with
and without inhibition. In all cases, the extension of one stage is
supposed not to affect the duration of any other stages and similarly, when
the start of a stage is advanced by the recall or hurry-call rules, the
durations of this and subsequent stages are supposed not to be affected.
The analysis could be modified readily to apply to junctions where this
does occur.

In order to ensure the independence of the action of priority rules
implemented together, the length of an extension is supposed to be less
than the time from the end of the last stage before buses gain right of way
to the end of the last stage in which buses have right of way. This means
that if a bus is granted priority by recall or hurry-call, it will not need
an extension as well in order to pass through the junction.

In both of the analyses presented in this thesis, the stream
containing buses is supposed to receive right of way during only one set of
contiguous stages in the sequence. Vincent and Hoppe (1970) reported an
application of responsive priority at a junction where the one stage during
which buses received right of way was included twice in the sequence. This
reduced the longest time between successive green periods for buses and was
found to reduce the delay experienced when compared with sequences where
this stage occurred only once but had a longer duration. The analyses
presented here could be extended to apply to junctions subject to control
like this. In Chapter 4, no further assumptions are made concerning the
nature of the sequence. However, in Chapter 5, each stream is supposed to
receive right of way during only one set of contiguous stages in the
sequence and further is supposed not to have right of way at some times. In
general there are many sequences for a junction which satisfy these
conditions: Tully (1976) has developed a procedure to generate all of them
so that the most suitable one can then be selected according to an
appropriate criterion. Throughout this thesis, the sequence is taken to be predetermined.

The rules of operation of vehicle actuated traffic signals in current use in Britain (Ministry of Transport, 1970) cause each stage to be extended further beyond its minimum green time whenever a vehicle is detected in a stream which has right of way during that stage until a maximum permissible duration has elapsed. A transition is then made to the next stage in the sequence for which a demand has been caused by the detection of a vehicle. Since capacities of streams are relevant principally during busy periods, they can be estimated under the assumption that there is sufficient demand for each stage to run for its maximum permissible duration in every cycle which is not interrupted to provide priority for a bus. No assumptions concerning the short-term variability of arrivals of non-priority vehicles are necessary in the analysis of the capacity of a stream of traffic (Allsop, 1972b). Thus the results presented in Chapter 4 are applicable to junctions which are controlled either by vehicle actuated or fixed time traffic signals and at which vehicles arrive in almost any manner, including the possibility that they arrive mainly in platoons due to the effects of adjacent signal-controlled junctions.

More restrictive assumptions are required in the analysis of delay presented in Chapter 5. These are that in each stream the arrivals of vehicles during disjoint intervals are mutually independent and are independent of the durations of the stages and that the period for which each stage runs is fixed except for when it is modified in the process of granting priority to a bus. Accordingly, the estimates of delay are appropriate only at junctions which are isolated from others and are controlled by traffic signals which, apart from the granting of priority, can be regarded as operating fixed-time. The further restriction is required that each stream has right of way during a single interval in each uninterrupted cycle.
CHAPTER 4
CAPACITY

4.1 Introduction

In Section 1.3, the capacity of a stream of traffic was defined to be the greatest mean rate, measured in pcu's, at which traffic can enter a junction from that stream. If the mean arrival rate, also measured in pcu's, exceeds the capacity, then a net accumulation of vehicles queueing in that stream will be caused. A stream in which this occurs is said to be oversaturated.

The discussion in Section 2.2 showed that there is no simple measure of costs and benefits appropriate to describe all the effects of a priority scheme. However, if any streams become oversaturated for any substantial period as a consequence of the implementation of a priority scheme, then the additional delay caused to vehicles in these streams is likely to be so large as to outweigh any benefits to buses, irrespective of the particular evaluation criterion adopted. Thus the provision of adequate capacity for all streams of traffic is a necessary part of any procedure to maximise the benefits of a priority scheme.

Methods to estimate the capacity of the streams of traffic at a junction which is controlled by fixed-time traffic signals have been developed by Allsop (1972b), Ohno and Mine (1973a, b) and Yagar (1974). These analyses lead to methods to determine signal-settings which maximise the common quantity by which the mean arrival rates in all streams could be multiplied without causing any stream to be oversaturated. While these signal-settings do not necessarily optimise any particular measure of benefit, they do ensure a reasonable allocation of time for each stream at
a junction to receive right of way. This has the advantage over explicit optimisation of measures such as delay that it can be applied even if the mean arrival rates are so high that at least one stream is oversaturated whatever signal-settings are implemented.

In this chapter, previous analyses of capacity are reviewed and extended to junctions where the durations of the stages are random variables. A general relationship is given between signal-settings implemented with different rules of operation of the junction in terms of the resulting capacity of the streams of traffic. Explicit formulae are derived for these capacities at a junction where priority is provided for buses by extension with each of hurry-call and recall, both with and without inhibition. These formulae can be used to investigate the consequences of the implementation of responsive priority for buses and of variations in the circumstances of implementation from those for which the priority scheme was designed. Expressions are deduced for signal-settings which would, if implemented without priority for buses, give rise to capacities similar to those arising from the signal-settings used as part of the priority scheme.

If some particular signal-settings are known to provide adequate capacity for all streams in the absence of responsive priority, then the traffic manager may wish to find signal-settings which, when implemented with a particular priority method, give rise to similar capacities. The problem of finding signal-settings which achieve this aim is formulated as the inverse of the process described above. This problem has been investigated by Allsop (1977b), who showed how appropriate values for the signal-settings could be expressed as a power series in the mean bus arrival rate. While this result guarantees the existence of the desired signal-settings when the mean bus arrival rate is sufficiently small, no indication is given of the range of arrival rates for which solutions
exist, or indeed for which the power series converges. A further analysis is presented here which delimits the range of bus arrival rates for which acceptable signal-settings exist. An alternative solution method is proposed and investigated. Two examples are given in Chapter 6 to illustrate the application of this method.

4.2 The capacity of a stream of traffic

4.2.1 Introduction.

The object of the analysis of the capacity of a stream of traffic at a signal-controlled road junction presented here is to establish a relationship between the capacity and the signal-settings used. The first step is to investigate the properties of the departure model. To this end, lemma 4.1 below shows the problem of estimating the capacity of a stream to be equivalent to finding the proportion of all time which is effectively green for that stream. The latter problem is then reduced to the investigation of various statistics of the stages in the control sequence.

Notation is introduced as it is required and retained throughout the remainder of this thesis. A complete list of the notation introduced in this way is given in Appendix 3. Consider a junction at which \( M \) streams of traffic discharge. For each stream \( j \) \( (1 \leq j \leq M) \), let

- \( s_j \) be the saturation departure rate (pcu's per second)
- \( q_j \) be the mean arrival rate (vehicles per second)
- \( k_j \) be the mean number of pcu's per vehicle
- \( K_j \) be the capacity of stream \( j \) (pcu's per second)
- \( \lambda_j \) be the effective proportion of all time for which stream \( j \) has right of way with saturation departure rate \( s_j \)
- \( g_j \) be the duration of an effective green period
- \( r_j \) be the duration of an effective red period.
The following result can now be stated concisely.

**Lemma 4.1**

If stream \( j \) experiences alternating periods during which it has right of way with saturation departure rate \( s_j \) and periods when no vehicles can proceed, then

\[
K_j = A_j s_j \quad \text{pcu's per second} \quad (4.1)
\]

**Proof**

To determine the capacity of the stream, consider the rate at which traffic can enter the junction if the mean arrival rate is sufficiently great that the queue is never empty. For this stream, time can be divided up into alternating effective red and green periods: a single pair of these periods is considered.

Suppose that the effective green period has duration \( g_j \) and that in time \( g^* \leq g_j \), \( n \) vehicles have departed, leaving a vehicle at the front of the queue with pcu equivalent \( t > s_j (g_j - g^*) \). According to the departure model, exactly \( s_j g^* \) pcu's have departed in time \( g^* \) and the next vehicle will depart with probability \( s_j (g_j - g^*)/t \). Denoting the number of pcu's departing in the effective green period by \( N \), the expected value of \( N \) conditional on the value of \( g_j \) is given by

\[
E(N|g_j) = s_j g^* + t s_j (g_j - g^*)/t = s_j g_j \quad (4.2)
\]

Taking expectations over \( g_j \), the mean number of pcu's to depart in a pair of effective red and green intervals is

\[
E(N) = s_j E(g_j) \quad . \quad (4.3)
\]
The mean duration of these periods is $E(r_j + g_j)$ so the greatest mean rate at which traffic can enter the junction from stream $j$ is

$$K_j = \frac{s_jE(g_j)}{E(r_j) + E(g_j)}$$

$$= \Lambda_j s_j \quad (4.1)$$

**4.2.2 Discussion.**

The assumption used in the proof of lemma 4.1 that vehicles would always be present in the queue implied that the stream is oversaturated: if this were not so, then the mean rate at which traffic enters the junction would be less than the capacity. As a consequence of this, the capacity of a stream provides an upper bound on the mean arrival rate for which the stream is not oversaturated of the form

$$q_j \leq \frac{\Lambda_j s_j}{K_j} \quad (4.4)$$

Various authors have noted that in many cases a junction will not operate satisfactorily if the arrival rate in a stream exceeds some given proportion $p_j$ of the capacity of that stream. This leads to the corresponding practical capacity constraint for each stream

$$q_j \leq p_j \frac{\Lambda_j s_j}{K_j} \quad (1 \leq j \leq M) \quad (4.5)$$

Webster and Cobbe (1966) suggest that the value $p_j = 0.9$ is suitable for use in normal circumstances, though any value in the range $[0, 1]$ is possible, where the limiting case where $p_j = 1$ corresponds to the constraint (4.4).
The departure model supposes that once the queue has dissipated, it will not form again until after the end of the effective green period. As a consequence of this, there is no upper bound on the number of vehicles which can enter the junction during a single effective green period. However, if the mean arrival rate in a stream exceeds the capacity, then once a queue has become established, the mean rate at which traffic can leave the queue is restricted by that capacity so the net size of the queue will increase without bound. Thus satisfaction of the capacity constraint \((4.4)\) is a necessary and sufficient condition for a stream to be undersaturated in all cases where the arrivals of vehicles are not co-ordinated to a large extent with the last part of the effective green periods.

The analysis of capacity presented here supposes that the arrivals of vehicles are sufficiently variable to ensure that \((4.4)\) is equivalent to undersaturation. This will certainly be the case at isolated junctions and at junctions in urban networks where other junctions are sufficiently far removed for traffic not to arrive almost exclusively in platoons.

If there is a minimum possible headway between successive vehicles arriving in a stream, then this assumption of independence can be relaxed. Indeed, in practice vehicles will not normally arrive at the stop-line more frequently than the saturation departure rate. If this is assumed to happen, then the analysis of capacity presented here can be applied regardless of any other properties of the arrival process. This includes the possibility that platoons of traffic are arranged to arrive during the effective green period as is often the case in systems of co-ordinated traffic signals.
If the traffic signals are subject to vehicle-actuated control, then
the persistent queues would ensure that each stage during which a heavily
loaded stream has right of way runs to the maximum value allowed for it.
Thus the expected value of the duration of the effective green period used
in lemma 4.1 can be calculated from the maximum permitted stage durations.
During busy periods, all other stages will normally run for their maximum
durations so the expected value of the duration of the effective red
period can also be calculated from maximum stage times. Any reduction of
demand in other streams can only result in a reduction in the duration of
the effective red period. Thus the procedure described yields the most
restrictive value which could arise.

In lemma 4.1, the stream under consideration was supposed to
experience alternating periods during which traffic can discharge at the
full saturation departure rate or can not depart at all. While this
formulation is sufficiently versatile to encompass most kinds of streams
occurring in practice, no account is taken by it of the possibility that
the saturation departure rate of a stream differs from stage to stage of
the cycle. A common instance in which this occurs is that of opposed
turning movements where vehicles are permitted to discharge through gaps in
the opposing stream during some stages and have exclusive right of way
during others.

The next result provides a natural extension to the result of lemma
4.1 for a stream which has right of way with various saturation
departure rates.
Corollary 4.2
Suppose that stream \( j \) experiences \( m \) saturation departure rates. Let \( f_{ij} (1 \leq i \leq m) \) be the proportions of \( s_j \) experienced and suppose that the saturation departure rate \( f_{ij} s_j \) obtains for a proportion \( \Lambda_{ij} \) of all time \((1 \leq i \leq m)\). Then the capacity of the stream is given by

\[
K_j = \sum_{i=1}^{m} \Lambda_{ij} f_{ij} s_j
\]  \( (4.6) \)

Proof
Since the periods during which the several saturation departure rates obtain are necessarily disjoint, the result of lemma 4.1 can be applied separately to the periods during which each of them obtains. Thus the contribution to the capacity from periods during which the saturation departure rate \( f_{ij} s_j \) obtains is \( \Lambda_{ij} f_{ij} s_j \). Summing these contributions over all values of \( i \) gives (4.6).

The quantity \( \sum_{i=1}^{m} \Lambda_{ij} f_{ij} \) can be interpreted as the effective proportion of all time for which the saturation departure rate \( s_j \) obtains. If \( \Lambda_j \) is identified in this way, then (4.1), the result of lemma 4.1, is recovered.

Ohno and Mine (1973a) gave detailed methods to estimate the mean duration of the parts of stages during which different proportions of the unobstructed saturation departure rate would obtain to a stream of opposed turning traffic. Similar methods were also applied to estimate the proportion of the saturation departure rate achieved during these periods. Both of these methods used the assumptions of identical service times for all vehicles and Poisson arrivals in the opposing streams. Yagar (1974) presented a similar but less detailed analysis in which a known proportion of the saturation departure rate was supposed to obtain throughout each stage.
Allsop (1977a) discussed various methods by which opposed turning vehicles could be accommodated at a junction. He observed that the possibility of vehicles passing through gaps in opposing streams is not normally important in the analysis of junction capacity. This is because the occurrence of suitable gaps in the opposing streams becomes rare as flows approach capacity and if the stream containing opposed turners is nearly saturated, then periods of exclusive right of way are desirable and can normally be provided.

4.2.3 Traffic capacity in relation to signal timings.

The result of corollary 4.2 showed that the capacity of stream \( j \) can be determined from \( \lambda_j \), the effective proportion of all time for which that stream has right of way. Some further notation is now introduced and expressions for \( \lambda_j \) are derived from statistics of the stages in the control sequence. The formulation developed is sufficiently flexible to encompass the bus priority methods to be investigated and agrees with that of Allsop (1972b) and Yagar (1974) when applied to junctions at which there is no priority. A general relationship is established between signal timings which give rise to similar capacities for the streams under different circumstances of implementation. In particular, signal-settings used in conjunction with responsive bus priority methods are related to signal-settings at the same junction in the absence of priority for buses.

Consider a junction controlled by traffic signals with \( m \) stages in the sequence. For convenience, these are taken to be numbered starting from the first stage in which the stream containing buses for which priority is provided has right of way. The number of the last stage in which this stream has right of way is denoted by \( k \). In the absence of priority, the sequence of stages from 1 to \( m \) will be repeated ceaselessly.
The effect of granting priority by extension is to increase the duration of stage \( k \). The effect of granting priority by hurry-call is to reduce the duration of some of the stages with numbers greater than \( k \). Neither of these priority rules change the order in which the stages occur. However, if priority by recall is granted during stage \( i \) \((k < i < m)\), then a transition is made to stage \( 1 \) as soon as this is permissible, thus eliminating stages \( j \) \((i < j < m)\) from that cycle. Accordingly, the model investigated here incorporates the possibility of each transition from stage \( i \) \((1 \leq i < m)\) to stage \( i+1 \) and from each stage \( i \) \((k < i < m)\) to stage \( 1 \). A cycle is now taken to be the period between the start of successive occurrences of stage \( 1 \).

The notation used to describe the operation of the signal controller is as follows. Let

- \( c \) be the duration of an uninterrupted cycle \((s)\)
- \( \lambda_{i0} \) be the duration of the transition period between stage \( i \) and stage \( i+1 \) \((1 \leq i < m)\) \((s)\)
- \( \lambda_{i1} \) be the duration of the transition period between stage \( i \) and stage \( 1 \) \((k < i < m)\) \((s)\)
- \( \lambda_{m0} \) be the duration of the transition from stage \( m \) to stage \( 1 \):

\[
\lambda_{m0} = \lambda_{m1} \quad (s)
\]

- \( L \) be the total duration of the transition periods in an uninterrupted cycle \((s)\)
- \( \lambda_o \) be \( L/c \), the proportion of an uninterrupted cycle which comprises transition periods
- \( \tau_{0c} \) be the mean duration of the transition periods in a cycle \((s)\)
- \( G_i \) be the minimum green time for stage \( i \) \((1 \leq i < m)\) \((s)\)
- \( \gamma_i \) be \( G_i/c \) \((1 \leq i < m)\)
- \( \lambda_{1c} \) be the maximum permissible duration of stage \( i \) \((1 \leq i < m)\) \((s)\)
\( \varepsilon_{\text{c}} \) be the mean over all cycles in which the junction is heavily loaded of the duration of stage \( i \) (\( 1 \leq i \leq m \)) \((s)\)

\( P_n \) be the probability that there are exactly \( n \) stages in a cycle \((k < n \leq m)\)

\( \delta_{ij} \) be the Kroneker delta, defined by

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{otherwise.} 
\end{cases}
\]

The permissible stage transitions are shown schematically in Figure 4.1 together with their durations. In this figure, each node represents a stage and each directed arc a permissible transition. Figure 4.1a shows the stage changes which occur during normal operation and when there is priority by extension or hurry-call. Figure 4.1b shows the additional stage changes which can be caused by a priority recall of stage 1.

Some relationships between these quantities follow immediately and are stated here.

**Lemma 4.3**

The mean duration of the transition periods in a cycle satisfies

\[
\varepsilon_0 = \frac{\kappa}{i=1} \lambda_{i0} + \sum_{n=k+1}^{m} \frac{\lambda_{i0}^{n-1}}{P_n} \lambda_{i1}^{n1} \tag{4.7}
\]

**Proof**

If there are \( n \) stages in a cycle, then the duration of the transition periods in that cycle is \( \sum_{i=1}^{n-1} \lambda_{i0}^{n-1} + \lambda_{i1}^{n1} \). The probability of this event is \( P_n \), which is non-zero only if \( k \leq n \leq m \). Taking expectations over \( n \) gives

\[
\varepsilon_{\text{c}} = \left[ \sum_{n=k+1}^{m} P_n \sum_{i=1}^{n-1} \lambda_{i0}^{n-1} \lambda_{i1}^{n1} \right] c
\]

Using \( \sum_{n=k+1}^{m} P_n = 1 \) and dividing both sides by \( c \) gives (4.7). \( \square \)
Figure 4.1a: Stage transitions occurring in normal operation, together with their associated durations.

Figure 4.1b: Stage transitions occurring when stage 1 is recalled, together with their associated durations.
Corollary 4.4

If there are \( m \) stages in every cycle, then \( \varepsilon_0 = \lambda_0 \).

Proof

Since there are \( m \) stages in every cycle, \( P_n = \delta_{nm} \), so (4.7) reduces to

\[
\varepsilon_0 = \sum_{i=1}^{m-1} \lambda_i + \lambda_m
\]
\[
= \sum_{i=1}^{m} \lambda_i
\]
\[
= \lambda_0 .
\]

If a stream has right of way in the stages before and after a transition, it will normally also have right of way throughout that transition period. Some streams, such as those containing opposed turning traffic, may receive a bonus of extra effective green time during some transitions. Stages are delimited by changes in the state of the controller rather than the start or end of the effective green periods of the streams, so some part of the transition periods which initiate and terminate the effective green period for each stream may be included in it.

The notation used to relate the behaviour of the streams of traffic to the operation of the signal controller is as follows. Let

- \( a_{ij} \) be the proportion of the saturation departure rate \( s_j \) achieved by stream \( j \) during stage \( i \) \((1 \leq i \leq m) \) \((1 \leq j \leq M)\)
- \( a_{0j} \) be the effective proportion of all time during transition periods which is green for stream \( j \) \((1 \leq j \leq M)\)
- \( a_{10j} \) be the effective proportion of the transition period between stages \( i \) and \( i+1 \) which is green for stream \( j \) \((1 \leq i < m) \) \((1 \leq j \leq M)\)
\( a_{11j} \) be the effective proportion of the transition period between stages 1 and 1 which is green for stream \( j \) (\( k < i \leq m \)) (\( 1 \leq j \leq M \)).

As with the duration of the transition periods, this notation is extended so that \( a_{m0j} = a_{m1j} \) (\( 1 \leq j \leq M \)).

In this notation, \( a_{ij} \) corresponds to the product of \( a_{ij} \) and \( f_{ij} \) of Yagar (1974). If stream \( j \) has right of way during stage \( i \) and the next stage, then whichever of \( a_{i0j}^\lambda \) and \( a_{i1j}^\lambda \) is appropriate normally takes the value 1. If stage \( i \) is the last stage during which stream \( j \) does not have right of way in an uninterrupted cycle, then \( a_{i0j}^\lambda \) \( a_{i1j}^\lambda \) is the difference in time between the start of the next stage and the start of the green indication for stream \( j \) less an allowance for the lag before the start of the effective green period. If stream \( j \) does not have right of way during stage \( i \) (\( k < i \leq m \)) but does during stage 1, then \( a_{i1j}^\lambda \) is determined in a similar manner. Finally, the values of \( a_{i0j}^\lambda \) and \( a_{i1j}^\lambda \) can be calculated for transitions which terminate the effective green period of stream \( j \) from any extra green indication plus the lag between the end of the green indication and the end of the effective green period.

Some relationships between these quantities follow immediately and are stated as preliminaries to the main result of this subsection, Theorem 4.7.

**Lemma 4.5**

The effective proportion of all time during transition periods which is green for stream \( j \) (\( 1 \leq j \leq M \)) satisfies

\[
\begin{align*}
\hat{a}_{0j} & = \sum_{i=1}^{k} a_{ij}^{i0} + \sum_{n=k+1}^{m} P \{ \sum_{i=k+1}^{n-1} a_{ij}^{i0} + a_{ij}^{i1} \} \\
\end{align*}
\]  

(4.8)
Proof
If there are \( n \) stages in a cycle, then the total effective green time for stream \( j \) during the transition periods of that cycle is
\[
\left( \sum_{i=1}^{n-1} a_{i0j} \lambda_{i0j} + a_{n1j} \lambda_{n1j} \right) c .
\]
Taking expectations over \( n \) and rearranging as in lemma 4.3 gives the desired result. 

\[
\text{Corollary 4.6}
\]
If there are \( m \) stages in every cycle, then
\[
a_{0j} = \sum_{i=1}^{m} a_{i0j} \left( \frac{\lambda_{i0j}}{\lambda_0} \right)
\]

Proof
From corollary 4.4, under these hypotheses, \( \epsilon_0 \sim \lambda_0 \). Proceeding as in the proof of that corollary then substituting for \( \epsilon_0 \) and dividing both sides by \( \lambda_0 \) gives the desired result.

\[
\text{Theorem 4.7}
\]
The capacity of stream \( j \) \((1 \leq j \leq M)\) is given by
\[
K_j = \left( \frac{\sum_{i=0}^{m} a_{ij} \epsilon_i}{\sum_{\ell=0}^{m} \epsilon_{\ell}} \right) s_j
\]

Proof
At any particular time the controller is either running a stage or undergoing a transition between stages. Thus the proportion of all time for which stage \( i \) \((1 \leq i \leq m)\) runs is \( \epsilon_i / \sum_{\ell=0}^{m} \epsilon_{\ell} \) and the effective proportion of all time for which stream \( j \) \((1 \leq j \leq M)\) has right of way during an occurrence of stage \( i \) is \( a_{ij} \epsilon_i / \sum_{\ell=0}^{m} \epsilon_{\ell} \). The proportion of all time which is occupied by transition periods is \( \epsilon_0 / \sum_{\ell=0}^{m} \epsilon_{\ell} \), where the value of \( \epsilon_0 \) is calculated from (4.7). The effective proportion of all time for
which stream \( j \) \((1 \leq j \leq M)\) has right of way during transition periods is
\[
a_{0j} \varepsilon_0 \left( \sum_{\ell=0}^{\lambda \ell} \varepsilon_{\ell} \right)\]
where the value of \( a_{0j} \varepsilon_0 \) is calculated from \((4.8)\).

Summing the effective proportions of all time occurring in each state over the states gives
\[
\Lambda_j = \frac{\sum_{i=0}^{m} a_{ij} \varepsilon_i}{m} \left( \sum_{\ell=0}^{\lambda \ell} \varepsilon_{\ell} \right) \quad (1 \leq j \leq M) \tag{4.10}
\]

Substituting this form for \( \Lambda_j \) into \((4.1)\) gives \((4.9)\).

**Corollary 4.8**

If there is no priority for buses, then
\[
K_j = \frac{\sum_{i=0}^{m} a_{ij} \lambda_i \varepsilon_j}{m} \quad (1 \leq j \leq M) \tag{4.11}
\]

**Proof**

Under these hypotheses, in busy periods \( \varepsilon_i = \lambda_i \) \((1 \leq i \leq M)\) and there will be \( m \) stages in every cycle. Furthermore, from corollaries 4.4 and 4.6, \( \varepsilon_0 = \lambda_0 \) and \( a_{0j} = a_{0j} \lambda_0 \) \((1 \leq j \leq M)\). Substituting these particular values into \((4.9)\) gives
\[
K_j = \left( \sum_{i=0}^{m} a_{ij} \lambda_i \right) \varepsilon_j \quad (1 \leq j \leq M)
\]

From the definition of \( \lambda_i \) \((0 \leq i \leq M)\) as exhaustive and mutually exclusive proportions of an undisturbed cycle, \( \sum_{\ell=0}^{\lambda \ell} \lambda_{\ell} = 1 \), so \((4.11)\) is established.
In the proof of theorem 4.7, the effective proportion of all time for which a stream has right of way was calculated from the effective proportion of each cycle for which this occurs. By contrast, the method used to establish the result of lemma 4.1 and hence that of corollary 4.2 considered the ratio of the duration of an effective green period to the time between the start of two successive effective green periods. There is no contradiction inherent in the use of these distinct methods; rather the difference emphasises the different considerations in each case. In lemma 4.1 the primary consideration was the periods of effective red and green experienced by a stream, whereas in theorem 4.7 the primary consideration was the durations of the stages in the control sequence. This approach to the capacity of a stream is flexible in that no additional hypotheses are required concerning the way in which each stream receives green indications. In particular, the result of theorem 4.7 is applicable without any restriction on the number of separate effective green periods experienced by a stream during a single cycle.

The result of corollary 4.8 is equivalent to that of Yagar (1974) since here the proportion of the saturation flow \( a_j \) achieved by stream \( j \) during stage \( i \) has been absorbed into \( a_{ij} \) rather than being displayed explicitly. If the values taken by \( a_{ij} (1 \leq i \leq m) \) in this result are restricted to either 1 or 0, indicating respectively unopposed right of way or no right of way, then Allsop's (1972b) result is recovered.

The result of theorem 4.7 shows that the capacity of stream \( j \) \((1 \leq j \leq M)\) can be determined exactly from the quantities \( e_i (1 \leq i \leq m) \) and \( P_n (k < n \leq m) \) provided that the values of the constants \( a_{ij}, a_{10j} (1 \leq i \leq m), (1 \leq j \leq M) \) and \( a_{11j} (k < i \leq m) (1 \leq j \leq M) \) are known. Expressions for
the former quantities are derived for a variety of bus priority rules in Section 4.4.

While this exact analysis is useful in its own right, the approximate method introduced next is rather more flexible. This is used to define a signal-setting policy which is more readily amenable to a mathematical analysis and which can be used to provide signal-settings which are suitable for use with responsive priority.

The following definition provides a formal relationship between signal-settings applied at the same junction with the same control sequence but under different rules of operation. Two sets of signal-settings, together with their circumstances of implementation, are said to be capacity-equivalent if the proportion of all time for which each stage runs is the same in each case during periods when the junction is heavily loaded. The next result is used extensively in Section 4.5.

**Lemma 4.9**

Any set of signal-settings \( \{\lambda_i\} \) together with their circumstances of implementation are capacity-equivalent to the unique set of signal-settings \( \{\lambda^*_i\} \) under fixed-time control where

\[
\begin{align*}
\lambda^*_0 &= \frac{e_0}{m \sum_{i=0}^{m} e_i} \\
\lambda^*_i &= \frac{\lambda^*_0}{e_0} e_i \\
&\quad (1 \leq i \leq m)
\end{align*}
\] (4.12)

**Proof**

Since the signal-settings \( \{\lambda^*_i\} \) are implemented under fixed-time control, the proportion of all time for which stage \( i \) \((1 \leq i \leq m)\) runs in these circumstances is just \( \lambda^*_i \) and the proportion of all time which is occupied

\[
\sum_{i=0}^{m} \lambda^*_i = 1
\]
by stage transitions is $\lambda^*_0$. The proportion of all time for which stage $\lambda^*_0$ runs when the signal-settings are $\{\lambda_i\}$ under the appropriate circumstances of implementation is $\varepsilon_i / \sum_{\ell=0}^{m} \varepsilon_{\ell}$ and the proportion of all time which is occupied by stage transitions is $\varepsilon_0 / \sum_{\ell=0}^{m} \varepsilon_{\ell}$. If the signal-settings are capacity-equivalent, then the corresponding proportions can be equated, giving

$$\lambda^*_i = \frac{\varepsilon_i}{\sum_{\ell=0}^{m} \varepsilon_{\ell}} \quad (0 \leq i \leq m) \quad (4.13)$$

and in particular

$$\lambda^*_0 = \frac{\varepsilon_0}{\sum_{\ell=0}^{m} \varepsilon_{\ell}} .$$

Rearranging this to give an expression for $\varepsilon_{\ell} / \sum_{\ell=0}^{m} \varepsilon_{\ell}$ and substituting this into (4.13) gives

$$\lambda^*_i = \frac{\lambda^*_0}{\varepsilon_0} \varepsilon_i \quad (0 \leq i \leq m) .$$

The value of each $\lambda^*_i$ $(0 \leq i \leq m)$ is unique since if $\{\lambda_i\}$ together with the circumstances of implementation were also capacity-equivalent to $\{\lambda^*_i\}$ under fixed-time operation, then arguing as before,

$$\lambda^*_i = \frac{\varepsilon_i}{\sum_{\ell=0}^{m} \varepsilon_{\ell}} \quad (0 \leq i \leq m)$$

$$= \lambda^*_i \quad (0 \leq i \leq m) \quad \text{from (4.13)} \quad [\text{]}$$
The capacities of the streams of traffic at a junction are not, in general, identical when the junction is subject to different control methods with capacity-equivalent signal-settings. The following results quantify any such discrepancies and identify some circumstances in which the capacities are identical.

Lemma 4.10

Suppose that the signal-settings \( \lambda_i \) (0 ≤ i ≤ m) give rise to mean durations \( \epsilon_i \) (0 ≤ i ≤ m) and probabilities \( P_n \) (k < n ≤ m) under some circumstances of implementation and similarly that the signal-settings \( \lambda_i^+ \) (0 ≤ i ≤ m) give rise to \( \epsilon_i^+ \) (0 ≤ i ≤ m) and \( P_n^+ \) (k < n ≤ m) under some other circumstances of implementation. If \( \{\lambda_i\} \) are capacity-equivalent to \( \{\lambda_i^+\} \), then the difference between \( K_j^- \), the capacity of stream \( j \) (1 ≤ j ≤ M) under the signal-settings \( \{\lambda_i\} \) and the corresponding capacity \( K_j^+ \) is given by

\[
K_j^+ - K_j = \left( \epsilon_i + \frac{1}{m} \sum_{i=1}^{k} a_{10j} \lambda_{10}^+ \right) \sum \frac{1}{m} \sum_{i=0}^{\epsilon_i} \sum_{\epsilon_i}^{\epsilon_i^+} L + \frac{1}{m} \sum_{i=0}^{\epsilon_i} \sum_{\epsilon_i}^{\epsilon_i^+} L
\]

\[
+ \frac{m}{n=k+1} \left( \frac{P_n^+}{m} - \frac{P_n}{m} \right) \left( \sum_{i=k+1}^{n-1} a_{10j} \lambda_{10}^+ + a_{n1j} \lambda_{n1}^+ \right) s_j \quad \text{(4.14)}
\]

Proof

From the results of theorem 4.7 and lemma 4.5,

\[
K_j^+ = \left( \sum_{i=1}^{m} a_{ij} \epsilon_i + \sum_{i=1}^{k} a_{10j} \lambda_{10}^+ \right) + \sum_{n=k+1}^{m} \frac{P_n^+}{m} \left( \sum_{i=k+1}^{n-1} a_{10j} \lambda_{10}^+ + a_{n1j} \lambda_{n1}^+ \right) s_j \quad \text{(1 ≤ j ≤ m)} \quad \text{(4.15)}
\]

and

\[
K_j = \left( \sum_{i=1}^{m} a_{ij} \epsilon_i + \sum_{i=1}^{k} a_{10j} \lambda_{10}^+ \right) + \sum_{n=k+1}^{m} \frac{P_n}{m} \left( \sum_{i=k+1}^{n-1} a_{10j} \lambda_{10}^+ + a_{n1j} \lambda_{n1}^+ \right) s_j \quad \text{(1 ≤ j ≤ m)} \quad \text{(4.16)}
\]
Subtracting (4.16) from (4.15) and using the capacity-equivalence relationships
\[ \frac{e_i}{\mathcal{L} = 0} e_i^+ = \frac{e_i^+}{\mathcal{L} = 0} e_i \quad (1 \leq i \leq m) \]

Together with the fact that
\[ a_{ij} \quad (1 \leq i \leq m) \quad (1 \leq j \leq M) \]
are constants, gives (4.14).

The most useful form of this result is given in the next result.

**Corollary 4.11**

Suppose that the signal-settings \( \{ \lambda_i^+ \} \), when implemented under fixed-time control, are capacity-equivalent to \( \{ \lambda_i \} \). Then

\[
K_j^+ - K_j = \left\{ \frac{e_0^+ - \lambda_0}{e_0} + \sum_{i=1}^{k} a_{i0} \lambda_{10} + \sum_{n=k+1}^{m} \prod_{n=n}^{p} \left[ \sum_{i=k+1}^{m} a_{i0} \lambda_{10} + \frac{\lambda_0}{e_0} \left( \sum_{i=k+1}^{n-1} a_{i0} \lambda_{10} + \lambda_n \right) \right] \right\} s_j \quad (4.17)
\]

\( (1 \leq j \leq m) \)

**Proof**

Since \( \{ \lambda_i^+ \} \) are implemented as fixed-time signal-settings, \( e_i^+ \lambda_i^+ \quad (0 \leq i \leq m) \)

So

\[
\frac{1}{m} \sum_{\mathcal{L} = 0} e_i^+ = \frac{1}{m} \sum_{\mathcal{L} = 0} \lambda_i^+ = 1 \quad (4.18)
\]

Now

\[
\frac{e_0}{m} \sum_{\mathcal{L} = 0} e_i^+ = \frac{e_0^+}{m} \sum_{\mathcal{L} = 0} \lambda_i^+ = \lambda_0^+ \]
so \[ \frac{1}{\Sigma_{\mathcal{L}=0}^{m} \mathcal{E}} = \frac{\lambda^+}{\varepsilon_0} . \] (4.19)

Using (4.18) and (4.19) in (4.14) together with \( P_n^t = \delta_{nm} \), gives

\[ K_j^+ - K_j = \left[ \left( 1 - \frac{\lambda^+_0}{\varepsilon_0} \right) \sum_{i=1}^{k} a_{i0j} \lambda_{i0} + \sum_{i=k+1}^{m} a_{i0j} \lambda_{i0} - \right. \]
\[ - \frac{m}{\sum_{n=k+1}^{m} \mathcal{P}} \sum_{i=k+1}^{m} \mathcal{P} \left[ \sum_{n=0}^{n-1} \frac{a_{ij0} \lambda_{ij0} + a_{nij} \lambda_{nij}}{\lambda_{i0} + \lambda_{n1}} \right] \right] \mathcal{S}_j \quad (1 \leq j \leq m) \]

Rearranging and using \( \sum_{n=k+1}^{m} \mathcal{P} = 1 \) gives (4.17).

The following results identify some circumstances in which there is no discrepancy between the capacities of streams under capacity-equivalent signal-settings.

**Corollary 4.12**

Under the hypotheses of lemma 4.10, if \( P_n^t = P_n \) \((k \leq n \leq m)\), then \( K_j^+ = K_j \) \((1 \leq j \leq M)\).

**Proof**

From the result of lemma 4.3,

\[ \varepsilon^+_0 = \sum_{i=1}^{k} \lambda_{i0} + \sum_{n=k+1}^{m} \mathcal{P} \left( \sum_{i=k+1}^{m} \lambda_{i0} \right) \]
\[ = \sum_{i=1}^{k} \lambda_{i0} + \sum_{n=k+1}^{m} \mathcal{P} \left( \sum_{i=k+1}^{m} \lambda_{i0} \right) \quad \text{(by hypothesis)} \]
\[ = \varepsilon_0 . \]
Since \( \{\lambda^+_i\} \) and \( \{\lambda_i\} \) are capacity-equivalent,

\[
\frac{1}{m} \sum_{\ell=0}^{+\infty} e^\ell = \frac{e_0}{m} \frac{1}{\sum_{\ell=0}^{+\infty} e^\ell} = \frac{1}{m} \frac{p_n^+}{\sum_{\ell=0}^{+\infty} e^\ell} \quad (4.20)
\]

and

\[
\frac{1}{m} \sum_{\ell=0}^{+\infty} e^\ell = \frac{p_n}{\sum_{\ell=0}^{+\infty} e^\ell} \quad (4.21)
\]

Using (4.20) and (4.21) in (4.14) gives \( K_j^+ - K_j^- = 0 \) \((1 \leq j \leq M)\).

**Corollary 4.13**

If \( \{\lambda^+_i\} \) are capacity-equivalent to \( \{\lambda_i\} \) and there are \( m \) stages in every cycle in both circumstances, then \( K_j^+ = K_j^- \) \((1 \leq j \leq M)\).

**Proof**

Since there are exactly \( m \) stages in each cycle in both cases,

\[
p_n^+ = \delta_{nm}
\]

so corollary 4.12 can be applied.

**4.2.4 Discussion.**

The results of lemma 4.10 and corollary 4.11 show that any discrepancy between the capacity of a stream under different capacity-equivalent signal-settings results entirely from the difference in utilisation of the transition periods. The sizes of any such discrepancies can be bounded above by the relationships

\[
|K_j^+ - K_j^-| \leq (|a_{0j}e_0^+| + |a_{0j}e_0^-|)s_j \quad (1 \leq j \leq M) \quad (4.22)
\]
In most cases \( |a_{ij}e^0| < \lambda_j \) as streams normally receive the majority of their effective green time during stages rather than during transition periods. In the unusual circumstances that a stream receives no extra green time in addition to that occurring during stages, (4.22) shows that the capacity is the same under capacity-equivalent signal-settings. A particular consequence of corollary 4.13 is that the capacity of each stream at a junction where priority is implemented by extension or hurry-call is equal to the capacity achieved under capacity-equivalent signal-settings implemented under fixed-time control.

The reason for the introduction of capacity-equivalence is that the problem of finding signal-settings which are suitable for implementation with a priority scheme can be stated conveniently in terms of this concept. Suppose that, either as a result of calculations or direct experience, some signal-settings are known to provide adequate capacity for each stream of traffic at a junction in the absence of priority for buses. One possible objective when calculating signal-settings for use with priority is to find ones which provide similar capacities for each stream to those arising from the known set. If they exist, signal-settings which are capacity-equivalent when implemented with priority for buses to the original signal-settings will achieve this objective.

The advantage of this formulation of the problem of finding suitable signal-settings is that it enables them to be calculated from ordinary ones without detailed reference to the layout of the junction or characteristics of the streams of traffic. This problem is investigated in detail in Section 4.5. Before then, methods for calculating signal-settings which maximise the common factor by which the mean arrival rates in all streams could be multiplied without causing oversaturation in any stream are
reviewed in Section 4.3. In Section 4.4, explicit formulae are derived to calculate $P_n (k<n<m)$ and $e_i (0<i<m)$ from signal-settings $\{\lambda_i\}$ when these are implemented with priority by extension, hurry-call or recall, both with and without inhibition. These quantities can then be used to estimate the capacity of each stream according to theorem 4.7 or to calculate capacity-equivalent fixed-time signal-settings according to lemma 4.9.

4.3 Maximisation of capacity at a junction

4.3.1 Introduction.

The problem of finding suitable signal-settings to implement at a junction is of considerable practical importance. In this section, some methods are reviewed which have been devised for use at junctions controlled by fixed-time or vehicle-actuated traffic signals. Various authors have used methods of mathematical programming to investigate this problem. In order to do this, an objective must be stated. The objectives considered by several of these authors are equivalent in the sense that they give rise to identical signal-settings. These objectives, which are defined in terms of the capacity of the streams which discharge at a junction, are the topic of this section. Some other objectives which are defined in terms of delay and other related quantities are discussed in Chapter 5.

Each of the procedures considered here is subject to an equivalent set of constraints which are stated together. The signal settings which satisfy all these constraints simultaneously comprise the feasible region: any problem for which the feasible region is not empty is said to be feasible. The constraints can be stated as follows.
Practical capacity constraints

\[ p_j K_j - \kappa_j q_j \geq 0 \quad (1 \leq j \leq M) \]  \hspace{1cm} (4.23)

Minimum stage duration constraint

\[ \lambda_i - \frac{G_i}{L} \lambda_0 \geq 0 \quad (1 \leq i \leq m) \]  \hspace{1cm} (4.24)

Maximum (specified) cycle-time constraints

\[ \lambda_0 - \frac{L}{G_0} \geq (=) 0 \]  \hspace{1cm} (4.25)

where \( G_0 \) is the maximum permissible (specified) cycle time. Finally, from the definition of \( \lambda_i \) (0 \( \leq i \leq m \)),

\[ \sum_{i=0}^{m} \lambda_i = 1 \]  \hspace{1cm} (4.26)

4.3.2 Allsop's formulation.

Allsop (1972b) proposed a method for finding signal-settings which considers only the more heavily loaded streams of traffic. Although the signal-settings are used explicitly only in the constraints and not in the objective function, the solution method used by Allsop does yield suitable values. Furthermore, information is available from this method concerning the sensitivity of the solution to changes such as relaxation of constraints and variations in saturation departure rates.

This formulation finds the maximum common factor, \( \mu \), by which the mean arrival rates in all streams can be multiplied whilst the problem remains feasible. This is posed as a linear programming problem (Vajda, 1961) as follows.
Maximise \[ \mu, \lambda_i \quad (0 \leq i \leq m) \]

Subject to

\[
\begin{align*}
\sum_{i=0}^{m} a_{ij} \lambda_i - \mu q_j & \geq 0 \quad (1 \leq j \leq M) \\
\lambda_i - \frac{G_i}{L} \lambda_0 & \geq 0 \quad (1 \leq i \leq m) \\
\lambda_0 - \frac{L}{G_0} & \geq (=) 0 \\
\sum_{i=0}^{m} \lambda_i & = 1
\end{align*}
\]

(4.27)

Allsop (1972b) solved this problem by the simplex method (Vajda), which yields the following information:

(a) \( \mu^* \), the maximum value of \( \mu \) for which the problem is feasible

(b) \( \lambda_i^* (0 \leq i \leq m) \), signal-settings which will accommodate the flows

\( \mu^* q_j \quad (1 \leq j \leq M) \)

(c) the values of the dual and slack variables (Vajda) associated with each of the constraints.

Further analysis showed how variations in the maximum or specified cycle time, minimum green time or arrival and saturation departure rates could cause changes in the maximum flow multiplier. Explicit formulae were given for the size of these changes. These are valid provided that the changes are sufficiently small that no constraint which was previously satisfied as a strict inequality is violated and that the cycle-time constraint is binding.

The last proviso, which was used only in the derivation of the formula for the effect of a change in a binding minimum green constraint, is normally satisfied. However, Davies, Jamieson and Reid (1980) gave a practical example where increasing the cycle time could reduce the maximum
flow multiplier for which the problem remains feasible. This example depended on an unusual geometrical arrangement at the junction which enabled each stream to have right of way throughout all of the transition periods contiguous to the stages during which it has right of way. As a consequence of this, the transition periods are better used than the stages. The optimal cycle time when the arrival and saturation departure rates in each stream are similar was shown to be equal to the sum of the durations of the transition periods, with all stage durations set to zero. Similar results will occur at any junction where the transition periods are sufficiently well used.

In cases like the one described above, the cycle-time constraint will not be binding at the solution to (4.27). As a consequence, Allsop's (1972b) formula for the effects of a change in a binding minimum green constraint is not applicable. A similar but more intricate analysis has been used (Heydecker, 1983) to derive a more general formula which applies to this case.

To each constraint which is satisfied as a strict inequality, there corresponds a non-zero slack variable. In the case of the cycle-time constraint, the value of this variable can be used to calculate the largest possible reduction in the maximum permissible cycle time which will not cause any reduction in the maximum flow multiplier. Similarly, in the case of a minimum stage duration constraint, it corresponds to the largest possible increase in the minimum permissible stage duration which will not alter the solution.

A rather more informative case is that of non-zero slack variables associated with capacity constraints. Here, the value of the slack variables can be used to calculate the greatest amount by which the mean arrival rate can be increased or the saturation departure rate can be decreased for that stream without affecting the value of the solution, $\mu^\#$. 
4.3.3 Alternative formulations.

The objective of Allsop's (1972b) formulation of the signal-setting problem is to maximise $\mu$, the common flow multiplier. If the resulting maximal value $\mu^*$ is greater than unity, then the junction is said to have a reserve capacity of $100(\mu^*-1)$ per cent or $100(1-1/\mu^*)$ per cent of the present flows. If $\mu^*$ is less unity, then the junction is said to be overloaded by $100(1-\mu^*)$ per cent or $100(1/\mu^*-1)$ per cent of the present flows. Since there is no requirement in the linear programme (4.27) that the junction should not be overloaded, Allsop's (1972b) formulation is applicable even in cases where some streams are necessarily oversaturated. For any problem where $\mu^* > 1$, Allsop's (1972b) objective is equivalent to maximising the reserve capacity and similarly if $\mu^* < 1$, to minimising the overload.

Ohno and Mine (1973a) proposed a similar formulation to find suitable signal-settings. In the first instance, they formulated the problem in terms of the proportion of the capacity of each stream which corresponds to the mean arrival rate there. This quantity is known as the degree of saturation of a stream and is denoted by $x_j$ for stream $j$. According to the result of corollary 4.8,

$$x_j = \frac{k_j q_j}{\sum_{i=0}^{m} a_{ij} \lambda_i} \quad (1 \leq j \leq M)$$

(4.28)
Ohno and Mine's (1973a) formulation is to

\[
\text{Minimise} \quad \text{Maximum} \quad x_j \\
\lambda_i \quad (0 \leq i \leq M) \quad j \quad (1 \leq j \leq M)
\]

Subject to

\[
\begin{align*}
\lambda_i - \frac{G_i}{L} \lambda_0 & \geq 0 \quad (1 \leq i \leq M) \\
\lambda_0 - \frac{L}{C_0} & \geq (=) 0 \\
\sum_{i=0}^{m} \lambda_i &= 1 \\
x_j & \geq 0 \quad (1 \leq j \leq M)
\end{align*}
\]

(4.29)

An equivalent formulation was preferred on the grounds that the constraints \(x_j \geq 0\) \((1 \leq j \leq M)\) are computationally inconvenient. For each stream \(j\) \((1 \leq j \leq M)\), \(z_j\), the reciprocal of the degree of saturation is called the degree of undersaturation. The new formulation is to

\[
\text{Maximise} \quad \text{Minimum} \quad z_j \\
\lambda_i \quad (0 \leq i \leq M) \quad j \quad (1 \leq j \leq M)
\]

Subject to

\[
\begin{align*}
\lambda_i - \frac{G_i}{L} \lambda_0 & \geq 0 \quad (1 \leq i \leq M) \\
\lambda_0 - \frac{L}{C_0} & \geq (=) 0 \\
\sum_{i=0}^{m} \lambda_i &= 1
\end{align*}
\]

(4.30)

This problem was solved by mixed-integer programming methods. Ohno and Mine (1973a) showed that if \(x^*\) is the solution to (4.29) and \(z^*\) to (4.30), then the two problems are equivalent in the sense that \(z^* = 1/x^*\).

Akçelik (1978) showed how the value of \(x^*\) could be calculated directly when the junction under consideration is of a sufficiently simple design. The resulting value was then interpreted as a measure of performance. Akçelik (1978) recommended that a maximum acceptable value for this be selected from the range \([0.8, 0.95]\), depending on the importance of the junction. This is in general agreement with Webster and Cobbe's recommendation that the maximum acceptable degree of saturation for a stream of traffic should be taken as 0.9.
Ohno and Mine's (1973a) formulation is to

\[
\begin{align*}
\text{Minimise} & \quad \text{Maximum} & x_j \\
\lambda_i & \quad (0 \leq i \leq m) & j \quad (1 \leq j \leq M)
\end{align*}
\]

Subject to

\[
\begin{align*}
\lambda_i - \frac{G_i}{L} \lambda_0 & \quad \geq 0 & (1 \leq i \leq m) \\
\lambda_0 - \frac{L}{G_0} & \quad \geq (=) 0 \\
\sum_{i=0}^{m} \lambda_i & \quad = 1 \\
x_j & \quad \geq 0 & (1 \leq j \leq M)
\end{align*}
\]

(4.29)

An equivalent formulation was preferred on the grounds that the constraints \(x_j \geq 0\) \((1 \leq j \leq M)\) are computationally inconvenient. For each stream \(j\) \((1 \leq j \leq M)\), \(z_j\), the reciprocal of the degree of saturation is called the degree of undersaturation. The new formulation is to

\[
\begin{align*}
\text{Maximise} & \quad \text{Minimum} & z_j \\
\lambda_i & \quad (0 \leq i \leq m) & j \quad (1 \leq j \leq M)
\end{align*}
\]

Subject to

\[
\begin{align*}
\lambda_i - \frac{G_i}{L} \lambda_0 & \quad \geq 0 & (1 \leq i \leq m) \\
\lambda_0 - \frac{L}{G_0} & \quad \geq (=) 0 \\
\sum_{i=0}^{m} \lambda_i & \quad = 1
\end{align*}
\]

(4.30)

This problem was solved by mixed-integer programming methods. Ohno and Mine (1973a) showed that if \(x^*\) is the solution to (4.29) and \(z^*\) to (4.30), then the two problems are equivalent in the sense that \(z^* = 1/x^*\).

Akçelik (1978) showed how the value of \(x^*\) could be calculated directly when the junction under consideration is of a sufficiently simple design. The resulting value was then interpreted as a measure of performance. Akçelik (1978) recommended that a maximum acceptable value for this be selected from the range \([0.8, 0.95]\), depending on the importance of the junction. This is in general agreement with Webster and Cobbe's recommendation that the maximum acceptable degree of saturation for a stream of traffic should be taken as 0.9.
Allsop's (1972b) analysis provided only for cases where each $a_{ij}$ ($1 \leq i \leq m$; $1 \leq j \leq m$) takes the value 0 or 1. Yagar (1974) extended this treatment to include the possibility of non-integer values for these variables. This has application to streams which are partially obstructed by others during part of the time for which they have right of way. Allsop (1977a) described an iterative method which incorporates opposed turning manoeuvres and mixed streams of traffic including opposed turners. Ohno and Mine (1973a) gave approximate formulae to calculate suitable values for $a_{ij}$ ($1 \leq i \leq m$) when stream $j$ contains only opposed turners.
4.4 Capacity at a junction where there is priority for buses

4.4.1 Introduction and preliminaries.

In this section, a detailed analysis is given of the duration and number of stages in the cycle at a junction where responsive priority is implemented for buses by any combination of the usual priority rules. The results of this analysis can be used to determine the capacity of each stream at a junction by use of the analysis developed in Section 4.2.

While the results presented here are specific to the priority rules investigated, the method of analysis can be applied to other rules. These results are derived under the assumption that the arrival of buses at the detector forms a Poisson process. Allsop (1977b) considered this and the more regular arrival pattern described below. He showed that any differences between the mean durations of the stages occurring with each of these arrival processes are of second order in the mean bus arrival rate.

The priority rules considered here are extension, hurry-call and recall. Each is considered separately, but since it operates in a distinct way, the results of the analysis of extension can be combined in a natural way with those of the analysis of either of the other two methods for junctions where priority rules are implemented in combination. The inhibition rule is also analysed separately in a manner which permits the use of the results with those of either of the analyses of hurry-call and recall.

This is a convenient point at which to discuss some of the ramifications of the model adopted for arrivals of buses at the detector. Writing $\tilde{p}'(t)$ for the probability that exactly $n$ events occur in a period of duration $t$, a Poisson process with mean rate $b$ has the probability distribution
\[ p_n(t) = \frac{e^{-bt} (bt)^n}{n!} \quad (n \geq 0) \]  

The following results, which are consequences of (4.31), are stated here for convenient reference. Since they are standard results (see, for example, Haight, 1963) no proofs are presented here.

**Lemma 4.14**

The mean and variance of \( n \), the number of events of a Poisson process of rate \( b \) in a period of duration \( t \), is given by

\[
\begin{align*}
E(n|t) &= bt \\
\text{Var}(n|t) &= bt
\end{align*}
\]  

When the process under consideration is that of vehicular arrivals, the duration the period between two successive events is called a **time headway**.

**Lemma 4.15**

The time headways between vehicles which arrive according to a Poisson process with rate \( b \) are distributed according to the negative exponential probability density function

\[ p(h) = be^{-bh} \quad (h \geq 0) \]

The distribution of partial headways formed by the time from an independent event to the next event of a Poisson process of rate \( b \) is distributed with the same probability density function, (4.33).

The mean and variance of the time headways are given by

\[
\begin{align*}
E(h) &= 1/b \\
\text{Var}(h) &= 1/b^2
\end{align*}
\]

A useful indication of the variability of a process is provided by the ratio of the variance of the number of events in a given sampling interval
to the mean number in the same interval. This quantity is called the index of dispersion of the process and is denoted by the symbol $I_\alpha$. An immediate consequence of lemma 4.14 is that for a Poisson process, $I_\alpha = 1$, regardless of the duration of the sampling interval.

Allsop (1977b) introduced the quasi-regular model of bus arrivals as an alternative to the Poisson model adopted here. In the quasi-regular model no more than one bus can arrive at the detector during each cycle and only the proportion of cycles during which a bus arrives is specified. The instants at which these arrivals occur are distributed uniformly over the duration of an uninterrupted cycle. A special case of this occurs when the probability of bus arrivals in consecutive cycles are independent and identical: this is called the Bernoulli arrival process. Thus some indication of the possibilities encompassed by the quasi-regular arrival process are indicated by considering the Bernoulli process.

**Lemma 4.16**

The mean and variance of $n$, the number of events of a Bernoulli process of rate $b$, occurring during $N$ cycles are given by

$$
\begin{align*}
E(n|N) &= Nb\bar{t} \\
\text{Var}(n|N) &= Nb\bar{t}(1-b\bar{t})
\end{align*}
$$

(4.35)

where $\bar{t}$ is the mean duration of a cycle.

The index of dispersion of this process cannot be calculated so readily as it was for the Poisson process. If the duration of a cycle were fixed, then the index of dispersion of a Bernoulli process calculated over any whole number of cycles would be given by $I_\alpha = 1-b\bar{t}$. If the duration of a cycle is not fixed, then this result is only approximate, but its accuracy improves as the number of cycles comprising the sampling interval increases. Finally, if the sampling interval does not comprise a whole
number of cycles, then the error in the estimates given by (4.35) cannot exceed that due to the arrival of 2 extra vehicles. In any case, 
\( \bar{a}^2 (T) \to 1 - bt \) as \( T \to \infty \), where \( T \) denotes the duration of the sampling interval. Thus the index of dispersion of a Bernoulli process differs from that of a Poisson process of the same rate by a term of first order in the rate.

In the remainder of this section, formulae are derived for various statistics of the duration of the stages at a junction where there is responsive priority for buses. Only the mean durations are required to apply the analysis of Section 4.2. However, other statistics are required by the analysis presented in Chapter 5, so they are derived here for convenience. The expressions given for the mean durations of the stages are similar to those given by Allsop (1977b) but are derived under less restrictive hypotheses concerning the durations of the extension and transition periods. Some general results are given which are more powerful than is justified by their use in this chapter. These will be used again in Chapter 5.

The following additional notation will be used in the remainder of this thesis. Let

- \( T_k \) be the length of time for which stage \( k \) is held when a bus is granted priority by extension (s)
- \( \beta/c \) be the mean rate of bus arrivals at the detector (bus/s)
- \( \tilde{p}_0 (\xi) \) be the probability that no buses arrive at the detector during a period of duration \( \xi c \).

Because of the assumption that the arrivals of buses at the detector form a Poisson process,

\[ \tilde{p}_0 (\xi) = \exp(-\beta \xi) \quad (4.36) \]
The assumption that the duration of an extension period is less than the time from the end of the last stage in one cycle to the end of the unextended green period for buses in the next cycle can now be written as

\[ \tau \leq \sum_{i=1}^{k} \lambda_{i} + \sum_{i=1}^{k-1} \lambda_{i0} + \text{Minimum } \lambda_{i1} \]  

(4.37)

This ensures that buses detected before stage 1 is called do not cause stage \( k \) to be extended. In particular, no bus which is granted priority by hurry-call or recall will also require an extension. Condition (4.37) is weaker than Allsop's (1977b) corresponding condition (6b):

\[ \tau \leq \lambda_{k} + \lambda_{k-1,0} \]

4.4.2 Priority by extension.

The action of the extension priority rule is restricted to the last part of the period during which buses receive right of way and depends on the time at which the bus arrives at the detector as follows.

Arrival less than \( \tau_{C} \) before the time at which stage \( k+1 \) would be called if no buses were detected: stage \( k \) is extended until \( \tau_{C} \) after the detection of the bus.

Arrival at any other time: no effect.

This rule forces the controller to run stage \( k \) until a time headway greater than \( \tau_{C} \) occurs in the bus arrival process. Possible variations on this rule include the restriction of priority by extension to no more than one bus in each cycle or imposing a maximum permissible duration on stage \( k \) giving rise to a time after which no more buses will be granted extensions.
The extension rule exhibits a memoryless property which can be exploited in order to estimate the moments of the duration of stage \( k \). This arises because the state of the controller depends on the time of detection only of the most recent bus. A more general system of this kind will be analysed and the required results for priority by extension will then be deduced.

Consider a two-state system which enters state 1 at time \( t=0 \) and which will make an instantaneous transition to state 0 at the end of a trial period unless a prolonging event occurs during the trial period. In the latter case, a further trial period is initiated, either immediately, or in some cases after a further time interval has elapsed. Cases where the start of the next trial period is delayed arise in Chapter 5. The interval between the initial instants of two successive trial periods is called the period of prolongation. A system which behaves in this manner with constant probability of prolongation and for which the duration of successive periods of prolongation are independent and identically distributed random variables will be called a 2-state M–R system. This is because it exhibits some of the properties of a Markov-renewal process (Činlar, 1969).

Principal quantities of interest in the study of M–R systems are the moments of the duration of the sojourn in each state. The next result relates these moments to those of the distribution of the periods of prolongation. It is followed immediately by an application to find the first two moments of the duration of stage \( k \) when priority for buses is provided by extension.

**Theorem 4.17**

Consider a 2-state M–R system where the probability that a prolonging event occurs during a single trial period is \( P_r \) and the moments of the period of prolongation are \( \nu \), \( c^n \) (\( n \geq 1 \)). The moments \( \mu_n \) of the duration of the
part of the sojourn in state 1 in excess of a single trial period satisfy the recursive relationship

$$\mu_n = \left( \frac{P}{1-P} \right)^n \sum_{j=1}^{n} \binom{n}{j} \nu_j \mu_{n-j} \quad (n \geq 1)$$

(4.38)

where \( \binom{n}{j} \) denotes the binomial coefficient \( \frac{n!}{j!(n-j)!} \).

Proof

Consider the system at the initial instant of a trial period. There are two kinds of event which can terminate that trial period:

(a) with probability \( P \), a prolonging event occurs, causing the initiation of a new trial period, the initial instant of which is separated from that of the previous one by a prolonging period;

(b) with probability \( (1-P) \), no prolonging event occurs and the sojourn in state 1 terminates at the end of the trial period.

Thus the only way in which the sojourn in state 1 can terminate is as a result of event (b): this will add one trial period to the duration of the sojourn, so the quantities of interest are the moments of the sum of the prolonging periods caused by events of type (a). Let \( t \) denote the duration of a prolonging period and let \( T_i \) denote the duration of the sojourn in state 1 after \( i \) prolonging events have occurred \((i \geq 0)\) less the duration of one trial period. Then

$$T_0^n = \begin{cases} 0 & \text{with probability } (1-P) \\ (t+T_i)^n & \text{with probability } P \end{cases} \quad (n \geq 1)$$

(4.39)

Taking expectations,
\[ E(T_0^n) = \mathbb{P}_x E[(t+T_1)^n] \]
\[ = \mathbb{P}_x E\left[ \sum_{j=0}^{n} \binom{n}{j} t^j (T_1^{n-j}) \right] \]
\[ = \mathbb{P}_x \sum_{j=0}^{n} \binom{n}{j} E(t^j) E(T_1^{n-j}) \]

since \( t \) and \( T \) are mutually independent \hspace{1cm} (4.40)

Now \( E(t^j) = \mu_n^j \) \( (j \geq 0) \) and, because events in each trial period are independent of those in all others,
\[ E(T_1^{n-j}) = E(T_0^{n-j}) \]
\[ = \mu_n^{n-j} \] \( (n \geq j) \).

Applying these identities to (4.40) gives
\[ \mu_n^c^n = \mathbb{P}_x \sum_{j=0}^{n} \binom{n}{j} \nu_j \mu_n^{n-j} c^n \]
\[ = \mathbb{P}_x (\mu_n + \sum_{j=1}^{n} \nu_j \mu_n^{n-j}) c^n \] \hspace{1cm} (n \geq 1) \hspace{1cm} (4.41)

so \[ \mu_n = \left( \frac{\mathbb{P}_x}{1-\mathbb{P}_x} \right) \sum_{j=1}^{n} \nu_j \mu_n^{n-j} \] \hspace{1cm} (n \geq 1) \hspace{1cm} (4.38)

Lemma 4.18

Let \( \nu_n \) be the \( n \)th moment of the truncated negative exponential distribution on \( [0, \omega] \) defined by the probability density function

\[ p(\xi) = \begin{cases} \frac{\beta e^{-\beta \xi}}{\int_0^{\omega} e^{-\beta \xi'} \, d\xi'} & (0 \leq \xi \leq \omega) \\ 0 & \text{otherwise} \end{cases} \]

Then \[ \nu_n = \frac{n!}{\beta^n} \left[ 1 - e^{-\beta \omega} \sum_{j=0}^{n} (\beta \omega)^j \right] \sum_{j=0}^{n} \frac{(\beta \omega)^j}{j!} \] \hspace{1cm} (n \geq 1) \hspace{1cm} (4.42)

Proof

Let \( I_n = \int_0^\omega \xi^n e^{-\beta \xi} \, d\xi \) \( (n \geq 0) \). Then

\[ \nu_n = \int_0^\omega \xi^n \frac{\beta e^{-\beta \xi}}{\int_0^\omega e^{-\beta \xi'} \, d\xi'} \, d\xi \]
\[ = \frac{I_n}{I_0} \] \hspace{1cm} (4.43)
Now \[ I_0 = - \frac{\omega}{\beta} \frac{1}{\beta} e^{-\beta \xi} \]

\[ = \frac{1}{\beta} [1 - e^{-\beta \omega}] \quad (4.44) \]

and \[ I_n = - \frac{\omega}{\beta} \left( \frac{n \xi}{\beta} e^{-\beta \xi} \right) + \int_{0}^{\xi} e^{-\beta \xi} d\xi \quad (by \ parts) \]

\[ = - \frac{\omega}{\beta} \left( \frac{n \xi}{\beta} e^{-\beta \xi} \right) + \frac{n}{\beta} \xi^{n-1} \quad (n \geq 1) \quad (4.45) \]

Applying (4.45) recursively \( n \) times and (4.44) to the remaining integral gives

\[ I_n = - \frac{\omega}{\beta} \frac{n \xi}{\beta} e^{-\beta \xi} + \frac{n \xi^{n-1}}{\beta^2} e^{-\beta \xi} + \ldots + \frac{n!}{\beta^{n+1}} e^{-\beta \xi} \]

\[ = - \frac{\omega}{\beta} \frac{n!}{\beta^{n+1}} e^{-\beta \xi} \sum_{j=0}^{n} \frac{(\beta \xi)^j}{j!} \quad (n \geq 0) \quad (4.46) \]

Substituting (4.46) into (4.43) gives (4.42).

---

**Corollary 4.19**

Suppose that priority by extension is provided for buses arriving at a mean rate of \( \beta / c \) and that stage \( k \) is held for a period of duration \( \tau_c \) after the detection of each bus. The first two moments, \( \mu_1 c \) and \( \mu_2 c^2 \), of the time by which the duration of stage \( k \) exceeds its unextended duration are given by
\[
\begin{align*}
\mu_1 c &= \left[ \frac{1 - e^{-\beta T(1+\beta T)}}{\beta e^{-\beta T}} \right] c \\
\mu_2 c^2 &= \frac{2}{\beta^2 c e^{-\beta T}} \left[ 1 - e^{-\beta T(1+2\beta T)} + e^{-2\beta T(\beta T+(\beta T)^2/2)} \right] c^2
\end{align*}
\] (4.47)

**Proof**

The process of extending stage \( k \) constitutes a 2-state M-R system in which state 1 corresponds to the controller running stage \( k \) and state 0 corresponds to the controller running any other stage or transition. The trial period is the interval during which the arrival of a bus at the detector will cause stage \( k \) to be extended and any such extension corresponds to a prolonging event. In this case, the period of prolongation is the time between the start of the trial period and the occurrence of the prolonging event. Independence of events in successive trial periods is an immediate consequence of the result of lemma 4.15. Thus the result of theorem 4.17 can be applied once the appropriate forms for \( P_x, \nu_1 \) and \( \nu_2 \) have been established.

In this case, a prolonging event occurs whenever a bus arrives at the detector during a trial period of duration \( \tau_0 \). Thus

\[
\begin{align*}
P_x &= 1 - \tilde{p}_0(\tau) \\
&= 1 - e^{-\beta T} \\
&\text{from (4.36)}
\end{align*}
\] (4.48)

From the result of lemma 4.15, the time from the start of a trial period to the time at which the next bus arrives at the detector is distributed with the negative exponential probability density function. Thus the duration of the periods of prolongation are independent and identically distributed with the truncated negative exponential probability density function on \([0, \tau_0]\). From the result of lemma 4.18,

\[
\begin{align*}
\nu_1 &= \left[ 1 - e^{-\beta T(1+\beta T)} \right] \\
&= \beta \left[ 1 - e^{-\beta T} \right] \\
&\text{from (4.49)}
\end{align*}
\]
Using (4.48) and (4.49) in (4.38) with $n=1$ gives

$$
\mu_1 = \frac{1-e^{-\beta \tau}}{e^{-\beta \tau}} \frac{[1-e^{-\beta(1+\beta \tau)}]}{\beta [1-e^{-\beta \tau}]} = \frac{[1-e^{-\beta(1+\beta \tau)}]}{\beta e^{-\beta \tau}} \quad (4.50)
$$

Now $\mu_2$ can be found from (4.38) with $n=2$ by using (4.48-50):

$$
\mu_2 = 2 \left( \frac{1-e^{-\beta \tau}}{e^{-\beta \tau}} \right) \frac{[1-e^{-\beta(1+\beta \tau)}]^2}{\beta^2 e^{-2\beta \tau} [1-e^{-\beta \tau}]} + \frac{[1-e^{-\beta(1+\beta \tau)/(\beta+t)}]}{\beta^2 [1-e^{-\beta \tau}]} \right)
\mu_2 = \frac{2}{\beta^2 e^{-2\beta \tau}} \left[ 1 - e^{-\beta(1+2\beta \tau)} + e^{-2\beta \tau}(\beta^2+(\beta \tau)^2/2) \right] \quad (4.51)
$$

Let $\nu_k c^2$ be the variance of the duration of stage $k$. The main result of this sub-section can now be deduced from the result of corollary 4.19.

**Corollary 4.20**

Suppose that priority by extension is provided for buses arriving at a mean rate of $\beta/c$. Then

$$
P_n = \delta \quad (1 \leq n \leq m)
$$

$$
\epsilon_i = \begin{cases}
\lambda_k + \frac{e^{\beta \tau(1+\beta \tau)}}{\beta} & (i=k) \\
\lambda_i & \text{otherwise}
\end{cases}
$$

$$
\nu_k = \frac{[e^{2\beta \tau}-2\beta \tau e^{\beta \tau}-1]}{\beta^2}
$$

**Proof**

The only effect of priority by extension is to increase the duration of stage $k$ in some cycles. Thus the duration of other stages is always the pre-set maximum value, so $\epsilon_i = \lambda_i \quad (1 \leq i < k, k < i \leq m)$. Since there are $m$
stages in every cycle, \( P_n = \delta_{nm} (1 \leq n \leq m) \) and from the result of corollary 4.4, \( \epsilon_0 = \lambda_0 \). Finally, from the result of corollary 4.19, the duration of stage \( k \) is increased by an amount with first two moments given by (4.47).

Thus

\[
\epsilon_k = \lambda_k + \left[ e^{\beta T} - (1 + \beta T) \right] / \beta
\]

and

\[
u_k = 2 \left[ e^{2\beta T} - e^{\beta T} (1 + 2\beta T) + (\beta T + (\beta T)^2 / 2) \right] / \beta^2 - \left[ e^{2\beta T} - 2e^{\beta T} (1 + \beta T) + (1 + 2\beta T + (\beta T)^2) \right] / \beta^2
\]

\[= \left[ e^{2\beta T} - 2\beta T e^{\beta T} - 1 \right] / \beta^2 \]

The methods used to establish the results of corollaries 4.19 and 4.20 are more powerful than was required for this purpose. They will be applied to other problems in this and the next chapter.

Results equivalent to (4.50) have been given by Adams (1936) and Tanner (1951) to estimate the mean delay incurred by pedestrians in crossing a road which carries traffic with negative-exponential headways. Tanner (1951) also gave a formula for the variance of this delay which is equivalent to that given in (4.52). Allsop (1977b) gave (4.50) and deduced the expression for \( \epsilon_k \) in (4.52) in his analysis of responsive priority for buses.

Several variants on the extension rule are possible. Firstly, this form of priority can be restricted so that no more than one bus may be granted priority by extension during each cycle. This variant of the extension priority rule is amenable to a similar analysis to that of corollaries 4.19 and 4.20 as follows.

**Lemma 4.21**

Suppose that the hypotheses of corollary 4.19 are satisfied with the exception that no more than one bus may be granted priority by extension in each cycle. Then the moments of the time by which the duration of stage \( k \)
exceeds its unextended duration are given by

\[ \mu_n c^n = \frac{n!}{\beta^n} \left[ 1 - e^{-\beta \tau} \sum_{j=0}^{n} \frac{\beta^j}{j!} \right] c^n \quad (n \geq 1) \] (4.53)

Proof

Since stage \( k \) can be extended no more than once, there are only two cases to consider. These are:

(a) with probability \((1-P_x)\), stage \( k \) is not extended;

(b) with probability \(P_x\), stage \( k \) is extended by an amount of time with the truncated negative-exponential probability density function on \([0, \tau c]\).

Each instance of case (a) contributes nothing to the moments of the time by which stage \( k \) exceeds its unextended duration. Accordingly the moments \( \mu_n \) can be calculated from the expressions (4.42) for \( u_n \) and (4.48) for \( P_x \). Thus

\[ \mu_n c^n = P_x u_n c^n \]

\[ = \frac{n!}{\beta^n} \left[ 1 - e^{-\beta \tau} \sum_{j=0}^{n} \frac{\beta^j}{j!} \right] c^n \quad (n \geq 1) \] (4.53)

The form of the first moment

\[ \mu_1 = \left[ 1 - e^{-\beta (1+\tau \xi)} \right] / \beta \] (4.54)

was given by Allsop (1977b) for this form of priority. Inspection of (4.50) shows that if any number of extensions are allowed in each cycle, then as the mean bus arrival rate increases, the mean duration of stage \( k \) increases without bound. Since each extension adds at most \( \tau c \) to the duration of stage \( k \), if at most one extension is granted in each cycle, then the mean duration of stage \( k \) cannot exceed \( (\lambda_k + \tau)c \). The following result improves on this upper bound.
Corollary 4.22

Under the hypotheses of lemma 4.21, the mean time by which the duration of stage $k$ exceeds its unextended duration satisfies the upper bound

$$\mu_1 \leq \tau/3$$  \hspace{1cm} (4.55)

Proof

Inspection of (4.54) shows that $\mu_1$ is non-negative and that as $\beta$ tends to either 0 or $\infty$, $\mu_1$ tends to 0. Differentiating (4.54) with respect to $\beta$ gives

$$\frac{d\mu_1}{d\beta} = e^{-\beta \tau} \frac{[1 + \beta \tau + (\beta \tau)^2] - 1}{\beta^3}$$  \hspace{1cm} (4.56)

which is a continuous function of $\beta$. Applying Rolle's theorem (Dieudonné, 1969, p 153) to $\mu_1(\beta)$, $\beta \in [0, \infty)$, there is a point ($\beta_0$, say) at which $\frac{d\mu_1}{d\beta} |_{\beta_0} = 0$ and at which $\mu_1$ achieves its maximum value. Now

$$\frac{d\mu_1}{d\beta} = 0 \iff [1 + \beta \tau + (\beta \tau)^2] - e^{\beta \tau} = 0$$  \hspace{1cm} (4.57)

$$\iff 1 - 2 \sum_{n=3}^{\infty} \frac{(\beta \tau)^{n-2}}{n!} = 0$$  \hspace{1cm} (4.58)

The left-hand side of (4.58) is a strictly decreasing function of $\beta$, so the root of this equation, $\beta_0$, is unique. Rearranging (4.57) and substituting for $e^{-\beta \tau}$ in (4.54) gives

$$\tilde{\mu}_1(\beta_0) = \frac{\beta_0 \tau}{1 + \beta_0 \tau + (\beta_0 \tau)^2} \tau = f(\beta_0 \tau) \tau$$  \hspace{1cm} (4.59)

where

$$f(x) \equiv \frac{x}{1 + x + x^2}$$  \hspace{1cm} (4.60)
Now $\bar{\mu}_1(\beta_0)$ can only be obtained explicitly from (4.59) if the value of $\beta_0$ is known. However, examination of the derivative of $f(x)$ given by (4.60) shows that $f(x) \leq 1/3 \ (0 \leq x < \infty)$. Using this in (4.59) gives $\bar{\mu}_1(\beta_0) \leq 1/3$, but $\beta_0$ is the value of $\beta$ which maximises $\bar{\mu}_1(\beta)$ so $\mu_1 \leq 1/3 \ (0 \leq \beta < \infty)$.

A second variant on the extension rule involves the use of several detectors placed serially on the carriageway traversed by buses. Each of these detectors calls an extension which is sufficiently long for a bus to reach the next detector or to cross the stop line in normal circumstances. In this case, the trial period corresponds to the time between the arrival of a bus at the outermost detector and the end of the last extension called by that bus. The duration of this period is a random variable which is not related to the length of a single extension in any simple way. If the distribution of this time were known, then the moments of the duration of stage $k$ could be determined from an analysis rather more intricate than that of lemma 4.18 et seq.

4.4.3 Priority by hurry-call.

The hurry-call priority rule enables the controller to reduce delays incurred by buses which arrive at the detector when the stream in which they travel does not have right of way. A particular feature of this priority rule is that it does not alter the order in which the stages of the sequence occur. The action of the hurry-call rule depends on the time at which the bus arrives at the detector as follows.

Arrival before the end of stage $k$: no effect.

Arrival during the transition period immediately preceding stage $i$ ($k < i \leq m$) or during the minimum green time for that stage: the duration of stage $i$ and of any subsequent stages in that cycle is reduced to the minimum permissible values.
Arrival during stage \( i \) (\( k \leq i \leq m \)) after the minimum green time has elapsed: stage \( i \) is terminated immediately and the durations of any subsequent stages in that cycle are reduced to the minimum permissible values.

The mean durations of the stages at a junction where there is priority by hurry-call can be deduced readily from results already presented here. These are given, together with the other statistics required for the analysis of capacity, in the following lemma.

**Lemma 4.23**

Suppose that priority by hurry-call is provided for buses arriving at a mean rate of \( \beta/c \). Then

\[
P_n = \delta_{nm} \quad \text{for} \quad (k \leq n \leq m) \\
E_i = \left\{ \begin{array}{ll}
\lambda_i \\
\gamma_i + \tilde{p}_0 \left[ \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j^0 \gamma_j \right] \left[ 1 - \tilde{p}_0 (\lambda_i + \gamma_i) \right] / \beta
\end{array} \right\} \quad (0 \leq i \leq k)
\]

\[
E_i = \left\{ \begin{array}{ll}
\lambda_i \\
\gamma_i + \tilde{p}_0 \left[ \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j^0 \gamma_j \right] \left[ 1 - \tilde{p}_0 (\lambda_i + \gamma_i) \right] / \beta
\end{array} \right\} \quad (k \leq i \leq m)
\]

**Proof**

Since the hurry-call priority rule does not affect the number of stages in the cycle, \( P_m = 1 \) and \( P_n = 0 \) (\( n \neq m \)) and so from the result of corollary 4.4, \( E_i = \lambda_i \) (\( 1 \leq i \leq k \)). The duration of the stages numbered up to \( k \) are unaffected by the action of the hurry-call rule, so \( E_i = \lambda_i \) (\( 1 \leq i \leq k \)). For each stage \( i \) (\( k \leq i \leq m \)), there are three distinct possibilities.

(a) With probability \( 1 - \tilde{p}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j^0 \gamma_j \right) \) a hurry-call is granted before the minimum green time for stage \( i \) has elapsed: the duration of stage \( i \) is exactly \( \gamma_i \).

(b) With probability \( \tilde{p}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j^0 \gamma_j \right) \left[ 1 - \tilde{p}_0 (\lambda_i + \gamma_i) \right] \) no hurry-call is granted before the minimum green time for stage \( i \) has elapsed but stage \( i \) is terminated later by a hurry-call: the duration of stage
i in excess of the minimum green time is distributed with the truncated negative-exponential probability density function on
\[0, (\lambda_i - \gamma_i) \theta]\.

(c) With probability
\[\tilde{P}_0\left(\sum_{j=k+1}^i \lambda_j + \sum_{j=k}^{i-1} \lambda_j \gamma_j \theta\right)\]
no hurry-call is granted before stage \(i\) ends: the duration of stage \(i\) is exactly \(\lambda_i \theta\).

The conditional expectation of the duration of stage \(i\) in case (b) can be calculated according to the result of lemma 4.18 as
\[\gamma_i \left[1 - \frac{\beta(\lambda_i - \gamma_i) + (1 + \beta(\lambda_i - \gamma_i))}{\beta(1 - \beta(\lambda_i - \gamma_i))}\right] \lambda_i \theta.
\]
Combining each of the three conditional expectations with the appropriate probability and using (4.36) gives

\[E_i = \sum_{j=k+1}^i \lambda_j + \sum_{j=k}^{i-1} \lambda_j \gamma_j \theta \gamma_i \left[1 - \frac{\beta(\lambda_i - \gamma_i) + (1 + \beta(\lambda_i - \gamma_i))}{\beta(1 - \beta(\lambda_i - \gamma_i))}\right] \lambda_i \theta \tilde{P}_0(\lambda_i - \gamma_i) \lambda_i \theta \]

\[= \gamma_i \tilde{P}_0(\sum_{j=k+1}^i \lambda_j + \sum_{j=k}^{i-1} \lambda_j \gamma_j \theta) \left[1 - \frac{\beta(\lambda_i - \gamma_i)}{\beta(1 - \beta(\lambda_i - \gamma_i))}\right] \lambda_i \theta \tilde{P}_0(\lambda_i - \gamma_i) \lambda_i \theta \]  

\(k < i \leq m\) \[

4.4.4 Priority by recall.

The recall priority rule is a direct alternative to the hurry-call rule at junctions where there are at least two stages during which buses do not have right of way: at other junctions, the two rules result in identical responses of the controller to the detection of a bus. The recall rule enables the controller to make a transition to the start of the first stage during which buses have right of way if a bus is detected when it does not have right of way. As a result of this, some stages may not occur and some streams of traffic may not have right of way in some cycles. The action of the recall rule depends on the time at which the bus arrives at the detector as follows.
Arrival before the end of stage \( k \): no effect.

Arrival during the transition period immediately preceding stage \( i \) (\( k < i \leq m \)) or during the minimum green time for that stage: the duration of stage \( i \) is reduced to the minimum permissible value and a transition is made to stage 1.

Arrival during stage \( i \) (\( k < i \leq m \)) after the minimum green time has elapsed: stage \( i \) is terminated immediately and a transition is made to stage 1.

The mean durations of the stages at a junction where there is priority by recall can be deduced in a manner similar to that used in the proof of Lemma 4.23 for the case of priority by hurry-call. These are given, together with the other statistics required for the analysis of capacity in the next lemma.

**Lemma 4.24**

Suppose that priority by recall is provided for buses arriving at mean rate of \( \beta/c \). Then

\[
  p_n = \begin{cases} 
    \tilde{p}_0 \left( \sum_{i=k+1}^{n-1} \lambda_i + \sum_{i=k}^{n-2} \lambda_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_{n-1,0} + \lambda_n \right) \right] & (k < n < m) \\
    \tilde{p}_0 \left( \sum_{i=k+1}^{n-1} \lambda_i + \sum_{i=k}^{n-2} \lambda_i \right) & (n = m) \\
    0 & \text{otherwise}
  \end{cases} \quad (4.62)
\]

\[
  \varepsilon_i = \begin{cases} 
    \lambda_i & (1 < i < k) \\
    \tilde{p}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-2} \lambda_j \right) \gamma_i + \tilde{p}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-2} \lambda_j + \gamma_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_{i-1,0} \right) \right] / \beta & (k < i \leq m)
  \end{cases} \quad (4.63)
\]
Proof

There will be \( n \) stages in a cycle \((k < n < m)\) if no bus is detected between the call of stage \( k+1 \) and the call of stage \( n \) but a bus is detected between the call of stage \( n \) and the call of stage \( n+1 \). These events have probability \( p_{0}^{\sum_{i=k+1}^{n-1} \lambda_{i}} + \sum_{i=k}^{n-2} \lambda_{i} \) and \( 1 - p_{0}^{\lambda_{n-1} + \lambda_{n}} \) respectively and their product gives the formula for \( P_{n}^{(k < n < m)} \). There will be \( m \) stages in a cycle if no bus is detected between the call of stage \( k+1 \) and the call of stage \( m \). Thus the formula for \( P_{m} \) is just the first of the above probabilities with \( n = m \). No other number of stages is possible, so \( P_{n} = 0 \) \((n < k, n > m)\) and (4.62) is established.

Using the expressions (4.62) for \( P_{n} \) in the general formula (4.7) for \( e_{0} \) gives

\[
e_{0} = \sum_{i=1}^{k} \lambda_{i}^{0} + \sum_{i=k+1}^{n-1} \lambda_{i}^{0} + \sum_{i=k}^{n-2} \lambda_{i}^{0} \left[ 1 - p_{0}^{\lambda_{n-1} + \lambda_{n}} \right] \left( \sum_{i=k+1}^{n-1} \lambda_{i}^{0} + \lambda_{n} \right) + \sum_{m=1}^{m-1} \sum_{m=1}^{n-1} \lambda_{m}^{0} + \lambda_{n}^{0} \left( \sum_{i=k+1}^{n} \lambda_{i}^{0} + \lambda_{n}^{0} \right) \left[ 1 - p_{0}^{\lambda_{n-1} + \lambda_{n}} \right] \left( \sum_{i=k+1}^{n} \lambda_{i}^{0} + \lambda_{n}^{0} \right) \quad \text{(4.62)}
\]

The duration of each stage \( i \) \((1 \leq i \leq k)\) is unaffected by the recall rule, so \( e_{i} = \lambda_{i}^{0} \) \((1 \leq i \leq k)\). For each stage \( i \) \((k < i \leq m)\), there are four distinct possibilities.

(a) With probability \( 1 - p_{0}^{\lambda_{i-1}^{0} + \lambda_{i-2}^{0}} \) a recall is granted before stage \( i \) is called: the duration of stage \( i \) is 0.

(b) With probability \( p_{0}^{\lambda_{i-1}^{0} + \lambda_{i-2}^{0}} \left[ 1 - p_{0}^{\lambda_{i-1}^{0} + \gamma_{i}} \right] \) a recall is granted to a bus which is detected during the transition period immediately preceding stage \( i \) or during the minimum green time for that stage: the duration of stage \( i \) is exactly \( \gamma_{i} \).

(c) With probability \( p_{0}^{\lambda_{i-1}^{0} + \lambda_{i-2}^{0}} \left[ 1 - p_{0}^{\lambda_{i-1}^{0} + \gamma_{i}} \right] \) no recall is granted before the minimum green time for stage \( i \) has elapsed but
stage \( i \) is terminated later by a recall: the duration of stage \( i \) in excess of the minimum green time is distributed with the truncated negative-exponential probability density function on \([0, (\lambda_i - \gamma_i) c]\).

(d) With probability \( \tilde{P}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{j=0} \lambda_j \right) \) no recall is granted before stage \( i \) ends: the duration of stage \( i \) is then exactly \( \lambda_i c \).

The conditional expectation of the duration of stage \( i \) in case (c) can again be calculated according to the result of lemma 4.18 as

\[
\tilde{Y}_i + \frac{[1 - e^{-\beta(\lambda_i - \gamma_i)}][1 + \beta(\lambda_i - \gamma_i)]]}{\beta[1 - e^{-\beta(\lambda_i - \gamma_i)}]}.
\]

Combining each of the four conditional expectations with the appropriate probability and using (4.36) gives

\[
\tilde{e}_i = \tilde{P}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{j=0} \lambda_j \right) \left[ \tilde{Y}_i + \tilde{P}_0 \left( \sum_{j=k+1}^{j=0} \lambda_j + \sum_{j=k}^{j=0} \lambda_j + \gamma_i \right) \right]
\]

\[
\left[ \tilde{Y}_i + \tilde{P}_0 \left( \lambda_i - \gamma_i \right) \left( 1 + \beta(\lambda_i - \gamma_i) \right) \right] + \tilde{P}_0 \left( \lambda_i - \gamma_i \right) \lambda_i \right]
\]

\[
\tilde{P}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{j=0} \lambda_j \right) \left[ \tilde{Y}_i + \tilde{P}_0 \left( \sum_{j=k+1}^{j=0} \lambda_j + \sum_{j=k}^{j=0} \lambda_j + \gamma_i \right) \right]
\]

\[
\left[ \tilde{Y}_i + \tilde{P}_0 \left( \lambda_i - \gamma_i \right) \left( 1 + \beta(\lambda_i - \gamma_i) \right) \right] + \tilde{P}_0 \left( \lambda_i - \gamma_i \right) \lambda_i \right]
\]

\[
(k < i \leq m)
\]

The result of lemma 4.24 provides sufficient information to apply the analysis of capacity presented in Section 4.2 and for the analysis in Section 4.5 to proceed. However, various other statistics will be required in Chapter 5. These are given in lemma 4.25 in terms of the following additional notation. Let

\( \eta_{1c} \) be the conditional mean duration of stage \( i \) given that there are exactly \( i \) stages in a cycle \((k < i \leq m)\)

\( \eta_{2ic}^2 \) be the conditional second moment of the duration of stage \( i \) given that there are exactly \( i \) stages in a cycle \((k < i \leq m)\)

\( P_n^a \) be the probability that stage \( n \) is interrupted by a priority recall \((k < m \leq m)\)

\( \eta_{1c}^a \) be the conditional mean duration of stage \( i \) given that it is interrupted by a priority recall \((k < i \leq m)\)

\( \eta_{2ic}^a \) be the conditional second moment of the duration of stage \( i \) given that it is interrupted by a priority recall \((k < i \leq m)\).
Lemma 4.25

Suppose that priority by recall is provided for buses arriving at a mean rate of $\beta/c$. Then

$$
\eta_1^a = \left\{ \begin{array}{l}
(1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1))^{-1}\{y_1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)\lambda_1 + \\
+ \tilde{P}_0(\lambda_{i-1},0+\gamma_1)[1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)/\beta] \} \\
+ \tilde{P}_0(\lambda_{i-1},0+\gamma_1)[1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)/\beta] \} \\
\end{array} \right. \\
(k<i\leq m) 
$$

$$
\eta_{2i}^a = \left\{ \begin{array}{l}
(1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1))^{-1}\{y_1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)\lambda_1 + \\
+ 2\tilde{P}_0(\lambda_{i-1},0+\gamma_1)[(y_1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)\lambda_1 + (1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)/\beta)] \} \\
+ 2\tilde{P}_0(\lambda_{i-1},0+\gamma_1)[(y_1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)\lambda_1 + (1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)/\beta)] \} \\
\end{array} \right. \\
(k<i\leq m) 
$$

$$
\eta_1 = \left\{ \begin{array}{l}
\eta_1^a \quad (k<i\leq m) \\
(1-\tilde{P}_0(\lambda_{m-1},0+\gamma_m)(\lambda_{m-1},0+\gamma_m)/\beta) \quad (i=m) \\
\end{array} \right. 
$$

$$
\eta_{2i} = \left\{ \begin{array}{l}
\eta_{2i}^a \quad (k<i\leq m) \\
(1-\tilde{P}_0(\lambda_{m-1},0+\gamma_m)(\lambda_{m-1},0+\gamma_m)/\beta) \quad (i=m) \\
\end{array} \right. 
$$

$$
p_n^a = \tilde{P}_0(\lambda_{n-1},0+\gamma_n)(\lambda_{n-1},0+\gamma_n)/\beta) \quad (k<n\leq m) 
$$

Proof

Stage $i$ $(k<i\leq m)$ will be interrupted by a priority recall if no bus is detected between the call of stage $k+1$ and the call of stage $i$ but a bus is detected between the call of stage $i$ and the end of that stage. In order to calculate the values of the conditional moments $\eta_1^a$ and $\eta_{2i}^a$ $(k<i\leq m)$, two distinct cases are considered.

(a) With conditional probability $1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)$ a recall is granted to a bus which is detected during the transition period immediately preceding stage $i$ or during the minimum green time for that stage: the duration of stage $i$ is exactly $\gamma_1 c$.

(b) With conditional probability $\tilde{P}_0(\lambda_{i-1},0+\gamma_1)[1-\tilde{P}_0(\lambda_{i-1},0+\gamma_1)]$ no recall is granted before the minimum green time for stage $i$ has elapsed but stage $i$ is terminated later by a recall; the duration of stage $i$ in excess of the minimum green time has the truncated negative-exponential distribution on $[0,(\lambda_{i-1},0+\gamma_1)c]$.
The first two conditional moments of stage $i$ in case (b) can be calculated from the result of lemma 4.18 as

$$
\gamma_i + \frac{[1-e^{-\beta(\lambda_i-\gamma_i)}(1+\beta(\lambda_i-\gamma_i))]}{\beta[1-e^{-\beta(\lambda_i-\gamma_i)}]}
$$

and

$$
\gamma_i + \frac{2[1-e^{-\beta(\lambda_i-\gamma_i)}[1+\beta(\lambda_i-\gamma_i)+\beta^2(\lambda_i-\gamma_i)^2/2]]}{\beta[1-e^{-\beta(\lambda_i-\gamma_i)}]}
$$

respectively. Combining the conditional expectations with the appropriate probabilities and using (4.36) gives

$$
\eta_i = \frac{[1-P_0(\lambda_i-1,0+\gamma_i)]}{[1-P_0(\lambda_i-1,0+\lambda_i)]} \gamma_i + \frac{\bar{P}_0(\lambda_i-1,0+\lambda_i)[1-P_0(\lambda_i-\gamma_i)]}{[1-P_0(\lambda_i-1,0+\lambda_i)]} \left[ \gamma_i + \frac{1-P_0(\lambda_i-\gamma_i)}{\beta[1-P_0(\lambda_i-\gamma_i)]} \right]
$$

Similarly

$$
\eta_{2i} = \frac{[1-P_0(\lambda_i-1,0+\gamma_i)]}{[1-P_0(\lambda_i-1,0+\lambda_i)]} \gamma_i + \frac{\bar{P}_0(\lambda_i-1,0+\lambda_i)[1-P_0(\lambda_i-\gamma_i)]}{[1-P_0(\lambda_i-1,0+\lambda_i)]} \left[ \gamma_i + \frac{2[1-P_0(\lambda_i-\gamma_i)[1+\beta(\lambda_i-\gamma_i)+\beta^2(\lambda_i-\gamma_i)^2/2]]}{\beta[1-P_0(\lambda_i-\gamma_i)]} \right]
$$

so (4.64) is established.

The values of $\eta_i$ and $\eta_{2i}$ ($k \leq m$) are identical to those of $\eta^a_i$ and $\eta^a_{2i}$ respectively since there are $i$ stages in a cycle ($k \leq m$) if

and only if stage $i$ is interrupted by a priority recall. The exceptional cases are those of $\eta_m$ and $\eta_{2m}$ since besides the two possibilities just described, there are $m$ stages in every cycle where no recall is granted.

The conditional probability that a recall is granted is $1-P_0(\lambda_{m-1,0+\lambda_m})$ and that no recall is granted is $P_0(\lambda_{m-1,0+\lambda_m})$. In the latter case, the duration of stage $m$ is exactly $\lambda_m$. Using this information together with (4.64) gives

$$
\eta_m = [1-P_0(\lambda_{m-1,0+\lambda_m})] \eta^a_m + P_0(\lambda_{m-1,0+\lambda_m}) \lambda_m
$$

$$
= \gamma_m + \bar{P}_0(\lambda_{m-1,0+\gamma_m})[1-P_0(\lambda_m-\gamma_m)]/\beta
$$

so (4.65) is established. Similarly
\[ \eta_{2m} = [1-\tilde{F}_0(\lambda_{m-1,0}+\lambda_m)]\eta_{2m}^a + \tilde{F}_0(\lambda_{m-1,0}+\lambda_m)\lambda_m^2 \]
\[ = \gamma_m^2 + 2\tilde{F}_0(\lambda_{m-1,0}+\gamma_m)\left[\frac{\gamma_m-\tilde{F}_0(\lambda_m-\gamma_m)}{\beta} + \frac{1-\tilde{F}_0(\lambda_m-\gamma_m)}{\beta^2}\right] \]
so (4.66) is established.

Finally, by the same argument, \( \eta_{n}^a = \eta_{n} \) for \( k < n < m \) and \( \eta_{m}^a \) can be calculated from the same formula as these, which was given in lemma 4.24.

4.4.5 The inhibition rule.

The purpose of inhibition rules is to prevent priority from being granted to many buses in quick succession. These rules are implemented where excessive disruption would otherwise be caused to non-priority traffic. The rule which is analysed here is the one recommended by the Department of Transport (1977) and applies equally to the hurry-call and recall priority rules. The action of the rule can be described as follows.

A bus may be granted priority by hurry-call (recall) only if no bus was granted priority by hurry-call (recall) during the previous cycle.

If priority by extension is implemented, then its availability is not restricted by this rule. Allsop (1977b) analysed a different inhibition rule in conjunction with priority by extension and recall. In that case, both priority methods were suspended until the end of the transition period following the next uncurtailed occurrence of stage \( m \) after priority had been granted by either method. Thus if a bus was granted priority by extension, then no further priority would be available until the end of that cycle. If a bus was granted priority by recall, then no further priority would be available until the end of the next cycle. This rule, and the various other possibilities, can be analysed in a manner similar to that presented here.

Inhibition is said to be in effect from the time at which a bus is granted priority by hurry-call (recall) until the end of the transition
period following the next uncurtailed occurrence of stage \( m \). Thus inhibition will be in effect throughout some cycles, although this has no effect on the durations of stages 1 to \( k \).

The following notation is introduced to describe the consequences of the inhibition rule. Let

\[
P_I \quad \text{be the probability that inhibition is in effect at the start of stage 1}
\]

\[
\varepsilon_{0j}^+ \quad \text{be the conditional mean duration of the transition periods in a cycle given that inhibition is not in effect at the start of the cycle}
\]

\[
\varepsilon_{ij}^+ \quad \text{be the conditional mean duration of stage } i \text{ given that inhibition is not in effect at the start of the cycle } (1 \leq i \leq m)
\]

\[
p_{ni}^+ \quad \text{be the conditional probability that there are exactly } n \text{ stages in a cycle given that inhibition is not in effect at the start of the cycle } (k \leq n \leq m).
\]

An expression for \( P_I \) can be derived irrespective of whether priority is implemented by hurry-call or recall.

Lemma 4.26

Suppose that priority by either hurry-call or recall is provided for buses arriving at a mean rate of \( \beta/c \). Then

\[
P_I = 1 - \left[2 - \bar{p}_0 \left( \sum_{j=k+1}^{m} \lambda_j + \sum_{j=k}^{m-1} \lambda_j \right) \right]^{-1}
\]  \hspace{1cm} (4.68)

Proof

If inhibition is in effect at the start of a cycle, then a bus was granted priority by hurry-call or recall during the previous cycle. The probabilities that priority is available and that a bus is detected during the appropriate period are \( 1-P_I \) and \( 1-\bar{p}_0 \left( \sum_{j=k+1}^{m} \lambda_j + \sum_{j=k}^{m-1} \lambda_j \right) \) respectively.
Thus
\[ P_I = (1-P_i) \left[ 1 - \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} \right) \right] \]  
which can be rearranged to give (4.68).

**Corollary 4.27**

The limiting value of \( P_I \) as the mean bus arrival rate increases without bound is given by
\[ \lim_{\beta \to \infty} P_I = 1/2 \]  
(4.70)

**Proof**

Since
\[ \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} > 0, \lim_{\beta \to \infty} \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} \right) = 0. \]

From (4.68), \( P_I \) varies continuously with \( \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} \right) \) and has no other dependence on \( \beta \), so
\[ \lim_{\beta \to \infty} P_I = 1 - 2^{-1} \]
\[ = 1/2 \]

The main results of this sub-section can now be stated. The first two of these analyse the combination of the hurry-call and inhibition rules while the others analyse the combination of recall and inhibition.

**Lemma 4.28**

Suppose that priority by hurry-call with inhibition is provided for buses arriving at a mean rate of \( \beta/c \). Then
\[ P_n = \frac{\delta}{nm} \]
\[ \epsilon_i = \begin{cases} \lambda_i & (0 \leq i \leq k) \\ \left\{ 1 - \left[ 2 - \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} \right) \right]^{-1} \lambda_j \right\} + \left[ 2 - \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} \right) \right]^{-1} \left( \gamma_i + \tilde{P}_0 \left( \sum_{j=k+1}^{\infty} \lambda_j + \sum_{j=k}^{\infty} \lambda_{j0} + \gamma_i \right) [1 - \tilde{P}_0 (\lambda_i - \gamma_i)] / \beta \right) & (k < i \leq m) \end{cases} \]  
(4.71)
Proof
As in the case of priority by hurry-call, there are \( m \) stages in every cycle, so \( P_n = \delta_{nm} \) and from the result of corollary 4.4, \( \varepsilon_0 = \lambda_0 \). Again, the duration of stage \( i \) \((1 \leq i \leq k)\) is unaffected, so \( \varepsilon_i = \lambda_i \) \((1 \leq i \leq k)\).

If inhibition is in effect at the start of a cycle, then the duration of stage \( i \) \((k < i \leq m)\) is exactly \( \lambda_i c_0 \). Otherwise, the mean duration of stage \( i \) \((k < i \leq m)\) is given by (4.61) as
\[
\varepsilon_i^+ = \gamma_i + \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=i}^{m} \lambda_j \gamma_j \beta \int [1 - \frac{\gamma_i}{\lambda_i - \gamma_i}] \beta \quad (k < i \leq m) \tag{4.72}
\]

Since the probability that inhibition is in effect at the start of a cycle is \( P_I \), the unconditional mean duration of stage \( i \) \((k < i \leq m)\) is given by
\[
\varepsilon_i = P_I \lambda_i + (1 - P_I) \varepsilon_i^+ \quad (k < i \leq m) \tag{4.73}
\]

Substituting the expression (4.68) for \( P_I \) and (4.72) for \( \varepsilon_i^+ \) into (4.73) gives the required expression for \( \varepsilon_i \) \((k < i \leq m)\).

Corollary 4.29
Under the hypotheses of lemma 4.28, the limiting values of \( \varepsilon_i \) \((k < i \leq m)\) as the mean bus arrival rate increases without bound are given by
\[
\lim_{\beta \to \infty} \varepsilon_i = \frac{\lambda_i + \gamma_i}{2} \quad (k < i \leq m) \tag{4.74}
\]

Proof
From (4.73), \( \varepsilon_i \) varies continuously with \( P_I \) and \( \varepsilon_i^+ \). From the result of corollary 4.27, the limiting value of \( P_I \) is 1/2 and since
\[
\sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{m} \lambda_j \gamma_j > 0 \quad (k < i \leq m) \quad \lim_{\beta \to \infty} \frac{\pi}{\beta} \int \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{m} \lambda_j \gamma_j \right) \beta \quad (k < i \leq m) ,
\]
using (4.72) gives \( \lim_{\beta \to \infty} \varepsilon_i^+ = \gamma_i \) \((k < i \leq m)\). Substituting these values into (4.73) and using the continuity of \( \varepsilon_i \) gives (4.74).
Lemma 4.30

Suppose that priority by recall with inhibition is provided for buses arriving at a mean rate of \( \beta/c \). Then

\[
P_n = \begin{cases} 
\left[2 - \tilde{P}_0 \left( j = x+1, \sum_{j=k}^m \lambda_j \right) \right]^{-1} \left[1 - \tilde{P}_0 \left( i = k+1, \sum_{i=k}^m \lambda_i \right) \right] & (k < n < m) \\
1 - 2 \tilde{P}_0 \left( j = x+1, \sum_{j=k}^m \lambda_j \right) & (n = m) \\
0 & \text{otherwise}
\end{cases}
\]

\[
\varepsilon_i = \begin{cases} 
\lambda_i & (i = 0) \\
\left\{1 - 2 \tilde{P}_0 \left( j = k+1, \sum_{j=k}^m \lambda_j \right) \right\}^{-1} \left[2 - \tilde{P}_0 \left( j = x+1, \sum_{j=k}^m \lambda_j \right) \right]^{-1} & (1 < i < k) \\
\left\{1 - 2 \tilde{P}_0 \left( j = k+1, \sum_{j=k}^m \lambda_j \right) \right\}^{-1} \left[2 - \tilde{P}_0 \left( j = x+1, \sum_{j=k}^m \lambda_j \right) \right]^{-1} & (i = k) \\
\tilde{P}_0 \left( j = k+1, \sum_{j=k}^m \lambda_j + \sum_{j=k}^m \lambda_j + \gamma_i \right) \left[1 - \tilde{P}_0 \left( \lambda_i + \gamma_i \right) / \beta \right] & (k < i < m)
\end{cases}
\]

Proof

If inhibition is in effect at the start of a cycle, then there will be exactly \( m \) stages in that cycle, the total duration of the transition periods will be \( \lambda_0 c \) and the duration of stage \( i \) (\( 1 < i < m \)) will be exactly \( \lambda_i c \). Otherwise, the probability that there are exactly \( n \) stages in a cycle is given by (4.62) as

\[
P_n = \begin{cases} 
\tilde{P}_0 \left( i = x+1, \sum_{i=k}^m \lambda_i \right) [1 - \tilde{P}_0 \left( \lambda_{n-1} + \lambda_n \right)] & (k < n < m) \\
\tilde{P}_0 \left( i = k+1, \sum_{i=k}^m \lambda_i + \sum_{i=k+1}^m \lambda_i \right) & (n = m) \\
0 & \text{otherwise}
\end{cases}
\]

Similarly, the conditional mean durations \( \varepsilon_i^+ \) (\( 0 < i < m \)) are given by (4.63) as

\[
\varepsilon_i^+ = \begin{cases} 
\lambda_i & (i = k) \\
\tilde{P}_0 \left( j = k+1, \sum_{j=k}^m \lambda_j + \sum_{j=k}^m \lambda_j + \gamma_i \right) \left[1 - \tilde{P}_0 \left( \lambda_i + \gamma_i \right) / \beta \right] & (k < i < m)
\end{cases}
\]

Since the probability that inhibition is in effect at the start of a cycle is \( P_i \), the unconditional probabilities \( P_n \) are given by

\[
P_n = P_i \delta_{nm} + (1 - P_i) P_n^+ \quad (1 < n < m)
\]
Substituting the expressions (4.68) for $P_i$ and (4.77) for $P^+_n$ into (4.79) gives (4.75). Similarly, the unconditional expectations $E_i$ ($0 \leq i \leq m$) are given by

$$E_i = P_i \lambda_i + (1 - P_i)E^+_i \quad (0 \leq i \leq m)$$

(4.80)

Substituting the expressions (4.68) and (4.78) into (4.80) gives (4.76). []

**Corollary 4.31**

Under the hypotheses of lemma 4.30, the limiting values of $P_n$ ($1 \leq n \leq m$) and $E_i$ ($0 \leq i \leq m$) as the mean bus arrival rate increases without bound are given by

$$\lim_{\beta \to \infty} P_n = (\delta_{nm} + \delta_{n,k+1})/2$$

(4.81)

and

$$\lim_{\beta \to \infty} E_i = \begin{cases} 
\lambda_i & (i = 0) \\
(\lambda_0 + \sum_{j=1}^{i} \lambda_j + \lambda_{k+1,i})/2 & (1 \leq i \leq k) \\
\lambda_{k+1} + \gamma_k & (i = k+1) \\
\lambda_i/2 & (k+1 \leq i \leq m)
\end{cases}$$

(4.82)

**Proof**

Since

$$\begin{align*}
\sum_{i=k+1}^{n-1} \lambda_i + \sum_{i=k}^{n-2} \lambda_i &= 0 \\
\sum_{i=k+1}^{n} \lambda_i + \sum_{i=k}^{n-1} \lambda_i &= 0
\end{align*}$$

(4.83)

taking the limit of (4.77) as the mean bus arrival rate increases without bound gives

$$\lim_{\beta \to \infty} P^+_n = \delta_{n,k+1}$$

(4.84)

Taking the limit of (4.79) as the mean bus arrival rate increases without bound and using (4.70) and (4.84) gives (4.81).

From (4.78) and (4.83)

$$\begin{cases} 
\sum_{j=1}^{i} \lambda_j + \lambda_{k+1,i} & (i = 0) \\
\lambda_i & (1 \leq i \leq k) \\
\gamma_k & (i = k+1) \\
0 & (k+1 \leq i \leq m)
\end{cases}$$

(4.85)

Taking the limit of (4.80) and using (4.70) and (4.85) gives (4.82). []
4.5 Signal-settings at a junction where there is priority for buses

4.5.1 Introduction.

The analysis presented in Section 4.4 showed how some of the likely effects of a priority system on capacity can be investigated before it is implemented. In Chapter 5, the same line of investigation is pursued to give estimates of the mean delay incurred by priority and non-priority vehicles and of other related performance measures. In this section, a problem of a different nature is considered: this is the design problem of finding signal-settings suitable for use at a junction where responsive priority is provided for buses.

The criterion adopted here for signal-settings to be suitable is that when they are implemented with the appropriate form of priority, the capacity of each stream of traffic is similar to that obtained when some given signal-settings are implemented in the absence of priority for buses. The problem to be investigated can be stated conveniently in terms of capacity-equivalence and is described fully in sub-section 4.5.2 together with formulations of related problems. The remainder of this section is then devoted to an application of this method to junctions where priority is implemented by each of various specific rules.

Throughout this section, signal-settings are supposed to be available which are suitable for use at a junction in the absence of responsive priority. In particular, these signal-settings are supposed to satisfy the minimum stage duration constraints (4.25) and the cycle-time constraint (4.26). There are no other restrictions on the origins of these signal-settings, so they could be calculated by any normal procedure such as those described in Section 4.3, or might simply be known to work well in practice. In the former case, if the given signal-settings maximise
capacity at the junction in the absence of priority, then the ones calculated by the method developed here will be similar to those which maximise capacity at the junction when implemented with priority for buses arriving at the appropriate mean rate. This property, which is discussed further in sub-section 4.5.2, is a consequence of the capacity-equivalence enjoyed by the two sets of signal-settings. The signal-settings calculated in this way will certainly maximise capacity if the priority method used is one for which capacity-equivalent signal-settings give rise to capacities for each stream which are identical to those arising from the original-signal-settings in the absence of priority.

4.5.2 A mathematical statement of the problem.

The problem considered in this section is that of finding signal-settings which, when implemented with the appropriate form of priority, are capacity-equivalent to some given ones when these are implemented in the absence of priority. This is the inverse to the problem considered in Section 4.4 of finding signal-settings which, when implemented without priority, are capacity-equivalent to some given ones when these are implemented with priority.

This problem arises during the design of a priority installation when the physical layout and stage sequence at the junction have been established and normal procedures have been employed to find signal-settings which are suitable for use in the absence of priority. Each of the procedures used this far in the design process would be required in the design of an ordinary signal-controlled junction: the next step is the topic investigated in this section. The methods developed here could be applied equally well to the signal-settings already in use at an existing junction which is to be modified to include priority for buses.
The following additional notation will be used in the formal statement of the problem to be investigated.  This emphasises the dependence of $\varepsilon_i$ ($0 \leq i \leq m$) on $\beta$ and $\lambda_j$ ($0 \leq j \leq m$).  Let

$$\tilde{\varepsilon}_i((\lambda_j),\beta)$$

be the functions which map the implemented signal-settings $\{\lambda_j\}$ and the mean bus arrival rate $\beta/c^*$ to the values $\varepsilon_i$ ($0 \leq i \leq m$).

$\lambda^*_i$ correspond to signal settings which are known to be suitable for use at the junction under consideration in the absence of priority for buses ($0 \leq i \leq m$).

A convenient formulation of the problem is given in the following lemma.

**Lemma 4.32**

The signal-settings $\{\lambda_j\}$, when implemented with responsive priority for buses arriving at a mean rate of $\beta/c^*$, are capacity-equivalent to the signal-settings $\{\lambda^*_j\}$ if and only if they satisfy simultaneously the $m$ equations

$$\frac{\tilde{\varepsilon}_i((\lambda_j),\beta)}{\tilde{\varepsilon}_0((\lambda_j),\beta)} - \frac{\lambda^*_i}{\lambda^*_0} = 0 \quad (1 \leq i \leq m)$$

(4.86)

Thus finding signal-settings which are suitable for use with priority for buses is equivalent to solving Equations (4.86) for $\{\lambda_j\}$.

**Proof**

Suppose that the signal-settings $\{\lambda_j\}$, when implemented with priority for buses, are capacity-equivalent to $\{\lambda^*_j\}$.  From the result of lemma 4.9, the relationships (4.12) are satisfied by $\tilde{\varepsilon}_i((\lambda_j),\beta)$ ($0 \leq i \leq m$) and $\lambda^*_i$ ($0 \leq i \leq m$).  In particular,

$$\lambda^*_i = \frac{\tilde{\varepsilon}_i((\lambda^*_j),\beta)}{\tilde{\varepsilon}_0((\lambda^*_j),\beta)} \quad (1 \leq i \leq m)$$

(4.87)

which can be rearranged to give (4.86).
Now suppose that \( \bar{\varepsilon}_i(\{\lambda_j\}, \beta) \) (0 \( \leq \) i \( \leq \) m) and \( \lambda^*_i \) (0 \( \leq \) i \( \leq \) m) satisfy (4.86): rearranging these equations gives (4.87). Summing (4.87) over \( i \) (0 \( \leq \) i \( \leq \) m) and using \( \sum_{i=0}^m \lambda^*_i = 1 \) gives

\[
1 = \sum_{i=0}^m \bar{\varepsilon}_i(\{\lambda_j\}, \beta) \overline{\lambda}_0(\{\lambda_j\}, \beta)
\]

(4.88)

\[
\Leftrightarrow \lambda^*_0 = \frac{\bar{\varepsilon}_0(\{\lambda_j\}, \beta)}{\sum_{i=0}^m \bar{\varepsilon}_i(\{\lambda_j\}, \beta)}
\]

(4.89)

Together, (4.87) and (4.89) are equivalent to (4.12), so \( \{\lambda_j\} \) when implemented with priority, are capacity-equivalent to \( \{\lambda^*_j\} \).

Some general results concerning solution of equations of the form of (4.86) are given in Appendix 1. Since solving (4.86) for \( \{\lambda^*_j\} \) provides signal-settings which are capacity-equivalent to \( \{\lambda^*_j\} \), this corresponds to approximate correction of the capacities of the streams for any changes due to the implementation of priority for buses. This problem was considered first by Allsop (1977b), who gave solutions for (4.86) as power series in \( \beta \) under a variety of different priority rules, but with the restriction that each of the transition periods has the same duration.

Vincent, Cooper and Wood (1978) described an iterative method to approximate the solutions of (4.86) when there is only one stage in which buses do not have right of way. A table was given of corrections calculated in this manner.

There are two possible sources of error in the correction procedure investigated here. The first occurs in cases where capacity-equivalent signal-settings do not give rise to identical capacities for all streams: the sizes of any such errors can be estimated from the result of lemma
4.10. The second occurs when the actual bus arrivals do not constitute a Poisson process with mean rate \( \beta/c^* \). If, as might be expected to happen in practice, the mean bus arrival rate differs from \( \beta/c^* \), then the methods of analysis presented in Section 4.4 can be applied to estimate the consequences of this. If the bus arrival process is not Poisson, then further analysis is required. However, Allsop (1977b) showed that the more regular quasi-regular bus arrival process gives rise to capacities which are the same to first order in the mean bus arrival rate as those arising from a Poisson process with the same mean rate.

If the original-signal settings are chosen to maximise capacity, then the capacity-equivalent ones will be similar to those which maximise capacity at the junction when there is priority for buses. The following results provide bounds for any reduction in \( \mu^* \), the maximum flow multiplier for which the signal-settings provide adequate capacity, which results from the first kind of error.

**Lemma 4.33**

Suppose that \( \{\lambda_i^*\} \) are calculated according to the methods described in sub-section 4.3.2 to maximise capacity at the junction and that the maximum flow multiplier for which each of the practical capacity constraints (4.23) is satisfied is \( \mu^* \). If \( \{\lambda_i\} \), when implemented with priority for buses, are capacity-equivalent to \( \{\lambda_i^*\} \), then the maximum flow multiplier for which each of the practical capacity constraints is satisfied when there is priority for buses and the signal-settings \( \{\lambda_i\} \) are implemented is at least

\[
\mu^* - \text{Maximum } \Delta K_j \frac{P_j}{\kappa_j d_j}
\]

where

\[
\Delta K_j = \left[ \frac{\epsilon_0 - \lambda^* - \lambda^*}{\epsilon_0} \right]^{i=1}_{i=n} 10^j 10^{k+1} n^{m} \left[ \frac{\sum_{i=k+1}^{m} a_{i0j} \lambda_i 10^{-}}{\epsilon_0} \right]^{n-1}_{i=k+1} \sum_{i=1}^{n} a_{i0j} \lambda_i 10^{-} \right]^{s_j} \]

(4.90)
Proof

All that is required is to verify that the flows \[ \mu^* - \max_{1 \leq i \leq M} (\Delta K_i \frac{P_i}{K_i q_j}) q_i \] (\(1 \leq i \leq M\)) satisfy the practical capacity constraints when the signal-settings \(\{\lambda_i\}\) are implemented with priority for buses. Since the signal-settings \(\{\lambda_i^*\}\) provide adequate practical capacity for the flows \(\mu^* q_i\) (\(1 \leq i \leq M\)) when implemented without priority,

\[ P_i K_i - \mu^* K_i q_i \geq 0 \quad (1 \leq i \leq M) \]

\(\iff\)

\[ P_i (K_i + \Delta K_i) - \mu^* K_i q_i \geq 0 \quad (1 \leq i \leq M) \] (from corollary 4.11)

\(\iff\)

\[ P_i K_i - [\mu^* - \Delta K_i \frac{P_i}{K_i q_j}] K_i q_i \geq 0 \quad (1 \leq i \leq M) \]

\[ \implies P_i K_i - [\mu^* - \max_{1 \leq i \leq M} (\Delta K_i \frac{P_i}{K_i q_j})] K_i q_i \geq 0 \quad (1 \leq i \leq M) \] (4.91)

Corollary 4.34

Under the hypotheses of lemma 4.33, if there are \(m\) stages in every cycle, then the signal-settings \(\{\lambda_i\}\) maximise capacity at the junction when there is priority for buses. The maximum flow multiplier for which all the practical capacity constraints are satisfied is the same in each case.

Proof

Let \(\mu^*\) be the maximum flow multiplier for which all the practical capacity constraints are satisfied when the signal-settings \(\{\lambda_i^*\}\) are implemented without priority. Similarly, let \(\mu'\) be the maximum flow multiplier associated with the signal-settings \(\{\lambda_i\}\) when these are implemented with priority. As there are \(m\) stages in each cycle, corollary 4.13 can be applied to show that \(\Delta K_i = 0\) (\(1 \leq i \leq M\)). Using this in (4.91) gives

\[ P_i K_i - \mu^* K_i q_i \geq 0 \quad (1 \leq i \leq M) \] (4.92)

\(\iff\)

\[ \mu' \geq \mu^* \] (4.93)
Now let \( \{ \lambda_i^i \} \) be signal-settings which, when implemented with priority, maximise capacity at the junction. Let \( \{ \lambda_i^+ \} \) be signal-settings which, when implemented in the absence of priority, are capacity-equivalent to \( \{ \lambda_i^i \} \). The values of \( \{ \lambda_i^+ \} \) can be calculated according to the methods described in Section 4.4. Arguing as before, since there are \( m \) stages in each cycle,

\[
p_{i_k}^+ - \mu_i^k q_i \geq 0 \quad (1 \leq i \leq M)
\]  

(4.94)

Since \( \mu^* \) is the maximum flow multiplier that can be accommodated by any fixed-time signal-settings, this gives \( \mu^i \leq \mu^* \). Combining this with (4.93) gives \( \mu^i = \mu^* \). From (4.92), the signal-settings \( \{ \lambda_i^i \} \) can accommodate flows with the multiplier \( \mu^i \), so they maximise capacity at the junction when there is priority.

The formulation of the signal-setting policy as one of finding roots of the simultaneous equations (4.86) has the advantage that normal analyses of capacity can be applied to the capacity-equivalent signal-settings \( \{ \lambda_j^* \} \). This enables information concerning capacity at the junction when there is priority for buses to be obtained readily.

A considerable literature exists on the problem of solving equations of the form of (4.86) (see, for example, Ortega and Rheinboldt, 1970).

Furthermore, the results of lemma 4.33 and corollary 4.34 show that if appropriate signal-settings can be found, then responsive priority for buses can be implemented without major reduction in capacity.

Against the advantages of the approach adopted here, some difficulties arise as a consequence of its indirect nature. In particular, in cases where there is no solution to Equations (4.86), the method yields no signal-settings at all. Furthermore, discrepancies between the capacity of each stream in each case do arise in many circumstances. Attention is paid in the remainder of this section to conditions under which solutions to (4.86) exist.
An alternative formulation of the capacity-maximisation problem is now posed and discussed. The capacity of each stream can be calculated from the mean durations of the stages according to the result of theorem 4.7. Thus an explicit problem analogous to that discussed in sub-section 4.3.2 can be stated as follows.

\[
\begin{align*}
\text{Maximise} & \quad \mu \\
\text{subject to} & \quad \mu, \lambda_i^\ell \quad (0 \leq \ell \leq m) \\
& \quad \sum_{i=0}^{\ell} a_{ij} \bar{e}_i^\ell({\lambda_i^\ell}, \beta) - \mu k_j q_j \geq 0 \quad (1 \leq j \leq M) \\
& \quad \lambda_i - \frac{G_i}{L} \lambda_0 \geq 0 \quad (1 \leq i \leq m) \\
& \quad \lambda_0 - \frac{L}{G_0} \geq (\geq) 0
\end{align*}
\]

(4.95)

While this problem retains the linear objective function of (4.27), the practical capacity constraints here are non-linear. Provided that the region defined by the constraints in (4.95) is convex, the problem is well posed. A particular advantage of this formulation of the signal-setting problem is that it will yield a result in all cases where the minimum green constraints are consistent with the cycle-time constraint. However, there is no immediate indication of whether the capacity of the junction has been reduced by the introduction of priority. Furthermore, this formulation is restricted to capacity maximisation by the choice of the objective function in (4.95). While other problems could be stated with non-linear objective functions, they would be considerably more difficult to solve than either of (4.86) or (4.95).
4.5.3 Priority by extension.

This form of priority, which was introduced in sub-section 4.4.2, alters only the duration of stage $k$. Expressions for the mean durations of the stages when this form of priority is implemented were given in the result of corollary 4.20. These are used in the next result to give explicit expressions for signal-settings which are suitable for use at a junction where priority by extension is implemented.

Lemma 4.35

Suppose that priority by extension is provided for buses arriving at a mean rate of $\beta/\alpha^*$. Then the signal-settings given by

$$
\lambda_i = \begin{cases} 
\lambda_k^* - \frac{[e^{\beta T} - (1+\beta T)]}{\beta} & \text{(i=k)} \\
\lambda_i^* & \text{(i\neq k)}
\end{cases} \quad (4.96)
$$

are capacity-equivalent to the signal-settings $\{\lambda_i^*\}$ when these are implemented in the absence of priority for buses.

Proof

Substituting the expressions (4.52) for $\varepsilon_i$ ($0 \leq i \leq m$) into (4.86) and rearranging gives (4.96).

The especially simple form of the correction to the signal-settings $\{\lambda_i^*\}$ given by (4.96) can be used to good advantage when priority by extension is implemented with other priority rules. Since this form of priority acts independently of each of the other rules considered here, the correction to the duration of stage $k$ can be made after corrections for the effects of the other priority methods.

The size of the correction to the duration of stage $k$ increases with the mean bus arrival rate. The following result provides an upper bound on
the mean bus arrival rates for which the signal-settings calculated from (4.96) are acceptable in the sense that they allow adequate time for stage $k$ even in cycles where no bus is granted an extension.

**Corollary 4.36**

The greatest mean bus arrival rate for which the signal-settings $\{\lambda_1\}$ calculated according to the result of lemma 4.35 are acceptable is given by $\beta_c$, the unique value of $\beta$ which satisfies the equation

$$\lambda_k^* - \frac{[e^{\beta T} - (1+\beta T)]}{\beta} - \gamma_k = 0$$  \hspace{1cm} (4.97)

**Proof**

In order for the signal-settings $\{\lambda_1\}$ to be acceptable, they must satisfy the minimum green constraints (4.24). Since $\lambda_1 = \lambda_1^*$ (i≠k), the only minimum green constraint which could be violated as a result of applying the correction (4.96) is the one for stage $k$. Thus the constraint under consideration is

$$\lambda_k - \gamma_k \geq 0$$

$$\iff \lambda_k^* - \frac{[e^{\beta T} - (1+\beta T)]}{\beta} - \gamma_k \geq 0 \text{ (from (4.96))}$$

$$\iff \lambda_k^* - \frac{\sum_{n=2}^{\infty} \frac{\beta^{n-1} T^n}{n!}}{\beta} - \gamma_k \geq 0$$  \hspace{1cm} (4.98)

The left-hand side of (4.98) decreases monotonically and without bound as $\beta$ increases, so the constraint will be satisfied as the equality (4.97) by a unique value $\beta_c$ of $\beta$. Furthermore, the constraint will be satisfied as a strict inequality for all $\beta$ less than $\beta_c$.  

The final result of this sub-section investigates signal-settings which are suitable for use when no more than one bus may be granted priority by extension in each cycle.
Lemma 4.37

Suppose that priority by extension is provided for buses arriving at a mean rate of $\beta/c^*$, and that no more than one extension may be granted in each cycle. Then the signal-settings given by

$$
\lambda_i = \begin{cases} 
\lambda_k^* - \frac{[1 - e^{-\beta T(1+\beta T)}]}{\beta} & (i=k) \\
\lambda_i^* & (i\neq k)
\end{cases}
$$

(4.99)

are capacity-equivalent to $\{\lambda_i^*\}$ when these are implemented without priority. If

$$
\lambda_k^* - \frac{\tau}{3} - \gamma_k \geq 0
$$

(4.100)

then the signal-settings given by (4.99) will be acceptable for all values of $\beta$. In any case, they will be acceptable for $\beta \leq \beta_c$ where $\beta_c$ is the value of $\beta$ which satisfies (4.97).

Proof

The expressions (4.99) are established in the same way as were (4.96) except that the correction to the duration of stage $k$ is here given by (4.54) rather than (4.50). From the result of corollary 4.22, $\lambda_k \geq \lambda_k^* - \frac{\tau}{3}$ for all $\beta \geq 0$. Thus if (4.100) is satisfied, then the minimum green time constraint for stage $k$ will be satisfied by the solutions of (4.99) for any $\beta \geq 0$. Finally, from the result of corollary 4.36,

$$
\beta \leq \beta_c
$$

$$
\Leftrightarrow \lambda_k^* - \frac{[e^{\beta T} - (1+\beta T)]}{\beta} - \gamma_k \geq 0
$$

$$
\Rightarrow \lambda_k^* - \frac{[1 - e^{-\beta T(1+\beta T)}]}{\beta} - \gamma_k \geq 0 \quad \text{(since } e^{\beta T} \geq 1 \text{ for } \beta \geq 0, \tau \geq 0) \quad (4.101)
$$

so the minimum green constraint for stage $k$ is satisfied by $\lambda_k$ given by (4.99).
4.5.4 Priority by hurry-call.

In this sub-section, the problem of finding signal-settings which are suitable for use with priority by hurry-call is considered. The approach used is to solve Equations (4.86) for \( \{\lambda_j \} \) when the particular forms (4.61) are used for \( \{\varepsilon_j \} \). Because there are \( m \) stages in each cycle with this priority rule, the result of corollary 4.13 can be applied to show that these capacity-equivalent signal-settings give rise to capacities for each stream which are identical to those arising from \( \{\lambda^* \} \) in the absence of priority. Explicit solutions are given for Equations (4.86) in this case and some further results are deduced concerning the range of mean bus arrival rates for which satisfactory solutions exist.

The basic result of this sub-section is that of

**Theorem 4.38**

Suppose that priority by hurry-call is provided for buses arriving at a mean rate of \( \beta/c^* \). Then the signal-settings \( \{\lambda_i \} \) which are capacity-equivalent to \( \{\lambda^*_i \} \) satisfy the equations

\[
\lambda_i = \left\{ \begin{array}{ll}
\lambda^*_i & (1 \leq i \leq k) \\
\frac{1}{1 - \frac{\beta - \sum_{j=k+1}^{\infty} \lambda_j}{\sum_{j=k}^{\infty} \lambda_j}} & (k < i \leq m)
\end{array} \right.
\]

(4.102)

If solutions to these equations exist, then they are unique and can be calculated directly from the given forms.

**Proof**

According to the result of lemma 4.32, the required signal-settings \( \{\lambda_i \} \) satisfy Equations (4.86). Substituting the specific forms (4.61) for \( \{\varepsilon_j \} \) when there is priority by hurry-call into (4.86) and multiplying through by \( \lambda^*_0 \) gives
\[ \lambda_i - \lambda_i^* = 0 \quad (1 \leq i \leq k) \]
\[ \gamma_i + \tilde{\beta}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j \right) [1 - \tilde{\beta}_0 (\lambda_i - \gamma_i)] / \beta - \lambda_i^* = 0 \quad (k < i \leq m) \] (4.103)

The first \( k \) of Equations (4.102) follow immediately. The last \( (m-k) \) of Equations (4.103) are equivalent to

\[ \tilde{\beta}_0 (\lambda_i) \tilde{\beta}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j \right) = \tilde{\beta}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j \right) - \beta (\lambda_i^* - \gamma_i) \quad (k < i \leq m) \] (4.104)

Using (4.36) to give explicit forms for the probabilities on the left-hand side of (4.104), taking logarithms and rearranging gives the last \( (m-k) \) of Equations (4.102).

Inspection of expressions (4.102) shows that each \( \lambda_i \) \((1 \leq i \leq k+1)\) is given explicitly in terms of the parameters. Furthermore, each \( \lambda_i \) \((k+1 < i \leq m)\) is given in terms of \( \lambda_j \) \((k < j < i)\). Thus provided that the required logarithms exist, the values \( \{\lambda_i\} \) can be calculated recursively in order of increasing index. The uniqueness of \( \{\lambda_i\} \) is immediate because all the calculations are explicit.

The next results investigate the derivatives of \( \lambda_i \) \((k < i \leq m)\) with respect to the mean bus arrival rate in the region where the expressions (4.102) can be evaluated. These results will then be used to investigate the range of mean bus rates for which suitable signal-settings can be calculated.

**Lemma 4.39**

Under the hypotheses of theorem 4.38, the derivatives with respect to the mean bus arrival rate of the signal-settings \( \lambda_i \) \((k < i \leq m)\) calculated from (4.102) are given by

\[ \frac{d\lambda_i}{d\beta} = \frac{3 \lambda_i}{\beta^2} + \sum_{j=k+1}^{i-1} \frac{3 \lambda_j}{\beta^2} \frac{d\lambda_j}{d\beta} \quad (k < i \leq m) \] (4.105)
where

\[
\frac{3\lambda_i}{3\beta} = \frac{1}{\beta^2} \log_e \left( \tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i) \right) + \frac{(\tilde{p}_0 (\Psi_i + \gamma_i) \frac{3}{3\lambda_i} (\lambda_i^* - \gamma_i))}{\beta \left[ \tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i) \right]} \tag{4.106}
\]

\[
\frac{3\lambda_i}{3\lambda_j} = \frac{\beta (\lambda_i^* - \gamma_i)}{\tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \quad (k < j < i), (k < i \leq m) \tag{4.107}
\]

and \( \Psi_i = \sum_{j=1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j \) \quad (k < i \leq m)

Proof

Since the expressions (4.102) are used for \( \lambda_i \) \quad (k < i \leq m), \quad \frac{3\lambda_i}{3\beta} = 0 \quad (k < i \leq m), (i < j \leq m). \) Thus the total derivative \( \frac{3\lambda_i}{3\beta} \quad (k < i \leq m) \) depends on the partial derivative \( \frac{3\lambda_i}{3\beta} \) only when \( k < j < i \) \quad so (4.105) is established. The expressions (4.106) and (4.107) can be established by differentiating (4.102) with respect to \( \beta \) and \( \lambda_j \) \quad (k < j < i) .

\[
[]
\]

Corollary 4.40

Under the hypotheses of theorem 4.38, the signal-settings \( \lambda_i \) \quad (k < i \leq m) \) given by (4.102) do not decrease as the mean bus arrival rate increases. Furthermore, for each \( i \) \quad (k < i \leq m), \quad \text{if} \quad \lambda_i^* - \gamma_i > 0 \quad \text{then} \quad \lambda_i \quad \text{increases strictly with} \quad \beta .

Proof

From expressions (4.102), \( \lambda_i = \lambda_i^* \quad \text{if} \quad 1 \leq i \leq k \) \quad or \quad if \quad k < i \leq m \quad and \quad \lambda_i^* - \gamma_i = 0 \).

Thus any such \( \lambda_i \) is constant and hence non-decreasing in \( \beta \). Now suppose that \( k < i \leq m \) \quad and \quad \lambda_i^* - \gamma_i > 0 \). Rearranging (4.106) and using (4.36) gives

\[
\frac{3\lambda_i}{3\beta} = \frac{(\lambda_i^* - \gamma_i) + \left[ \frac{\tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)}{\beta (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \right] \log_e \left[ \frac{\tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)}{\tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \right] + \frac{(\tilde{p}_0 (\Psi_i + \gamma_i))}{\beta (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \log_e \left[ \frac{\tilde{p}_0 (\Psi_i + \gamma_i)}{\tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \right]}{\beta (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i)} \tag{4.108}
\]

In order for \( \lambda_i \) \quad given by (4.102) to exist, \( \tilde{p}_0 (\Psi_i + \gamma_i) - \beta (\lambda_i^* - \gamma_i) > 0 \). Furthermore, from (4.36), \( \tilde{p}_0 (\Psi_i + \gamma_i) < 1 \) \quad for all \( \beta > 0 \).
Now consider the function \( f(\xi) = \xi \log e^\xi \). This has derivative given by
\[
\frac{df}{d\xi} = 1 + \log e^\xi \quad \text{which satisfies} \quad 1 + \log e^\xi < 1 \quad \text{for} \quad 0 < \xi < 1.
\]
Applying the mean value theorem (Ortega and Rheinboldt, 1970) to \( f(\xi) \) gives that for each
\( i \quad (k < i \leq m) \), there exists \( \xi_i \) in the interval
\[
[\tilde{\rho}_0(\Psi_i + \gamma_i) - \beta(\lambda_i^* - \gamma_i), \tilde{\rho}_0(\Psi_i + \gamma_i)]
\]
with
\[
[\tilde{\rho}_0(\Psi_i + \gamma_i) - \beta(\lambda_i^* - \gamma_i)] \log e [\tilde{\rho}_0(\Psi_i + \gamma_i) - \beta(\lambda_i^* - \gamma_i)] - \tilde{\rho}_0(\Psi_i + \gamma_i) \log e [\tilde{\rho}_0(\Psi_i + \gamma_i)]
\]
\[
= - (\lambda_i^* - \gamma_i)(1 + \log e \xi_i) \quad \text{for} \quad (k < i \leq m)
\]
\[
> - (\lambda_i^* - \gamma_i)
\]
Using (4.109) in (4.108) and again that \( \tilde{\rho}_0(\Psi_i + \gamma_i) - \beta(\lambda_i^* - \gamma_i) > 0 \) gives
\[
\frac{\partial \lambda_i}{\partial \beta} > 0 \quad \text{for} \quad (k < i \leq m) \quad (4.110)
\]
Similarly, using \( \lambda_i^* - \gamma_i > 0 \) and \( \tilde{\rho}_0(\Psi_i + \gamma_i) - \beta(\lambda_i^* - \gamma_i) > 0 \) in (4.107) gives
\[
\frac{\partial \lambda_i}{\partial \lambda_j} > 0 \quad \text{for} \quad (k < j < i), (k < i \leq m) \quad (4.111)
\]
Using (4.110) and (4.111) in (4.105) gives that if \( \lambda_i^* - \gamma_i > 0 \), then
\[
\frac{d\lambda_i}{d\beta} \geq 0 \quad \text{for} \quad (k < i \leq m) \quad (4.112)
\]

**Lemma 4.41**

For each \( i \quad (k < i \leq m) \), if \( \lambda_i^* - \gamma_i > 0 \), then there is a unique value \( \beta_i \) of \( \beta \) which satisfies the equation
\[
\tilde{\rho}_0(\sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k}^{i-1} \lambda_j + \gamma_i) - \beta(\lambda_i^* - \gamma_i) = 0 \quad (4.113)
\]
where \( \lambda_j \quad (k < j < i) \) are given by (4.102). Furthermore, if \( \lambda_i^* - \gamma_i > 0 \), \( \lambda_i^* - \gamma_i > 0 \)
and \( k < j < i \), then \( \beta_j > \beta_i \).

**Proof**

The proof is by induction on \( i \). Note that if \( \lambda_i^* - \gamma_i = 0 \) for each \( i \quad (k < i \leq m) \), then the hurry-call facility has been disabled by the choice of the minimum green times. In this case no index \( i \) satisfies the hypotheses of the lemma, so there is nothing to prove.
Suppose now that the result holds for all indices not exceeding some \( n \) for which \( \lambda_n^{\#} \gamma_n > 0 \). Thus if \( k < \ell < i \leq n \), \( \lambda_{\ell}^{\#} \gamma_{\ell} > 0 \) and \( \lambda_i^{\#} \gamma_i > 0 \), then each of \( \beta_{\ell} \) and \( \beta_i \) is the unique solution to the appropriate version of (4.113) and \( \beta_{\ell} < \beta_i \). If \( n = m \) or \( \lambda_{j}^{\#} \gamma_j = 0 \) (\( n < j \leq m \)), then the proof is complete. Otherwise, let \( N = \min \{ j | n < j \leq m, \lambda_j^{\#} \gamma_j > 0 \} \).

As a consequence of the inductive hypothesis, for each \( i \leq N \), \( \lambda_i \) given by (4.102) is well defined for \( \beta < \beta_n \). By the continuity of the expressions (4.102) and of the left-hand side of (4.113), as \( \beta \to \beta_n^{\pm} \),

\[
\beta_0 \left( \sum_{j=k+1}^{\ell} \lambda_j + \sum_{j=\ell}^{n} \lambda_j^{\#} \gamma_j \right) - \beta \gamma_n \to 0,
\]

so from (4.102), \( \lambda_n \to \infty \). Since \( N \geq n \), as \( \beta \to \beta_n^{\pm} \), \( \beta_0 \left( \sum_{j=k+1}^{\ell} \lambda_j + \sum_{j=\ell}^{n} \lambda_j^{\#} \gamma_j \right) \to 0 \); but by hypothesis, \( \beta \gamma_n \to 0 \) so the left-hand side of (4.113) with \( i = N \) is strictly negative when \( \beta = \beta_n \). As the left-hand side of (4.113) is continuous in \( \beta \) and takes the value 1 when \( \beta = 0 \), (4.113) is satisfied when \( i = N \) by some \( \beta_n \), \( 0 < \beta_n < \beta_n \).

This proves the existence of \( \beta_n < \beta_n \): the inductive hypothesis gives \( \beta_n < \beta_i \) for each \( i \) (\( k < i < n \)) where \( \lambda_i^{\#} \gamma_i > 0 \), so the inequalities are also established. From the result of corollary 4.40, each \( \lambda_j \) (\( k < j < N \)) is a non-decreasing function of \( \beta \), so the left-hand side of (4.113) is a strictly decreasing function of \( \beta \) and thus has no more than one root for each \( i \).

For the initial step when \( N = \min \{ j | k < j \leq m, \lambda_j^{\#} \gamma_j > 0 \} \), only existence and uniqueness of \( \beta_N \) need be proved since there is no \( \beta_i \) of lesser index with which to compare \( \beta_N \). Here, the left-hand side of (4.113) is a continuous and unbounded function of \( \beta \) defined on \( [0, \infty) \). It takes the value 1 when \( \beta = 0 \) and tends to \( -\infty \) as \( \beta \) increases without bound. Thus the required solution of (4.113) exists. Uniqueness follows as in the inductive step.
Corollary 4.42

Under the hypotheses of theorem 4.38, \( \beta_c \), the least upper bound on values of \( \beta \) for which solutions to (4.102) exist, satisfies Equation (4.113) with \( i=j \) where

\[
J = \max \{ j | k < j \text{ m}, \lambda_j^*-\gamma_j > 0 \}
\]  

(4.114)

Proof

The upper bound on values of \( \beta \) for which solutions to all of the Equations (4.102) exist is the least value for which any of the inequalities

\[
\beta_c \; \tilde{P}_0 \left( \sum_{j=k+1}^{i-1} \lambda_j^* + \sum_{j=k}^{i-1} \lambda_j^0 + \gamma_i^* \right) - \beta \left( \gamma_i^* - \gamma_i \right) > 0 \quad (k \leq m)
\]  

(4.115)

is violated. From the result of lemma 4.41, the required value is \( \beta_j \) where \( J \) is given by (4.114).

The result of corollary 4.42 provides a method of calculating the maximum mean bus arrival rate for which priority by hurry-call can be implemented at a junction without causing any loss in capacity. Since Equations (4.102) and (4.113) depend on \( \{ \gamma_i \} \), the possibility arises of using these minimum green times to control the range of mean bus arrival rates for which solutions to (4.102) exist. If a minimum green time for a stage is increased, then the amount of priority which can be given to buses is reduced. This is because any bus which arrives at the detector before the end of the minimum permissible time for that stage will gain right of way only after some additional time has elapsed. The following results investigate the effect of changes in minimum green times on the value of \( \beta_c \).

Theorem 4.43

Under the hypotheses of theorem 4.38, the value of \( \beta_c \) varies with the minimum green times \( \gamma_h \) (for \( k < h < J \)) according to the differential equations
\[
\frac{\partial \beta_c}{\partial \gamma_h} = \left[ \beta \tilde{P}_0^\gamma \sum_{j=0}^{J} \left( \lambda_j \cdot (1, 0)^{Y_j} \right) \right] - \beta \sum_{i=0}^{\lambda_h} \left( \lambda_i \cdot (1, 0)^{Y_i} \right) \tilde{P}_0 \left[ \sum_{j=0}^{J} \left( \lambda_j \cdot (1, 0)^{Y_j} \right) \right] + \beta \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \right]
\]

where \( J \) is given by (4.114).

**Proof**

According to the result of corollary 4.42, \( \beta_c \) satisfies (4.113) when \( i = J \)
given by (4.114) and \( \lambda_j \) (\( k < j < J \)) are given by (4.102). Thus

\[
\lambda_i = \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} - \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} - \beta \left( \lambda_i \cdot (1, 0)^{Y_i} \right) \right]
\]

\( (k < i < J) \)

\[
\Rightarrow \tilde{P}_0 \left[ \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} \right] = \tilde{P}_0 \left[ \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} \right] - \beta \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} - \beta \left( \lambda_i \cdot (1, 0)^{Y_i} \right) \right]
\]

\( (k < i < J) \)

\[
\Rightarrow \tilde{P}_0 \left[ \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} \right] = \tilde{P}_0 \left[ \sum_{j=k+1}^{j=i-1} \lambda_j \cdot (1, 0)^{Y_j} - \beta \left( \lambda_i \cdot (1, 0)^{Y_i} \right) \right]
\]

\( (k < i < J) \)

Summing (4.117) over \( i \) (\( k < i < J \)) and using (4.113) with \( i = J \) gives

\[
\left[ \beta \tilde{P}_0 \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} - \beta \tilde{P}_0 \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \right] = 0
\]

(4.118)

all other terms of the summation over \( i \) cancelling in pairs. Treating the

left-hand side of (4.118) as a function \( F(\beta, \{ Y_j \}) \) which has a root when

\( \beta = \tilde{\beta}_c \) (\( \{ Y_j \} \)) and differentiating with respect to \( \beta \) gives

\[
\frac{\partial F}{\partial \beta} = -\sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \right] - \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \right] - \beta \tilde{P}_0 \left[ \sum_{j=0}^{J} \lambda_j \cdot (1, 0)^{Y_j} \right]
\]

(4.119)

Using (4.118) gives

\[
\frac{\partial F}{\partial \beta} \bigg|_{\beta = \tilde{\beta}_c} =
\]

(4.120)
Similarly,
\[ \frac{\partial F}{\partial Y_h} = -\beta \bar{p}_0 \left[ \sum_{j=k+1}^{J} (\lambda_j,0^+,Y_j) \right] + \beta \bar{p}_0 \left[ \sum_{j=h+1}^{J} (\lambda_j,1,0^+,Y_j) \right] + \beta \bar{p}_0 \left[ \sum_{j=i+1}^{J} (\lambda_j,1,0^+,Y_j) \right] \]
\[ + \beta \bar{p}_0 \left[ \sum_{j=h+1}^{J} (\lambda_j,0^+,Y_j) \right] \] (k < h ≤ J) \hfill (4.121)

and
\[ \frac{\partial F}{\partial Y_h} \bigg|_{\beta=\beta_c} = \left[ \beta \left[ \bar{p}_0 \left[ \sum_{j=h+1}^{J} (\lambda_j,0^+,Y_j) \right] - \beta \bar{p}_0 \left[ \sum_{j=i+1}^{J} (\lambda_j,1,0^+,Y_j) \right] \right] \right] \bigg|_{\beta=\beta_c} \] (k < h ≤ J) \hfill (4.122)

Now, from the implicit function theorem A1.1, the derivative of \( \beta_c \) with respect to \( Y_h \) can be expressed as
\[ \frac{d\beta_c}{dY_h} = -\left[ \frac{\partial F}{\partial Y_h} \bigg|_{\beta=\beta_c} \right] \cdot \left[ \frac{\partial F}{\partial Y_h} \bigg|_{\beta=\beta_c} \right]^{-1} \] (k < h ≤ J) \hfill (4.123)

where the derivative is total rather than partial because \( Y_h \) can be varied independently of the other variables upon which \( F \) depends. Substituting (4.120) and (4.122) into (4.123) gives (4.116).

Corollary 4.44

Under the hypotheses of theorem 4.38, \( \beta_c \) increases strictly with any increase in \( Y_h \) (k < h ≤ J) where \( J \) is given by (4.114).

Proof

Using (4.120) and (4.121) in (4.123) and rearranging gives
\[ \frac{d\beta_c}{dY_h} = \left[ \beta \bar{p}_0 \left[ \sum_{j=k+1}^{J} (\lambda_j,0^+,Y_j) \right] - \beta \bar{p}_0 \left[ \sum_{j=i+1}^{J} (\lambda_j,1,0^+,Y_j) \right] \right] \bigg|_{\beta=\beta_c} \] (k < h ≤ J) \hfill (4.124)

The denominator of (4.124) is strictly positive, and since \( \bar{p}_0(\xi) ≤ 1 \) for \( \xi ≥ 0 \), so is the numerator. Thus \( \frac{d\beta_c}{dY_h} > 0 \) (k < h ≤ J).

According to the result of corollary 4.42, Equations (4.102) can be solved for \( \{\lambda_j\} \) whenever the mean bus arrival rate is less than \( \beta_c / c^* \). However, as was noted in the proof of lemma 4.41, \( \lambda_j → \infty \) as \( \beta → \beta_c^- \), so the
duration of cycles in which no hurry-call is granted increases without bound as the mean bus arrival rate for which correction is made increases towards $\beta_c/c^*$. In most cases, there will be an upper limit on the cycle time above which cycles are considered to be excessively long. The next result provides a formal statement of the extra restriction on the mean bus arrival rate for which acceptable solutions to (4.102) exist which is consequent upon this consideration.

Lemma 4.45

Under the hypotheses of theorem 4.38 and the additional hypothesis that $\lambda_1^* - \gamma_i > 0$ for some $i$ (k<i≤m), let $\beta_\ell$ be the greatest mean bus arrival rate below which signal-settings $\{\lambda_i^*\}$ exist which are capacity-equivalent to $\{\lambda_i^\ell\}$ and which give rise to cycles of duration no greater than $c_\ell c^*$ in all circumstances. Then $\beta_\ell$ is the unique value of $\beta$ which satisfies

$$\sum_{i=1}^{m} \lambda_i + \lambda_0^* - c_\ell c^* = 0$$

(4.125)

where $\lambda_i$ (1≤i≤m) are given by (4.102).

Proof

Since granting priority by hurry-call reduces the duration of a cycle, the maximum duration occurs when no hurry-call is granted. The duration of an uninterrupted cycle where the signal-settings implemented are capacity-equivalent to $\{\lambda_i^\ell\}$ is $\sum_{i=1}^{m} \lambda_i + \lambda_0^* c^*$ where $\{\lambda_i^\ell\}$ are given by (4.102). This is no greater than $c_\ell$ if and only if the inequality

$$\sum_{i=1}^{m} \lambda_i + \lambda_0^* - c_\ell c^* \leq 0$$

(4.126)

is satisfied. Since $c_\ell c^*$ and the values $\lambda_i = \lambda_i^\ell$ (1≤i≤m) satisfy (4.102) when $\beta=0$, this certainly occurs when $\beta=0$. Now $\lambda_\ell \to \infty$ as $\beta + \beta_c^-$, so (4.126) is violated wherever $\beta$ is sufficiently close to $\beta_c$. The values of $\{\lambda_i^\ell\}$ given by (4.102) vary continuously with $\beta<\beta_c$ and from the
result of corollary 4.40 do not decrease as $\beta$ increases. Since for some $i \ (k<i\leq m) \quad \lambda^* \gamma_1 > 0$, the left-hand side of (4.126) increases strictly with $\beta$ so there is a single interval $[0, \beta^*]$ in which (4.126) is satisfied. The inequality (4.126) is satisfied as an equality only when $\beta = \beta^*$, so this is the unique solution of (4.125).

4.5.5 Priority by recall.

The problem of finding suitable signal-settings for use with priority by recall is considered in this sub-section. The approach used is similar to that of sub-section 4.5.4 and most of the results presented here correspond to results appropriate to priority by hurry-call. However, the analysis in this sub-section is less extensive than that in the previous one: this is a consequence of the expressions for the signal-settings derived here not being in closed form but rather requiring solution by some iterative method. A discussion of properties which would be sufficiently strong to guarantee the existence and uniqueness of the desired signal-settings and the convergence of a proposed solution method is given here with reference to the general results of Appendix 1.

Again, the basic result of the sub-section is given first in

Theorem 4.46

Suppose that priority by recall is provided for buses arriving at a mean rate of $\beta/c^*$. Then the signal-settings $\lambda_i$ which are capacity-equivalent to the set $\{\lambda_i^*\}$ satisfy the equations

$$
\lambda_i = \begin{cases} 
\frac{\lambda^*}{\lambda_0} \left[ \sum_{j=1}^{\bar{P}_1} \lambda_{j+1} \gamma_j + \sum_{n=k+1}^{m} \tilde{P}_0 \sum_{j=k+1}^{n-1} \lambda_{j+1} \sum_{j=k}^{n-2} \lambda_j \right] \\
\frac{1}{\beta} \log_e \left( \frac{\bar{P}_1(\{\lambda_i\})}{\bar{G}_i(\{\lambda_j\})} \right) 
\end{cases}
$$

where

$$
\left( \begin{array}{c}
\lambda^* \\
\lambda_0 \\
\bar{P}_1 \\
\bar{G}_i \\
\end{array} \right) = \left( \begin{array}{c}
\sum_{j=1}^{\bar{P}_1} \lambda_{j+1} \gamma_j \\
\sum_{n=k+1}^{m} \tilde{P}_0 \sum_{j=k+1}^{n-1} \lambda_{j+1} \sum_{j=k}^{n-2} \lambda_j \\
\sum_{j=1}^{\bar{P}_1} \lambda_{j+1} \gamma_j + \sum_{n=k+1}^{m} \tilde{P}_0 \sum_{j=k+1}^{n-1} \lambda_{j+1} \sum_{j=k}^{n-2} \lambda_j \\
1 \\
\end{array} \right)
$$

(4.127)
\[ \tilde{\tilde{P}}_i(\{\lambda_j\}) = \lambda_{i-1} \tilde{p}_i(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0) + \beta \lambda_i^* \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0) \times \left( \lambda_{n-1},0,1,1 \right) \quad (k \leq i \leq m) \]

\[ \tilde{\tilde{G}}_i(\{\lambda_j\}) = \lambda_{i-1} \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0) \left[ \beta Y_i + \tilde{p}_0(\lambda_{i-1},0,Y_i) \right] - \beta \lambda_i^* \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0)(\lambda_{n-1},0,1,1) \quad (k \leq i \leq m) \]

\[ \tilde{\tilde{P}}_0(\lambda_i) = \sum_{i=1}^{m} \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0) \left( \lambda_{n-1},0,1,1 \right) + \beta \lambda_i \quad \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0)(\lambda_{n-1},0,1,1) \quad (k \leq i \leq m) \]

Proof

According to the result of lemma 4.32, the required signal-settings \{\lambda_i\} satisfy (4.86). Multiplying both sides of (4.86) by \( \varepsilon_0 \) and substituting the specific forms (4.63) for \( \varepsilon_i \) (0 ≤ i ≤ k) when priority by recall is provided gives the first k of equations (4.127). For each \( i \) (k ≤ i ≤ m), the expressions (4.63) can be rearranged to give

\[ \varepsilon_0 = \sum_{j=1}^{k+1} \lambda_j^0 + \lambda_{k+1,1} + \sum_{j=k+1}^{k+1} \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0)(\lambda_{n-1},0,1,1) + \sum_{i=1}^{m} \tilde{p}_0(\lambda_i) \left[ \beta Y_i + \tilde{p}_0(\lambda_{i-1},0,Y_i) \right] - \beta \lambda_i \quad \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0)(\lambda_{n-1},0,1,1) \quad (k \leq i \leq m) \]

and

\[ \varepsilon_i = \tilde{p}_0(\sum_{j=k+1}^{k+1} \lambda_j + \sum_{j=k+1}^{k+1} \lambda_j^0)(\lambda_{n-1},0,1,1) \]

Substituting (4.129) into (4.86) and rearranging with the use of (4.36) gives the last (m − k) of equations (4.127).

Unlike Equations (4.102), equations (4.127) do not generally give closed forms for \{\lambda_i\}. Although each \lambda_i is expressed as an explicit function of \lambda_j (k ≤ i ≤ m, j ≠ i), use is made of all the values of these variables, so the recursive method used to solve equations (4.102) cannot be applied here. However, the forms (4.127) are convenient for application of the Gauss-Seidel solution method as described in Section A1.3. This is because successive values of the variables \lambda_i (1 ≤ i ≤ m) can be calculated directly from (4.127).

The dependence of \lambda_i (k ≤ i ≤ m) on \lambda_j (i ≤ j ≤ m) is only through and the effect of changes in \lambda_j (k ≤ j ≤ m) on the value of \varepsilon_0 is expected to be small compared to their effect on the values of \varepsilon_i (k ≤ i ≤ m). Thus a natural way in which to apply the Gauss-Seidel method is to start with the
expression for \( \lambda_{k+1} \) and then to proceed to successively higher indices of \( \lambda \). This solution method correspond to iterated application of the recursive solution method used in sub-section 4.5.4 for equations (4.102).

The next results prove that a unique solution to equations (4.127) exists and that the Gauss-Seidel method converges to it provided that the mean bus arrival rate is sufficiently small.

**Lemma 4.47**

Solving equations (4.127) for \( \lambda_i \) (1 \( \leq i \leq m \)) is equivalent to solving the equations

\[
\lambda_i - \frac{1}{\beta} \log_e \left[ \frac{\tilde{F}_i(\{\lambda_j\})}{\tilde{G}_i(\{\lambda_j\})} \right] = 0 \quad (k < i \leq m)
\]

(4.130)

for \( \lambda_i \) (k \( < i \leq m \)) where \( \tilde{F}_i(\{\lambda_j\}) \) and \( \tilde{G}_i(\{\lambda_j\}) \) are given by (4.128).

The elements of the Jacobian matrix \( J \) of the left-hand side of (4.130) are given by

\[
J_{i\ell} = D_{i\ell} - L_{i\ell} - U_{i\ell} \quad (k < i \leq m), (k < \ell \leq m)
\]

(4.131)

where

\[
D_{i\ell} = \delta_{i\ell} \quad (k < i \leq m), (k < \ell \leq m)
\]

\[
L_{i\ell} = \begin{cases} 
\frac{B \lambda_i}{\lambda_i} \sum_{j=1}^{\ell} \lambda_j + \lambda_{k+1,1} + \sum_{n=k+2}^{\ell} \tilde{F}_0(j_{k+1}+1, j_{k+1}+j_{k} \lambda_{j_0}) (\lambda_{n-1,0} + \lambda_{n,1} - \lambda_{n-1,1}) \\
0 & (k < i \leq m), (k < \ell < i) \\
0 & (k < i \leq m), (i < \ell \leq m)
\end{cases}
\]

(4.132)

\[
U_{i\ell} = \begin{cases} 
-\beta \lambda_i \sum_{n=\ell+1}^{m} \tilde{F}_0(i_{k+1}+1, j_{i+1}+1, j_{i}+n_{i-1}+1, j_{i+1}+j_{i}+\lambda_{j_0}) (\lambda_{n-1,0} + \lambda_{n,1} - \lambda_{n-1,1}) \\
0 & (k < i \leq m), (i < \ell < i) \\
0 & (k < i \leq m), (i < \ell \leq m)
\end{cases}
\]

(4.132)

**Proof**

Since the first \( k \) of equations (4.127) give \( \lambda_i \) (1 \( < i \leq k \)) as explicit functions of \( \lambda_i \) (k \( < i \leq m \)), only the last (m-k) of Equations (4.127) need
be solved simultaneously. Rearranging these last \((m-k)\) equations gives (4.130) as required.

Differentiating the forms of \(F_i\) and \(G_i\) \((k \leq m)\) given in (4.128) with respect to \(\lambda_\ell\) \((k \leq m)\) gives

\[
\frac{\partial F_i}{\partial \lambda_\ell} = \begin{cases} 
-\beta F_i & \text{if } (k \leq m), (k < \ell < i) \\
0 & \text{if } (k \leq m), (\ell = i) \\
\beta z_i^* \sum_{j=1}^m \sum_{j=k+1}^{i-1} \left( \sum_{j=1}^i \lambda_j + \sum_{j=k+1}^{i-1} \lambda_j \right) (\lambda_{n-1} + \lambda_n, 1 - \lambda_{n-1}, 1) & \text{if } (k < i \leq m), (i < \ell \leq m) \\
\beta^2 \lambda_i^* \sum_{j=1}^m \sum_{j=k+1}^{i-1} \sum_{j=1}^{i-1} \left( \sum_{j=1}^i \lambda_j + \sum_{j=k+1}^{i-1} \lambda_j \right) (\lambda_{n-1} + \lambda_n, 1 - \lambda_{n-1}, 1) & \text{if } (k \leq m), (k < \ell < i)
\end{cases}
\]

Differentiating the left-hand side of (4.130) with respect to \(\lambda_\ell\) \((k \leq m)\) gives

\[
J_{i\ell} = \delta_{i\ell} + \beta \frac{\partial G_i}{\partial \lambda_\ell} \frac{1}{G_i} - \frac{\partial F_i}{\partial \lambda_\ell} \frac{1}{F_i} \quad (k \leq m), (k < \ell \leq m)
\]  
(4.134)

Substituting for the partial derivatives of \(F_i\) and \(G_i\) from (4.133) in (4.134) and separating into diagonal, strictly lower and strictly upper triangular matrices \(D, L\) and \(U\) respectively gives (4.131) and (4.132).

Theorem 4.48 (Local existence of solutions to (4.127))

For some strictly positive \(\beta_e\), a unique solution to Equations (4.127) exists for each \(\beta\) in the interval \([0, \beta_e]\). Furthermore, the solutions vary continuously with \(\beta\) in this interval.

Proof

According to the result of lemma 4.47, the result is established if
solutions to equations (4.130) are shown to have the desired properties. When $\beta = 0$, $F_i - G_i = \lambda_1^*$ ($k < i \leq m$), so the left-hand sides of (4.130) are indeterminate. Applying the rule of Count de l'Hoiple when $\beta = 0$, (4.130) reduces to

$$\lambda_i - \left[ \frac{3F_i}{\beta} \frac{1}{F_i} - \frac{2G_i}{\beta} \frac{1}{G_i} \right] = 0 \quad (k < i \leq m)$$

$$\iff \lambda_i - \lambda_i^* = 0 \quad (k < i \leq m) \quad (4.135)$$

Thus, as would be expected, the values $\lambda_i = \lambda_i^*$ ($k < i \leq m$) satisfy (4.127) when $\beta = 0$.

Now from (4.132), when $\beta = 0$, $L_{\ell \ell} = U_{\ell \ell} = 0$ ($k < \ell \leq m$), so $\det(J)_{\beta = 0} = 1$. Furthermore, all the elements of $J$ vary continuously with $\beta$ and $\lambda_i$ ($k < i \leq m$), so the implicit function theorem A1.1 can be applied. This establishes the existence of solutions to (4.130) which vary continuously with $\beta$ in some neighbourhood of $\beta = 0$, and a fortiori in some interval $[0, \beta_e)$ where $\beta_e > 0$.

**Theorem 4.49** (Local convergence of the Gauss-Seidel method to the solutions of (4.127))

For some strictly positive $\gamma \leq \beta_e$ of theorem 4.48, for each $\beta$ in the interval $[0, \beta_e)$ the Gauss-Seidel method starting with $\lambda_i^{(0)} = \lambda_i^*$ ($k < i \leq m$) converges to $\lambda_i(\beta)$ ($k < i \leq m$), the solutions of equations (4.127).

**Proof**

Again, the proof considers the equivalent problem of solving equations (4.130). From the result of theorem 4.48, the functions $\lambda_i(\beta)$ are well defined and continuous for $\beta$ in $[0, \beta_e)$. In order for these solutions to exist, each of the quotients $F_i / G_i$ ($k < i \leq m$) must be positive. By inspection of (4.128), $\tilde{F}_i(\beta) > 0$ ($k < i \leq m$) for all $\beta$ in $[0, \beta_e)$, so
$\tilde{c}_i(\beta) > 0$ for all $\beta$ in $[0, \beta_e)$. Thus the elements of $L$ and $U$ given by (4.132) vary continuously with $\beta$ in the range $[0, \beta_e)$. Since $\det(D-U) = 1$, the elements of the matrix $H = (D-L)^{-1}U$ vary continuously with $\beta$.

Now when $\beta = 0$, $U_{i\ell} = L_{i\ell} = 0$ $(k < i \leq m), (k < \ell \leq m)$ so $H|_{\beta = 0}$ is identically equal to the zero matrix and has all eigenvalues equal to 0. Thus $\rho(H)|_{\beta = 0} = 0$ and corollary A1.5 can be applied to give the desired result.

The results of theorems 4.48 and 4.49 are local in the sense that no indication is given of the sizes of $\beta_e$ and $\beta_G$. In practice, the problem of solving (4.127) appears to be well-conditioned in that $\beta_e$ appears to correspond to $\beta_c$ of corollary A1.2 with $\lambda_{\infty}$ as $\beta \to \beta_e^-$. The next lemma gives a condition which is sufficiently strong to guarantee that this occurs. Furthermore, $\beta_G$ appears to be as large as $\beta_e$.

The next result provides a least upper bound on the mean bus arrival rate for which solutions to (4.127) exist. This corresponds to lemma 4.41 and corollary 4.42 of sub-section 4.5.4. This result is incomplete in the sense that it depends on the non-singularity of the matrix $J$ given by (4.131) and (4.132).

**Lemma 4.50**

Under the hypotheses of theorem 4.46 and the additional hypothesis that $\det(J) \neq 0$ where $J$ is given by (4.131) and (4.132), $\beta_c$, the least upper bound on $\beta$ for which solutions to (4.127) exist, satisfies the equation

$$
\sum_{m-1}^{m-2} \lambda_0^{-j} \rho_0(j = k+1, j = \sum_{k=0}^{m-2} \lambda_0^{j}) \left[ \beta \rho_0(m-1, 0) \lambda_0^{m-1} \right] - \beta \sum_{m-1}^{m-2} \rho_0(j = k+1, j = \sum_{k=0}^{m-2} \lambda_0^{j}) \left( \lambda_0^{n-1} + \lambda_0^{n-2} \lambda_0^{n-1} \right) = 0
$$

(4.136)

where $\lambda_i$ $(k < i \leq m)$ satisfy (4.127). Furthermore, $\tilde{\lambda}_i(\beta)$ $(1 \leq i \leq m)$ are uniquely defined continuous functions of $\beta$ in $[0, \beta_c)$. 
Proof

A necessary condition for solutions of (4.127) to exist is that each of the required logarithms exists. Using the notation (4.128), since \( F_i > 0 \) (\( k < l \leq m \)) for \( \beta > 0 \), this requirement is equivalent to \( G_i > 0 \) (\( k < l \leq m \)). Let \( \beta_c \) be the least positive value of \( \beta \) for which any \( G_i = 0 \). Since each \( G_i \) (\( k < l \leq m \)) is non-zero in \([0, \beta_c)\), each of the expressions (4.127) is finite. Together with the local existence theorem 4.48 and the hypothesis that \( \det(J) \neq 0 \), this is sufficient to apply the corollary A1.2 of the implicit function theorem to show that \( \lambda_i(\beta) \) (\( 1 \leq i \leq m \)) are uniquely defined continuous functions of \( \beta \) in \([0, \beta_c)\).

Now suppose that \( G_i = 0 \) at \( \beta = \beta_c \) for some \( k < l \leq m \). As \( \beta + \beta_c^- \), \( G_i \to 0 \) and \( \lambda \to \infty \) so \( G_i \to -\beta \lambda^k \sum_{j=1}^{\infty} \lambda_j^{j+1} \lambda^{k+1} \alpha_j < 0 \). Since \( G_m > 0 \) when \( \beta = 0 \) and \( G_m \) varies continuously with \( \beta \) in \([0, \beta_c)\), \( G_m = 0 \) for some \( \beta \) in \([0, \beta_c)\), contradicting the minimality of \( \beta_c \). Thus \( G_m = 0 \) when \( \beta = \beta_c \), so (4.126) is satisfied.

Due to the intricate form of the elements of the Jacobian matrix \( J \) given by (4.131) and (4.132), there is little hope of finding a general analytic proof that the additional hypothesis of non-singularity in lemma 4.50 is satisfied. Similarly, no general analytic proof is available that the matrix \( H \) satisfies the hypotheses of theorem A1.3, so only local convergence results for the Gauss-Seidel method can be provided. Two special cases in which the calculation of \( \{\lambda_i\} \) is explicit are considered next. Here, the desired properties are established directly.

Lemma 4.51

If there is only one stage during which the stream containing buses does not receive right of way, then Equations (4.102) and (4.127) reduce to the
same form:

$$
\lambda_i^* = \begin{cases} 
\lambda_i^+ & (1 \leq i \leq m) \\
-\lambda_{m-1,0}^* - \frac{1}{\beta} \log_e [\tilde{P}_0(\lambda_{m-1,0}^* + \lambda_m^*) - \beta(\lambda_m^* + \gamma_m^*)] & (i = m)
\end{cases}
$$

(4.137)

from which \{\lambda_i\} can be calculated explicitly.

Proof

Since there is only one stage during which the stream containing buses does not receive right of way, \(m=k+1\). Using this in (4.127) together with (4.36) gives (4.137). Similarly, (4.102) reduces to (4.137) under these hypotheses, so all the analysis of sub-section 4.5.4 can be applied and in particular, \{\lambda_i\} can be calculated directly.

Lemma 4.52

Under the hypotheses of theorem 4.46, if

$$
\lambda_{i1} = \sum_{j=1}^{m} \lambda_{j0}^* 
$$

(4.138)

then \{\lambda_i\} can be calculated explicitly from the formulae

$$
\lambda_i = \begin{cases} 
\lambda_i^+ & (1 \leq i \leq k) \\
-\frac{\Sigma_{j=k+1}^{i-1} \lambda_{j0}^* + \Sigma_{j=k+1}^{i-2} \lambda_{j0}^*}{\tilde{P}_0(\Sigma_{j=k+1}^{i-1} \lambda_{j0}^* + \Sigma_{j=k+1}^{i-2} \lambda_{j0}^*)} [\beta \gamma_{i-1} + \tilde{P}_0(\lambda_{i-1,0}^* + \gamma_{i-1})] - \beta \lambda_i^* & (k \leq i \leq m)
\end{cases}
$$

(4.139)

Proof

From (4.138), \(\lambda_{k+1,j} = \sum_{j=k+1}^{m} \lambda_{j0}^*\) and \(\lambda_{i-1,0}^* + \lambda_{i-1,i}^* - \lambda_{i-1,i}^* = 0\) \((k+1 \leq i \leq m)\). Using these forms in (4.127) together with the definition of \(\lambda_0^*\) and rearranging gives (4.139). Inspection of these expressions shows that each \(\lambda_i\) \((1 \leq i \leq m)\) is expressed either explicitly or in terms of \(\lambda_j\) \((k < j < i)\). Thus the values \(\lambda_i\) \((1 \leq i \leq m)\) can be calculated recursively, as in the case of priority by hurry-call.
If the additional hypotheses of lemma 4.52 are satisfied, then the analysis of the signal-settings \( \{ \lambda_i \} \) can proceed in a manner analogous to that in sub-section 4.5.4. However, these hypotheses are restrictive and unlikely to be satisfied in practice. Instead, the general equations (4.127) are analysed, although this analysis is incomplete in some respects.

The next result provides information about the behaviour of the solutions of equations (4.127) as the mean bus arrival rate varies.

**Theorem 4.53**

Under the hypotheses of theorem 4.46, the derivatives of the solutions \( \{ \lambda_i \} \) of equations (4.127) with respect to the mean bus arrival rate are given by

\[
\begin{align*}
\frac{d\lambda_i}{d\beta} &= \frac{\lambda_i^*}{\beta} \frac{\partial \epsilon_0}{\partial \beta} \quad (1 \leq i \leq k) \\
\lambda_i &= \frac{\partial \lambda_i}{\partial \beta} + \sum_{j=k+1}^m \frac{\partial \lambda_i}{\partial \lambda_j} \frac{\partial \lambda_j}{\partial \beta} + \frac{\partial \lambda_i}{\partial \epsilon_0} \frac{\partial \epsilon_0}{\partial \beta} \quad (k \leq i \leq m)
\end{align*}
\]

where

\[
\begin{align*}
\frac{\partial \lambda_i}{\partial \beta} &= -\left( \frac{\partial \epsilon_i}{\partial \lambda_i} \right)^{-1} \frac{\partial \epsilon_i}{\partial \beta} \quad (k \leq i \leq m) \\
\frac{\partial \lambda_i}{\partial \epsilon_0} &= -\left( \frac{\partial \epsilon_i}{\partial \lambda_i} \right)^{-1} \frac{\lambda_i^*}{\lambda_i^0} \quad (1 \leq i \leq m)
\end{align*}
\]

\[
\begin{align*}
\frac{\partial \epsilon_0}{\partial \beta} &= -\sum_{n=k+2}^\infty \left[ \Psi_{n-1} + \lambda_{n-1} \right] \tilde{p}_0 \left[ \Psi_{n-1} + \lambda_{n-1} \right] \left( \ lambda_{n-1,0} + \lambda_{n-1} - \lambda_{n-1} \right) \\
\frac{\partial \epsilon_i}{\partial \lambda_i} &= \tilde{p}_0 \left( \Psi_i \right) \tilde{p}_0 \left( \lambda_i \right) \quad (k \leq i \leq m) \\
\frac{\partial \epsilon_i}{\partial \beta} &= -\left( \Psi_{i-1} + \lambda_{i-1} \right) \tilde{p}_0 \left( \Psi_{i-1} + \lambda_{i-1} \right) \left( \Psi_{i+1,0} - \gamma_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_i - \gamma_i \right) \right] / \beta \\
&\quad - \left( \lambda_{i-1,0} + \gamma_i \right) \tilde{p}_0 \left( \Psi_{i+1,0} - \gamma_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_i - \gamma_i \right) \right] / \beta \\
&\quad - \tilde{p}_0 \left( \Psi_{i+1,0} - \gamma_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_i - \gamma_i \right) \right] / \beta^2 \quad (k \leq i \leq m)
\end{align*}
\]

\[
\begin{align*}
\frac{\partial \epsilon_i}{\partial \lambda_j} &= -\beta \tilde{p}_0 \left( \Psi_{i-1} + \lambda_{i-1} \right) \left( \Psi_{i+1,0} + \gamma_i \right) \left[ 1 - \tilde{p}_0 \left( \lambda_i - \gamma_i \right) \right] / \beta \quad (k \leq i \leq m) \\
&\quad (k < j < i)
\end{align*}
\]
and \[ \psi_i = \begin{cases} -\lambda_{k-1} & (i=k) \\ \sum_{j=k+1}^{i-1} \lambda_j + \sum_{j=k+1}^{i} \lambda_j & (k<i\leq m) \end{cases} \]

**Proof**

From the result of theorem 4.46, the signal settings \( \{\lambda_i\} \) satisfy the capacity-equivalence relations (4.86) which are equivalent to

\[ \tilde{e}_i(\{\lambda_j\}, \beta) - \frac{\lambda_i}{\lambda_0} e_0 = 0 \quad (1\leq i\leq m) \quad (4.143) \]

where \( e_i \) and \( e_0 \) are given by (4.63). The forms (4.142) can be established by differentiation of these expressions for \( e_i \) and \( e_0 \).

Similarly,

\[ \frac{\partial e_i}{\partial \lambda_j} = \begin{cases} 0 & (k<j\leq m), (k<i<j) \\ \text{or } (1\leq i\leq k), (1\leq j\leq m), (j\neq i) \\ 1 & (1\leq j\leq k), (i=j) \end{cases} \quad (4.144) \]

Applying the implicit function theorem A1.1 to the left-hand side of (4.143) as a function of \( \lambda_i \) \((1\leq i\leq m)\), \( e_0 \) and \( \beta \) gives (4.141) and, as a result of (4.144),

\[ \frac{\partial e_i}{\partial \beta} = 0 \quad (1\leq i\leq k) \quad (4.145) \]

\[ \frac{\partial \lambda_i}{\partial \lambda_j} = 0 \quad \{ (k<j\leq m), (k<i<j) \} \quad \text{or } (1\leq i\leq k), (1\leq j\leq m), (j\neq i) \quad (4.146) \]

The total derivatives \( \frac{d\lambda_i}{d\beta} \) \((1\leq i\leq m)\) can be expanded as

\[ \frac{d\lambda_i}{d\beta} = \frac{\partial \lambda_i}{\partial \beta} + \sum_{j=1}^{i-1} \frac{\partial \lambda_i}{\partial \lambda_j} \frac{d\lambda_j}{d\beta} + \sum_{j=i+1}^{m} \frac{\partial \lambda_i}{\partial \lambda_j} \frac{d\lambda_j}{d\beta} + \frac{\partial \lambda_i}{\partial e_0} \frac{d e_0}{d\beta} \quad (1\leq i\leq m) \quad (4.146) \]

which, together with (4.144) and (4.145) gives (4.140).
Corollary 4.54

Under the hypotheses of theorem 4.46 and the additional hypotheses
\[ \lambda_{i-1,0} + \lambda_{i,1} \geq \lambda_{i-1,1} \quad \text{for } (k+1 \leq i \leq m) \]  
(4.147)

the signal settings \( \lambda_i \) (1 \leq i \leq k) do not increase with increasing mean bus arrival rate. Furthermore, if \( m > k+1 \) and at least one of (4.147) is satisfied as a strict inequality, then each \( \lambda_i \) (1 \leq i \leq k) decreases strictly with increasing mean bus arrival rate.

Proof

From (4.142) and (4.147), \( \frac{\partial e_0}{\partial \beta} \leq 0 \). Using this in (4.140) gives \( \frac{\partial \lambda_i}{\partial \beta} \leq 0 \) (1 \leq i \leq k), so \( \lambda_i \) (1 \leq i \leq k) do not increase as \( \beta \) increases. If \( \lambda_{i-1,0} + \lambda_{i,1} > \lambda_{i-1,1} > 0 \) for some \( i \) (k \leq i \leq m), then from (4.142) \( \frac{\partial e_0}{\partial \beta} < 0 \) giving \( \frac{\partial \lambda_i}{\partial \beta} < 0 \) (1 \leq i \leq k) so \( \lambda_i \) (1 \leq i \leq k) decrease strictly with increasing \( \beta \).

No general result applies to the sign of \( \frac{\partial \lambda_i}{\partial \beta} \) (k \leq i \leq m). As might be expected, all but the last term of (4.140) are positive but under the additional hypotheses (4.147), the last term is negative. Using this observation, realistic examples can be constructed in which the last term dominates. In particular, if \( \lambda_{k+1} \) is small, \( m > k+1 \) and
\[ m-1,0 + \lambda_{m-1,1} > \lambda_{m,1} > \lambda_{m,0} \]
then \( \frac{\partial \lambda_{k+1}}{\partial \beta} \) can be negative. An example where this occurs is given in Appendix 2.

The lack of monotonicity of the solutions to (4.127) means that the duration of uninterrupted cycles may not increase with the mean bus arrival rate. Thus while restrictions on the duration of uninterrupted cycles will give rise to an upper bound on the mean bus arrival rate for which signal-settings can be found which are capacity-equivalent to a given set, unlike the case considered in lemma 4.45, equation (4.125) may not have a unique solution when \( \lambda_i \) (1 \leq i \leq m) satisfy (4.127). Rather, the value of the upper bound \( \beta^*_c \) associated with the maximum acceptable duration of an uninterrupted cycle, \( c^*_c \), is here the least positive solution of (4.125).
4.5.6 The inhibition rule.

The last part of this section is concerned with the effect of the inhibition rule on solutions to equations (4.86). No existence or convergence results are given here since the only ones available are similar to those of sub-section 4.5.5 in that they are local and non-constructive. Instead, the analysis is restricted to an exposition of the equations to be solved and an evaluation of limiting values of the solutions as the mean bus arrival rate increases without bound. An analysis is given of various possible restrictions on the mean bus arrival rate for which acceptable solutions exist.

The main result of this sub-section is given in

**Theorem 4.55**

If priority by hurry-call (recall) with inhibition is provided for buses arriving at a mean rate of $\beta/e^*$, then the signal-settings $\{\lambda^*_i\}$ which are capacity-equivalent to the set $\{\lambda_i\}$ satisfy the equations

\[
\frac{P_I}{P_I} \frac{\lambda^*_i}{\lambda_0^*} + (1-P_I) \lambda^*_i \quad (0 \leq i \leq m) \tag{4.148}
\]

where $P_I$ is given by (4.68) and $\lambda^*_i$ (0$\leq i \leq m$) are given by (4.61) (resp. (4.63)).

**Proof**

As in the proof of lemmas 4.28 and 4.30, the quantities $\epsilon^+_i$ (0$\leq i \leq m$) are related to the conditional expectations $\epsilon^+_i$ by the formula

\[
\epsilon^+_i = P_I \lambda^*_i + (1-P_I) \epsilon^+_i \quad (0 \leq i \leq m) \tag{4.149}
\]

where $P_I$ is given by (4.68). The values of $\epsilon^+_i$ (0$\leq i \leq m$) for the priority
method in use are given by (4.61) (resp. (4.63)). Substitution of (4.149) into (4.86) in accordance with the result of lemma 4.32 gives (4.148). []

Since \( \tilde{P}_i(\{\lambda_i\}) \) given by (4.68) is a continuously differentiable function, local existence and convergence results for solutions to equations (4.148) could be modelled on theorems 4.48 and 4.49. However, this would not provide any definite information concerning the behaviour of the solutions.

The next results give the limiting values of the solutions to (4.148) as the mean bus arrival rate increases without bound. Provided that the appropriate Jacobian matrix is non-singular for positive \( \beta \), the existence of these limiting values guarantees the existence of solutions to (4.148) in accordance with case (a) of corollary A1.2. Once again, no analytical proof is available except for special values of \( \beta \) that the Jacobian matrix is non-singular.

**Lemma 4.56**

If priority by hurry-call with inhibition is provided for buses arriving at a mean rate of \( \beta/c^* \), then the solutions of (4.148) satisfy

\[
\lim_{\beta \to \infty} \lambda_i = \begin{cases} 
\lambda_i^* & (1 \leq i \leq k) \\
\frac{2\lambda_i^*-\gamma_i}{2} & (k < i \leq m)
\end{cases}
\]

(4.150)

**Proof**

According to the result of corollary 4.29,

\[
\lim_{\beta \to \infty} \varepsilon_i = (\lambda_i + \gamma_i)/2 \\
(1 \leq i \leq m)
\]

(4.74)

Using this together with \( \varepsilon_i = \lambda_i \) (0 \leq i \leq k) from (4.71) in (4.86) gives the limiting forms of (4.148) as
\[ 0 = \begin{cases} \lambda_i - \lambda_i^* & (1 \leq i \leq k) \\ (\lambda_i^* + \gamma_i)/2 - \lambda_i^* & (k < i \leq m) \end{cases} \]  

(4.151)

These equations are satisfied by the values of \( \lambda_i \) (1 \leq i \leq m) given in (4.150).

Differentiating the right-hand side of (4.151) with respect to \( \lambda_j \) (1 \leq j \leq m) and using the continuous differentiability of the functions involved gives the following limit for the elements of the Jacobian matrix \( J \)

\[
\lim_{\beta \to \infty} J_{ij} = \begin{cases} 
0 & (i \neq j) \\
1 & (i = j), (1 \leq j \leq k) \\
1/2 & (i = j), (k < j \leq m) 
\end{cases}
\]  

(4.152)

Thus \( \lim_{\beta \to \infty} \det(J) = 2^{(k-m)} > 0 \) so the implicit function theorem A1.1 can be applied to guarantee the continuity of the values \( \lambda_i \) (1 \leq i \leq m) in the limit as \( \beta \to \infty \).

[ ]

**Lemma 4.57**

If priority by recall with inhibition is provided for buses arriving at a mean rate of \( \beta/c^k \), then the solutions of (4.148) satisfy

\[
\lim_{\beta \to \infty} \lambda_i = \begin{cases} 
\frac{(\lambda_i^0 + \sum_{j=1}^{k} \lambda_j^0 + \lambda_{k+1,1}) \lambda_i^*}{2 \lambda_0^*} & (1 \leq i \leq k) \\
\frac{\lambda_i^*}{\lambda_0^*} - \gamma_{k+1} & (i = k+1) \\
\frac{(\lambda_i^0 + \sum_{j=1}^{k} \lambda_j^0 + \lambda_{k+1,1}) \lambda_i^*}{\lambda_0^*} & (k+1 < i \leq m) 
\end{cases}
\]  

(4.153)

**Proof**

The proof is as for the last result. Using the limiting values (4.82) in (4.86) gives the limiting forms of (4.148) as
\[
0 = \begin{cases}
\frac{2\lambda_i}{(\lambda_0^* + j \lambda_j^* + k + 1,1)} & (1 \leq i \leq k) \\
\lambda_0^* & (i = k + 1) \\
\lambda_i & (k + 1 \leq i \leq m)
\end{cases}
\] (4.154)

These equations are satisfied by the values of \( \lambda_i \) \( (1 \leq i \leq m) \) given in (4.153). Continuity of the values of \( \lambda_i \) \( (1 \leq i \leq m) \) in the limit as \( \beta \to \infty \) follows as before from

\[
\lim_{\beta \to \infty} \det(J) = \frac{2^k}{(\lambda_0^* + j \lambda_j^* + k + 1,1)^m}.
\]

Provided that the appropriate Jacobian matrix is non-singular, solutions to equations (4.148) exist for all mean bus arrival rates. However, there are two particular reasons for which the solutions could be unsuitable for use in practice. The first is that a minimum stage duration constraint may be violated. If the extension priority rule is implemented in conjunction with the methods described in this sub-section, then the result of corollary 4.36 shows that the minimum constraint for the duration of stage \( k \) will certainly give rise to an upper bound on the mean bus arrival rate for which acceptable solutions to (4.148) exist. The second possibility is that the duration of an uninterrupted cycle may become unacceptably long as compensation is made for increasing mean bus arrival rates.

The next results indicate possible ways in which minimum stage duration constraints can restrict the range of mean bus arrival rates for which acceptable solutions exist. The first, which applies to priority by hurry-call with inhibition, is negative in the sense that it shows that no minimum green constraint which is satisfied by \( \{\lambda_i^*\} \) can be violated by capacity-equivalent signal settings \( \{\lambda_i\} \). Thus if priority is provided by extension as well, then the only minimum constraint which can be violated by \( \{\lambda_i\} \) is that associated with stage \( k \).
Lemma 4.58

If priority by hurry-call is provided for buses arriving at any mean rate and each of the minimum stage duration constraints
\[ \lambda_i^* - \gamma_i \geq 0 \quad (1 \leq i \leq m) \]  
(4.155)
is satisfied when \( \lambda_i^* = \lambda_i \) (1 \leq i \leq k), then each of these constraints is satisfied by the solutions of (4.148).

Proof

From (4.61), \( \lambda_0^* = \lambda_k^* \) and \( \lambda_1^* = \lambda_1 \) (1 \leq i \leq k), so for any mean bus arrival rate, the values \( \lambda_i^* = \lambda_i \) (1 \leq i \leq k) satisfy the first \( k \) of equations (4.148). Since by hypothesis \( \lambda_i^* \) (1 \leq i \leq k) satisfy (4.155), the first \( k \) of these inequalities are satisfied at all mean bus arrival rates. Furthermore, from (4.148),
\[ \lambda_i^* = P_i \lambda_i + (1-P_i) \varepsilon_i^+ \quad (k \leq i \leq m) \]  
(4.156)
and since \( P_i \) is a probability, \( 0 \leq P_i \leq 1 \). Thus for each \( i \) (k \leq i \leq m), either \( \lambda_i \geq \lambda_i^* \) or \( \varepsilon_i^+ \geq \lambda_i^* \). Now the duration of stage \( i \) cannot exceed \( \lambda_i \varepsilon_i^+ \) (k \leq i \leq m), so \( \lambda_i \geq \varepsilon_i^+ \) and thus in any case \( \lambda_i \geq \lambda_i^* \) (k \leq i \leq m) and from the hypotheses, \( \lambda_i \geq \gamma_i \) (k \leq i \leq m).

Lemma 4.59

Suppose that priority by recall with inhibition is provided for buses arriving at a mean rate of \( \beta/e^* \) and that \( \lambda_i^* - \gamma_i \geq 0 \) (1 \leq i \leq k). For each \( i \) (1 \leq i \leq k), if
\[ (\lambda_i^* + \sum_{j=1}^{k} \lambda_{j0} + \lambda_{k+1,1} \lambda_i^* \frac{2 \lambda_0^*}{\lambda_0^*} - \gamma_i < 0 \]  
(4.157)
then there is a finite maximal value \( \beta_i \) of \( \beta \) for which the solutions of (4.148) satisfy \( \lambda_i - \gamma_i \geq 0 \).

Proof

Suppose that for some \( i \) (1 \leq i \leq k), (4.157) is satisfied. From the result of lemma 4.57,
\[
\lim_{\beta \to 0} \lambda_i - \gamma_i = (\lambda_i^* + \frac{k}{\lambda_0} \sum_{j=1}^{\infty} \lambda_j + \gamma_{k+1,1}) \frac{\lambda_i^*}{2\lambda_0^*} - \gamma_i < 0 \tag{4.158}
\]

Since
\[
\lim_{\beta \to 0} \lambda_i - \gamma_i = \lambda_i^* - \gamma_i > 0 \tag{4.159}
\]
and \(\lambda_i\) varies continuously with \(\beta\), there is a finite maximal value \(\beta_i\) of \(\beta\) such that \(\lambda_i - \gamma_i > 0\) for all \(\beta\) in \([0,\beta_i]\).

The last results of this section provide tests to detect whether or not the range of mean bus arrival rates for which acceptable solutions to (4.148) exist is restricted by a specified maximum cycle duration. When the inhibition rule is used, every cycle during which inhibition is in effect will comprise uninterrupted stages. By contrast, when the inhibition rule is not used, this will only occur if no bus arrives at the detector between the times at which stages \(k+1\) and 1 are called.

Lemma 4.60

Suppose that priority by hurry-call with inhibition is provided for buses arriving at a mean rate of \(\beta/c^*\) and that \(c > c^*\). If

\[
\frac{c_t}{c^*} - [1 + \sum_{i=k+1}^{M} (\lambda_i^* - \gamma_i)] < 0 \tag{4.160}
\]

then there is a finite maximal value \(\beta_t\) such that the solution of (4.148) satisfy the cycle-time constraint

\[
\frac{c_t}{c^*} - \sum_{i=1}^{M} (\lambda_i + \gamma_{i,0}) \geq 0 \tag{4.161}
\]

for all \(\beta\) in \([0,\beta_t]\).
Proof

From the result of lemma 4.56
\[
\lim_{\beta \to \infty} \frac{c_t}{c^*} = \sum_{i=1}^{m} \left( \lambda_i + \lambda_{i0} \right) = \frac{c_t}{c^*} - [1 + \sum_{i=k+1}^{m} (\lambda^* - \gamma_i)] < 0 \quad \text{(by hypothesis)} \quad (4.162)
\]

Thus (4.161) is violated in the limit as the mean bus arrival rate increases without bound. Since \( c_t > c^* \) and the signal settings \( \lambda_i = \lambda^*_i \) \((1 \leq i \leq m)\) satisfy (4.148) when \( \beta = 0 \), the constraint (4.161) is satisfied in that limit. From the continuity of the solutions to (4.148), \( \beta_t \), the maximal value for which (4.161) is satisfied for each \( \beta \) in \([0, \beta_t]\) is well defined and finite.

Lemma 4.61

Suppose that priority is provided by recall with inhibition for buses arriving at mean rate of \( \beta / c^* \) and that \( c_t > c^* \). If
\[
\frac{c_t}{c^*} - \frac{1}{\lambda^*_0} (\lambda^*_0 + \sum_{j=1}^{k} \lambda^*_j + \lambda^*_{k+1}) (\sum_{i=1}^{k} \lambda^*_i + \sum_{i=k+1}^{m} \lambda^*_i) - \gamma_{k+1} - \lambda^*_0 < 0 \quad (4.163)
\]
then there is a finite maximal value \( \beta_t \) such that the solutions of (4.148) satisfy the cycle-time constraint (4.161) for each \( \beta \) in \([0, \beta_t]\).

Proof

From the result of lemma 4.57,
\[
\lim_{\beta \to \infty} \frac{c_t}{c^*} = \sum_{i=1}^{m} (\lambda_i + \lambda_{i0}) = \frac{c_t}{c^*} - \frac{1}{\lambda^*_0} (\lambda^*_0 + \sum_{j=1}^{k} \lambda^*_j + \lambda^*_{k+1}) (\sum_{i=1}^{k} \lambda^*_i + \sum_{i=k+1}^{m} \lambda^*_i) - \gamma_{k+1} - \lambda^*_{i=1} \lambda_{i0} < 0 \quad \text{(by hypothesis)} \quad (4.164)
\]

Thus (4.161) is violated in the limit as the mean bus arrival rate increases without bound. The remainder of the proof is as for lemma 4.60.
CHAPTER 5
DELAY

5.1 Introduction

Estimates of the mean delay incurred by vehicles at signal-controlled junctions can provide useful indications of the performance of the control policy implemented there. These estimates can be used to investigate the likely effects of proposed modifications such as the introduction of priority for buses or changes in the signal-settings used. This and other performance indicators, such as higher moments of delay and the mean number of stops incurred by vehicles, were discussed in Section 2.2 as possible components of evaluation criteria.

In this chapter, the problem of estimating mean delay is considered. When responsive priority for buses is implemented, the durations of the effective red and green periods experienced by the streams of traffic are random variables. This violates the requirement of usual delay formulae that these periods be of fixed duration. Similar restrictions apply to formulae developed to estimate other performance indicators, including the mean number of stops incurred by vehicles.

Two distinct categories of vehicles are identified here and analysed separately. The first category is that of non-priority vehicles: in this case the state of the traffic signals and the arrivals of vehicles are mutually independent. A formula is derived to estimate the mean delay incurred by a stream of traffic comprising vehicles in this category. The method used in the derivation of this formula is quite versatile and is used to derive a related formula to estimate the mean number of times each vehicle has to stop in the process of passing through the junction. A
formula is derived to estimate the mean overflow, a quantity which is required for use in each of the formulae described above.

All of these formulae reduce to forms which are similar to established ones when applied to fixed-time traffic signals with the normal rules of operation. In general, the use of these formulae requires a considerable amount of information concerning the durations of the effective red and green periods. Explicit formulae for the quantities required are given for the two combinations of priority rules which are used most commonly in practice. These are priority by extension and recall and priority by extension and recall with inhibition. Other combinations of priority rules could be analysed in a similar manner.

The second category of vehicles is those for which priority is provided. Responsive priority methods operate by altering the way in which the traffic signals change in order to allow these vehicles to pass with reduced delay. Thus the state of traffic signals is not independent of the arrivals of vehicles in this category. An approximate method is used to derive formulae which are appropriate to estimate the mean delay incurred by these vehicles both in the absence of priority and when priority is provided by each of the two combinations of priority rules mentioned above.

The analyses presented in this chapter are subject to more restrictions than were those in the previous chapter. Here the traffic signals are supposed to operate according to some fixed-time plan which is disrupted only when priority is granted to priority vehicles. Further assumptions are made concerning the short-term variability of the arrivals of non-priority vehicles when analysing the effects of the priority methods on them.
5.2 A delay formula for non-priority vehicles

5.2.1 Introduction and preliminaries.

A considerable literature has grown up on methods to estimate the mean delay incurred by vehicles at traffic signals. However, most of these methods are appropriate only when the durations of the effective red and green intervals are fixed. This requirement confines their application to fixed-time traffic signals. In particular, it makes them unsuitable for use at junctions where responsive priority is provided for buses.

In this section, various methods which have been used to estimate delay are reviewed. One of the more versatile methods of analysis is then used to derive a formula which is appropriate to estimate the mean delay incurred by vehicles when the durations of the effective red and green periods are random variables. A similar method is used to estimate the mean number of stops incurred by vehicles under these circumstances.

An approximate formula is derived to estimate the mean overflow and methods are given to estimate the covariances of the overflow with the durations of the subsequent effective red and green periods. All of these formulae are appropriate for any case where the durations of effective red and green periods are not correlated with the arrivals of the vehicles which are controlled by them. These formulae can be used together to estimate the mean delay incurred by vehicles in a stream of traffic from statistics of the durations of the effective red and green periods.

This is a convenient point at which to introduce the model of vehicular arrivals to be adopted in the derivation of the formulae described above. Suppose that batches of vehicles arrive according to the Poisson law. Suppose further that the number of vehicles in these batches are independent and identically distributed over the strictly positive
integers. The vehicular arrivals are then said to constitute a compound Poisson process.

The principal properties of this arrival model are stated in

Lemma 5.1
Suppose that vehicular arrivals at a point constitute compound Poisson process for which the mean rate of arrival of batches is $\lambda$ and the probability that there are $n$ vehicles in a batch is $a_n \ (n \geq 1)$. Let $\phi(z)$ be the probability generating function of $n$, defined by

$$\phi(z) = \sum_{n=1}^{\infty} a_n z^n \quad (5.1)$$

Then the probability generating function $\phi(z,t)$ of $\tilde{A}(t)$, the number of vehicles arriving in an interval of duration $t$, is given by

$$\phi(z,t) = e^{\lambda t[\phi(z)-1]} \quad (5.2)$$

Furthermore, the mean and variance of $\tilde{A}(t)$ are given by

$$\begin{align*}
E(A|t) &= \lambda t \phi'(1) \\
Var(A|t) &= \lambda t \phi'(1) + \lambda t \phi''(1) \\
\end{align*} \quad (5.3)$$

Proof

The form (5.2) for the probability generating function of $\tilde{A}(t)$ is given by Feller (1957, p 270). Differentiating (5.2) with respect to $z$ gives

$$\phi'(z,t) = \lambda t \phi'(z) e^{\lambda t[\phi(z)-1]}$$

and

$$\phi''(z,t) = \left( \lambda t \phi'(z) \right)^2 + \lambda t \phi''(z) e^{\lambda t[\phi(z)-1]}$$

Using the standard results (Feller, 1957, p 250)

$$\begin{align*}
E(A|t) &= \phi'(1,t) \\
Var(A|t) &= \phi''(1,t) + \phi'(1,t) - [\phi'(1,t)]^2 \\
\end{align*}$$

and that $\phi(1)=1$ gives (5.3).
The compound Poisson arrival process has the important property that the numbers of vehicles arriving during arbitrary disjoint intervals are mutually independent. Indeed, it is the most general arrival pattern which exhibits this property (Feller, 1957, p 270). The next result illustrates the facility of this process to represent a wide range of arrival patterns. Recall that the index of dispersion, $I_a$, is defined to be equal to the ratio $\text{Var}[\tilde{A}(T)]/E[\tilde{A}(T)]$ for some sampling interval of duration $T$.

**Corollary 5.2**

For any $q^* \geq 0$ and $I^* \geq 1$, there is a compound Poisson arrival process with mean rate $q^*$ and index of dispersion $I^*$.

**Proof**

From (5.3), the $I_a$-ratio of a compound Poisson arrival process is given by

$$ I_a = 1 + \frac{\phi''(1)}{\phi'(1)} $$

which is independent of the duration of the sampling interval used.

Let $n^*$ be the least integer such that $n^* > I^*$ and set

$$ a_n = \begin{cases} \frac{(n^*-1)(n^*+1-I^*)-(I^*-1)}{(n^*-1)(n^*+1-I^*)} & (n=1) \\ \frac{(I^*-1)}{(n^*-1)(n^*+1-I^*)} & (n=n^*) \\ 0 & \text{otherwise} \end{cases} $$

Then $\Phi_a = a_1 + \ldots + a_{n^*}$

$$ = 1 $$

Furthermore

$$ a_{n^*} = \frac{(I^*-1)}{(n^*-1)(n^*+1-I^*)} $$

$$ < \frac{1}{(n^*+1-I^*)} $$

since $(I^*-1) < (n^*-1)$

$$ < 1 $$

since $n^*-I^* > 0$
and \( a_{n^*} = \frac{(I^*-1)}{(n^*-1)(n^*+1-I^*)} \)

\[ > 0 \]

since \( I^*-1 > 0 \), \( n^*-1 > 0 \), \( I^* > 1 \). Finally, \( 0 < a_i < 1 \) because \( a_i + a_n \neq 1 \). Thus (5.5) defines a probability distribution.

Using (5.5) as the probability distribution of the batch sizes in a compound Poisson process, (5.1) and (5.4) give

\[ I_a = 1 + \frac{n^*(n^*-1)a_{n^*}}{a_1 + n^*a_{n^*}} = I^* \]

Now set \( \lambda = q^* \frac{(n^*+1-I^*)}{n^*} \). Since \( n^* > I^* \) and \( q^* > 0 \), \( \lambda > 0 \). From (5.1) and (5.3), the mean arrival rate is given by

\[ q = \lambda (a_1 + n^*a_{n^*}) = q^* \quad \text{(from (5.5))} \]

Note that if \( a_n = \delta \ln \), then \( \phi(z) \equiv z \) and the simple Poisson arrival process is recovered. The construction given in the proof of corollary 5.2 sets \( a_n = \delta \ln \) wherever \( I^* = 1 \).

The departure model used in Chapter 4 was subject only to weak hypotheses. For the estimation of delay, the times taken for different vehicles to depart while the queue persists are taken to be independent and identically distributed random variables. The variability of these departure times can arise in two distinct ways. A stream may contain several classes of vehicles, each of which has a different mean speed value per vehicle. Furthermore, some variability may exist within each class of vehicle. Suppose that a stream comprises \( n \) classes of vehicles. Let

\[ \kappa/s \] be the mean departure time for vehicles in that stream

\[ \sigma^2/s^2 \] be the variance of the departure times of vehicles in that stream

\[ \kappa_i/s \] be the mean departure time for vehicles in class \( i \) \((1 \leq i \leq n)\)
\( \sigma_i^2/s^2 \) be the variance of the departure times of vehicles in class \( i \) \((1 \leq i \leq n)\)

\( \pi_i \) be the proportion of vehicles in the stream which are in class \( i \) \((1 \leq i \leq n)\).

The quantities \( \kappa \) and \( \sigma \) can be found from \( \kappa_i, \sigma_i \) and \( \pi_i \) \((1 \leq i \leq n)\) from the result of

**Lemma 5.3**

The mean and variance of the departure times of vehicles in a stream of traffic are related to those of vehicles in the individual classes by the formulae

\[
\frac{\kappa}{s} = \sum_{i=1}^{n} \pi_i \frac{\kappa_i}{s} \\
\frac{\sigma^2}{s^2} = \sum_{i=1}^{n} \pi_i \frac{\sigma_i^2}{s^2} + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \pi_i \pi_j (\kappa_i - \kappa_j)^2/s^2
\]

(5.6)

**Proof**

Let \( S \) be the time taken for a single vehicle to depart. Then by definition, the conditional expected value of \( S \) given the class of the vehicle is

\[
E(S|i) = \kappa_i/s
\]

so \( E(S) = \sum_{i=1}^{n} \pi_i \kappa_i/s \).

Similarly,

\[
E(S^2) = \sum_{i=1}^{n} \pi_i (\kappa_i^2 + \sigma_i^2)/s^2
\]

so \( \text{Var}(S) = \sum_{i=1}^{n} \pi_i (\kappa_i^2 + \sigma_i^2)/s^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i \pi_j \kappa_i \kappa_j/s^2 \)

\[
= \sum_{i=1}^{n} \pi_i \sigma_i^2/s^2 + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \pi_i \pi_j (\kappa_i - \kappa_j)^2/s^2
\]

\[
= \sum_{i=1}^{n} \pi_i \sigma_i^2/s^2 + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \pi_i \pi_j (\kappa_i - \kappa_j)^2/s^2
\]
The next result gives a lower bound for $\sigma^2/s^2$ which can be used if $\sigma_i^2/s^2$ (1 ≤ i ≤ n) are not all known. Indeed, the lower bound gives the exact value when each of these quantities is zero, indicating that all the vehicles in each class require the same time to depart.

**Corollary 5.4**

The variance of the departure times of vehicles in a stream of traffic satisfies the lower bound

$$\sigma^2/s^2 \geq \sum_{i=2}^{n} \sum_{j=1}^{i-1} \pi_i \pi_j (\kappa_i - \kappa_j)^2 / s^2$$

(5.7)

where equality holds if and only if $\sigma_i = 0$ (1 ≤ i ≤ n).

This departure model is used here in two complementary ways. The first is to find statistics of the length of time taken for a given number of vehicles to depart from the queue and the second is to estimate statistics of the number of vehicles which depart in a period of given duration. These statistics, which are given in the next result, will be used throughout this chapter. Let

$\tilde{S}(n)$ be the time taken for $n$ vehicles to depart from the queue during an effective green period

$\tilde{N}(g)$ be the number of vehicles to depart in an effective green period of duration $g$

$C_s$ be the coefficient of variation of the departure times from the queue, defined by $C_s = \sigma_\kappa$.

**Lemma 5.5**

Suppose that the times taken for vehicles to depart from the queue are independent and identically distributed random variables with mean $\kappa/s$ and variance $\sigma^2/s^2$. Then the mean and variance of the time taken for $n$
vehicles to depart from the queue are given by

\[
\begin{align*}
E[S(n)] &= \frac{nk}{s} \\
Var[S(n)] &= \frac{ng^2}{s^2} = \frac{nk^2}{s^2} C_s^2 \\
\end{align*}
\]  
(n≥1)  \tag{5.8}

The mean and variance of the number of vehicles to depart during an effective green period throughout which the queue persists are given approximately by

\[
\begin{align*}
E[\bar{N}(g)] &= (\frac{s}{k}) E(g) \\
Var[\bar{N}(g)] &= (\frac{s}{k})^2 \text{Var}(g) + (\frac{s}{k}) C_s^2 E(g) \\
\end{align*}
\]  \tag{5.9}

Proof

The first result, (5.8), follows from the hypothesis that the times taken by different vehicles to depart are independent and identically distributed.

Now consider the conditional moments of the number of vehicles to depart during a single effective green period throughout which the queue persists given that the duration of the effective green period is \( g \). From Cox (1962, p 40)

\[
\begin{align*}
E[\bar{N}(g)|g] &= (\frac{s}{k}) g \\
E[\bar{N}(g)^2|g] &= (\frac{s}{k})^2 g^2 + (\frac{s}{k}) C_s^2 g \\
\end{align*}
\]  \tag{5.10}

where these approximations are asymptotically correct as \( g \to \infty \). Taking expectations over \( g \) in (5.10) leads to (5.9). [ ]

An immediate consequence of (5.8) is that

\[
\frac{Var[S(n)]}{[E[S(n)]]^2} = \frac{C_s^2}{n} \tag{5.11}
\]

The approximations (5.9) are used only in parts of the analysis which are themselves approximate so the use of exact results here would not in itself lead to an exact analysis. As a result of the model of vehicular
departures at the end of an effective green period, the error in the estimate of $\mathbb{E}[\hat{N}(g)]$ is independent of the distribution of $g$ and is of order $\frac{C^2}{s}$.

The following additional notation will be used in this chapter to describe the queueing behaviour of a single stream of traffic. Let

$\mathbf{t}$ be the time since the end of the last effective green period (s)
$\mathbf{d}$ be the mean delay incurred by a vehicle (s)
$\mathbf{D}_r$ be the total delay incurred by vehicles during an effective red period (vehicle seconds)
$\mathbf{D}_g$ be the total delay incurred by vehicles during an effective green period (vehicle seconds)
$\mathbf{D}^\#_g$ be the total delay incurred by vehicles during an effective green period of unlimited duration (vehicle seconds)
$\mathbf{D}$ be the total delay incurred by vehicles during an effective red period and the subsequent effective green period (vehicle seconds)
$\mathbf{Q}(t)$ be the number of vehicles in the queue, excluding any whose departure time is currently elapsing, at time $t$ (vehicles)
$\mathbf{R}_n$ be the duration of the effective red period numbered $n$, counting from some suitable origin (s)
$\mathbf{G}_n$ be the duration of the effective green period which follows on the effective red period number $n$ (s)
$\mathbf{Q}_n$ be the overflow at the start of the effective red period numbered $n$ (vehicles).

5.2.2 Review.

A considerable number of formulae have been devised to estimate the mean delay incurred by vehicles passing through a signal-controlled
junction. Most of the differences between these formulae are due to the different models used to represent the behaviour of traffic. An extensive review of this topic is given by McNeil and Weiss (1974) and detailed reviews of estimates appropriate to traffic in streams which are controlled by fixed-time signals are given by Allsop (1970, 1972a). The review presented in this sub-section considers material which is relevant to this thesis.

Many models have been used to represent the arrival of vehicles at a junction. Perhaps the simplest and least realistic of these is that of arrivals with constant headways, known as the regular arrivals model. Many observations have been published which indicate that this model does not provide a particularly good representation of road traffic. Adams (1936), for example, considered the distribution of the number of vehicles in a stream of freely flowing light traffic which passed an isolated point during 10-second intervals. He found that this could be approximated closely by the Poisson counting distribution (4.31). Other more recent observations such as those of Williams and Emmerson (1961) and Miller (1964) have shown that the $I_a$ values of traffic counts taken over periods of a few minutes duration normally lie in the interval $[1,2]$, a typical value being 1.5. This contrasts with the regular arrivals model which necessarily gives rise to $I_a$ values close to 0.

The binomial process has been used to represent arrivals which are somewhat irregular. This gives rise to $I_a$ values which are less than 1 but which can be varied within the interval $(0,1)$ by suitable choice of the time interval used to generate the vehicles. A common choice for this time interval is the departure time of a single vehicle, giving rise to the discrete-time model. This model is particularly convenient to analyse because during effective green periods, the number of vehicles in the queue either remains unchanged or is decreased by 1 in each departure period.
However, the model is awkward to use unless the durations of the effective red and green periods are integral multiples of the departure time of a single vehicle. This departure time is itself taken to be constant so \( C = 0 \). Finally, the \( I_a \) value measured over any whole number of departure intervals is determined by the arrival and departure rates as \( 1 - \frac{Kq}{s} \).

The simple Poisson model of arrivals has been used by many authors because of the convenient property that the numbers of vehicles to arrive in arbitrary disjoint intervals are independent. This model gives an \( I_a \) value of 1 irrespective of the sampling interval. The compound Poisson model is similarly convenient and provides \( I_a \) values which are also independent of the sampling interval and which can be given any value greater than or equal to 1. More sophisticated models have been developed by fitting probability density functions to observed headway data (see, for example, Ovuworie, Darzentas and McDowell, 1980). However, these have not lead to any tractable analyses of delay.

In this review, all estimates are applicable to streams of traffic which experience alternating effective red and green periods which are of constant duration. Thus these estimates apply mainly to fixed-time traffic signals at which each stream receives a single period of right of way in each cycle. In the interests of clarity, subscripts are omitted from quantities such as \( r_n \) and \( g_n \) in these cases.

The first estimates of delay to be published were derived using the model of regular arrivals and constant departure times. Clayton (1940) and Wardrop (1952) gave similar approximate formulae for this case. Allsop (1970, 1972a) gave the exact formula

\[
d = \frac{s}{2(s-Kq)c}(l_x-K_s^2) + \frac{Kq(2s-Kq)+\kappa\theta(s-Kq)}{12q^2s^2}
\]

for some \( \theta \) in the interval \( \left( -\frac{1}{3\sqrt{3}}, \frac{2}{3\sqrt{3}} \right) \).
Wardrop argued that any such formula would give a lower bound for the delay incurred in practice since any irregularity in vehicular arrivals will tend to cause an overflow at the end of the effective green periods. Any vehicles which are held over in this way will be delayed for an extra effective red period, thus increasing the mean delay. Two distinct methods which have been used to allow for the irregularity of vehicular arrivals at a junction are discussed in turn.

Allsop (1972a) proposed the following model for the queueing process at traffic signals in order to derive a delay formula after Webster (1958). Suppose that a notional queue is interposed between the source of vehicles and the stop-line. Suppose further that the service time for each vehicle at the notional queue is equal to the long-term mean service time of vehicles at the stop-line. The total delay in this composite queueing system can be found by adding the delays incurred at each of the two queues. The Pollaczek-Khintchine formula (Kendall, 1951) can be applied to give the mean delay incurred at the interposed queue by vehicles which arrive there according to the Poisson law. Since the mean service time at this queue is \( \kappa/s \lambda \), the mean delay is given by

\[
d_i = \frac{x}{(1-x)} \frac{\kappa}{2s \lambda} = \frac{x^2}{2q(1-x)}
\]

(5.13)

When the input to this interposed queue is Poisson, the output consists of groups of vehicles separated by single service times alternating with gaps of duration exceeding \( \kappa/s \lambda \) by an amount which has the negative exponential distribution with mean \( 1/q \). The number of vehicles in the groups has the Borel-Tanner distribution (Tanner, 1961). The mean delay incurred by this traffic at the stop-line is given approximately by the formula
\[ d_t \approx \frac{s r^2}{2(s-kq)c} \quad (5.14) \]

which is a simplified form of (5.12). Thus the mean delay incurred at the composite queue is given approximately by

\[ d = \frac{s r^2}{2(s-kq)c} + \frac{x^2}{2q(1-x)} \quad (5.15) \]

In order to investigate the verity of this composite representation of a traffic-signal queue, Webster (1958) compared the estimates of delay given by (5.15) with those derived from digital simulation. The simulation, which used a binomial arrival pattern, indicated that the expression (5.15) over-estimates delay. An empirical correction term was fitted to the results of the simulation to give the estimate

\[ d = \frac{s r^2}{2(s-kq)c} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{k^2 q^2}\right)^{1/3} x^{(2+5\alpha)} \quad (5.16) \]

which is known as Webster's 3-term formula. The value taken by the third term was found to correspond to between 5 and 15 per cent of \( d \), so Webster suggested that the approximation

\[ d \approx \frac{9}{10} \left( \frac{s r^2}{2(s-kq)c} + \frac{x^2}{2q(1-x)} \right) \quad (5.17) \]

be used in practice. This is known as Webster's 2-term (or simplified) formula.

Because of the algebraic simplicity and computational convenience of (5.17), this formula has seen much practical use and has been the subject of several further investigations. Hutchinson (1972) introduced a multiplicative factor of \( \frac{1}{a} \) to the second term of (5.17) to allow for arrival processes other than the simple Poisson one. While no theoretical justification was offered for this modification, it yields a formula which agrees with (5.17) when the simple Poisson arrival process is used and is similar to (5.12) when the regular arrival process is used. However, the exact formula corresponding to (5.13) for the case of compound Poisson...
arrivals can be shown (Heydecker, 1982) to be \( r \frac{(T_a-1)+x}{2q(1-x)} \) rather than Hutchinson's suggested \( \frac{I_a x^2}{2q(1-x)} \).

Murchland (1977) and Gallivan (1982) have derived similar formulae which are appropriate for streams which receive more than one effective green period in each cycle. Allsop (1977a) gave corresponding formulae for a stream which receives right of way with different saturation departure rates during various stages in the sequence. All of these formulae resemble Webster's 2-term formula (5.17) in as much as they comprise the sum of an expression to approximate the delay incurred at the traffic signal by vehicles with a regular arrival pattern and the Pollaczek-Khintchine formula (5.13). Robertson (1969) made similar use of the Pollaczek-Khintchine formula to add an extra component of delay to that derived from the simulation of a simplified traffic model.

An alternative method to allow for the irregularity of vehicular arrivals was introduced by Wisten (Beckmann, McGuire and Wisten, 1956). This method, which has been adopted by many authors since then, leads to an expression for the mean delay which involves the mean overflow. The method is described in detail in the next sub-section where it is used to derive a formula to estimate delay when the durations of the effective red and green periods are random variables.

The essence of this method is to derive an expression for the expected value of the total delay incurred by vehicles during an effective red period and the next effective green period. The mean delay is then calculated from this estimate and the mean number of arrivals during the same period using the formula

\[
d = \frac{E(D)}{E[L_n(q_n + q_n)]}
\]

(5.18)

Winston, investigating the discrete time model, used this method to derive
the expression
\[
d = \frac{sr}{2(s-kq)c} \left[ (r + \kappa/s) + \frac{2E(Q)}{q} \right]
\] (5.19)

Miller (1963) used an approximate analysis founded on Winsten's method to estimate the mean delay incurred by vehicles with a more general arrival pattern but with constant departure times. Two apparently weak conditions on the arrivals were used: these were that the numbers of vehicles arriving during consecutive pairs of effective red and green periods are mutually independent and that the arrivals of vehicles are not correlated with the state of the traffic signals. Miller's (1963) estimate of the mean delay is
\[
d = \frac{sr}{2(s-kq)c} \left[ (r + \frac{2E(Q)}{q} + \frac{K}{s})((I_a+1) + \frac{Kg}{s}) \right]
\] (5.20)

McNeil (1968) investigated the model of compound Poisson arrivals with independent and identically distributed departure times. He used a definition of delay which includes the service time of a vehicle if it has already incurred some delay but not otherwise. As with the model adopted here, McNeil supposed that once the queue has dissipated, it will not form again before the end of the effective green period. The mean delay in this case is given by
\[
d = \frac{sr}{2(s-kq)c} \left[ (r + \frac{2E(Q)}{q} + \frac{K}{s-kq})((I_a+1) + \frac{Kg}{s})((c^2-1)) \right]
\] (5.21)

Jacobson and Sheffi (1981) used (5.21) as the basis for an estimate of the mean delay incurred by non-priority vehicles when responsive priority is provided for buses. They used numerical methods to take expectations of the value of (5.21) over the distributions of the effective red and green times. They interpreted the result as the mean delay incurred by vehicles at a junction where the durations of the effective red and green periods are random variables with the distributions used. However, the process used is to find the mean delay under fixed-time operation and then to take expectations with respect to the duration of the effective red and green periods. Thus the resulting estimate of delay is appropriate to fixed-time
traffic signals where there is some uncertainty as to the durations of the effective red and green periods. Jacobson and Sheffi's estimate of delay can be written formally as

\[ d = \mathbb{E}\left\{ \frac{\mathbb{E}[D|x_n, r_n^2, q_n, Q_n]}{\mathbb{E}[x(x+g_n)|r_n^2, q_n]} \right\} \] (5.22)

In general, (5.22) will differ from (5.18) since the operations of expectation and division are performed in a different order in each case. However, in the special case where \( r_n + g_n = c \) where \( c \) is a constant, the denominator of (5.22) is a constant and can be removed from the outer expectation as a factor. This yields an expression which is identical to (5.18).

Before a delay formula of the form of (5.19) and (5.21) can be used, the mean overflow \( \mathbb{E}(Q) \) must be found. Winsten (Beckmann, McGuire and Winsten, 1956) recommended the use of an iterative numerical method to find this quantity. Since then, various authors have investigated this problem with the use of several different techniques.

In the case of the discrete time model, the particularly simple relation

\[ Q_{n+1} = \text{Max}[0, Q_n + \tilde{A}(r_n + g_n) - \tilde{N}(g_n)] \] (5.23)

holds exactly. Newell (1960) used methods similar to those developed by Bailey (1954) to find a formula for the probability generating function of the overflow in terms of the complex roots of a function. This formula can be developed following the methods of Boudreau, Griffin and Kac (1962) to give the expression

\[ \mathbb{E}(Q) = \sum_{j=1}^{G-1} (1-z_j)^{-1} - \frac{1}{2} \left\{ \frac{q_s}{\kappa c} + q_c - ((I_\alpha - 1)x + 1)(1-x)^{-1} \right\} \] (5.24)

where \( G = \frac{q_s}{\kappa c} \) and \( \{z_j\} \) are the \( G-1 \) complex solutions of the equation

\[ 0 = z^G - \left[ 1 + \frac{K_G(z-1)}{s} \right]^{q_s/c} \] (5.25)

which have modulus less than 1.
Haight (1959) analysed the model of simple Poisson arrivals and constant departure times for vehicles which have been delayed. He supposed that once the queue has dissipated, no further vehicles are delayed. He gave explicit formulae for the probability distribution of $Q_n$ in a form which is conditional on $Q_{n-1}$, $r_{n-1}$ and $g_{n-1}$. Miller (1969) used this conditional distribution for transition probabilities to approximate iteratively the unconditional probability distribution of $Q_n$ when $r_n$ and $g_n$ have fixed values. This method was found to converge slowly so is not particularly useful in itself. However, the results of this procedure were used to suggest the empirical formulae

$$E(Q_n) = \frac{\text{Exp}[-1.33(\frac{1-x}{x})^{\frac{s_g}{k}}]}{2(1-x)} \quad (5.26)$$

and $P(Q_n=0) = 1 - \text{Exp}[-1.58(\frac{1-x}{x})^{\frac{s_g}{k}}] \quad (5.27)$

Darroch (1964) analysed a versatile extension of the discrete-time model. He allowed for the number of arrivals during the time periods to be independent and identically distributed random variables with probability generating function $\phi(z)$. This admits both the binomial and the compound Poisson arrival models as special cases. The departure model treated each time period during which the queue persists as a Bernoulli trial where success corresponds to the departure of one vehicle. This gives rise to independent departure times with identical geometric distributions. As in the model used by Haight (1959), once the queue has dissipated, no further vehicles are delayed. A formula was given for the probability generating function of the overflow which, in the case of constant departure times, yields the expression

$$E(Q_n) = (1 - \frac{K_g}{s})^G \sum_{j=1}^{G-1} \left[1 - \frac{z_j}{\phi(z_j)}\right]^{-1} + \frac{G-cq}{1-cq/s} \left. \frac{d}{dz} \left[ \frac{[\phi(z)]^{(G-1)}[z-\phi(z)]}{z^G [\phi(z)]^{cs/k}} \right] \right|_{z=1} \quad (5.28)$$
where \( G = \frac{gs}{\kappa} \) and \( z_j \) are the \( G-1 \) complex solutions of the equation

\[
0 = z^G - \phi(z, c)
\]  \hspace{1cm} (5.29)

which have modulus less than 1. Because of the hypothesised independence of arrivals during successive departure periods, \( \phi(z, c) = [\phi(z)]^{cs/\kappa} \).

Several authors have noted that the relation (5.23) is equivalent to the recurrence relation of a bulk-service queue with maximum service batch size \( gs/\kappa \). This many vehicles could always depart in an effective green period of duration \( g \) under the assumptions of the usual departure model irrespective of the times at which they arrive. Thus the bulk-service model can be used to provide an upper bound for the mean overflow. Indeed, (5.24) is appropriate for this with any arrival process for which the number of arrivals in successive cycles are independent and identically distributed random variables with probability generating function \( \phi(z, c) \). In general, \( \{z_j\} \) are the complex solutions of the Equation (5.29) which have modulus less than 1 and \( G \) is one greater than the number of solutions which have this property.

Miller (1963) proceeded directly from (5.23) using methods similar to those of Kendall (1951). He considered the model of constant departure times and required that the number of vehicles arriving during consecutive pairs of effective red and green periods be mutually independent. This analysis lead to the estimate

\[
E(Q_n) = \text{Max}[0, \frac{(2x-1)I_a}{2(1-x)}]
\]  \hspace{1cm} (5.30)

McNeil (1968) deduced from Miller's (1963) method a strict upper bound for \( E(Q_n) \) which is given by

\[
E(Q_n) < \frac{xI_a}{2(1-x)}
\]  \hspace{1cm} (5.31)

He noted that the derivations of (5.30) and (5.31) could be extended to the case of independent and identically distributed departure times.
McNeil (1968) investigated the model of compound Poisson arrivals with independent and identically distributed departure times. With the bulk-service approximation, this departure model is equivalent to supposing that the capacity of the service batches are independent random variables with identical probability generating functions, $\Psi(z)$, say. McNeil (1968) supposed that there was a finite maximum to the size of the service batches, $N$, say. He derived an expression for the probability generating function of the overflow and deduced an expression for the mean value which reduces to the appropriate form of (5.24) when the departure times are all identical. Following Crommelin (1932) he showed that this expression is equivalent to

$$E(Q_n) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{1}{\Psi(z)} \left[ z^{n} \right]_{z=0}$$

where $\Phi(z,c)$ is the probability generating function of the number of arrivals in a cycle. The series (5.32) was shown to converge faster as $N$ increases.

Griffiths (1981) considered the case of simple Poisson arrivals and constant departure times for vehicles at an uncontrolled pedestrian crossing. Using the additional assumption that the arrivals of pedestrians forms a Poisson process, he found explicit formulae for the probability density functions of the durations of the effective red and green periods. A general form was derived for the probability generating function of $\tilde{Q}(r)$, the number of vehicles in the queue at the start of an effective green period. The derivation used the mutual independence of the durations of all the effective red and green periods and the additional assumption that $N$ is finite.

Using the particular forms of the probability density functions of the durations of the effective red and green periods that arise in the
pedestrian crossing case, Griffiths found the limiting form of the probability generating function of \( \bar{Q}(r) \) as \( N \) increases without bound. He then deduced an expression for \( E[\bar{Q}(r)] \) in a form which requires only the solution of a single real equation.

The overflow is closely related to the quantity \( \bar{Q}(r) \) because the size of the queue increases according to a simple Poisson counting process during the effective red periods. In particular, the mean values satisfy the relationship \( E(Q_n) = E[\bar{Q}(r)] - qE(r) \).

Several microscopic simulation models have been developed to investigate the consequences of providing responsive priority for buses. These models have been used to provide estimates of the mean delay incurred by priority and non-priority vehicles and could, in principle, be used to give estimates of other related quantities.

Wood (1978) developed the ROSIM simulation model (Needham, 1970) to incorporate a representation of responsive priority for buses. The resulting model was used by Vincent, Cooper and Wood (1978) to investigate the effects of providing responsive priority by each of three combinations of rules at an isolated cross-roads junction controlled by 2-stage traffic signals. El-Reedy and Ashworth (1978) constructed a similar simulation model which used cyclic traffic arrival patterns derived from the TRANSYT simulation program (Robertson, 1969). This model was used to investigate the effects of providing responsive priority for buses at a single junction within a network of co-ordinated traffic signals. Similar use of a microscopic simulation model was made by Cottinet et al (1980) who investigated an isolated cross-roads, an arterial road with 5 cross-roads junctions and a network of 8 such junctions.

The mean number of times a vehicle has to stop in order to pass through a signal-controlled junction can be used to estimate wear and tear
caused to vehicles and annoyance caused to drivers. Furthermore, formulae such as (2.4) make use of this quantity in order to estimate the rate at which fuel is consumed at a junction (Robertson, Lucas and Baker, 1980). The number of times a vehicle has to stop can be estimated from models of traffic behaviour in a similar but rather more straightforward way to values of the mean delay.

Webster (1958) estimated the proportion of vehicles that have to stop at least once ($P_H$) and the mean number of stops incurred by a vehicle ($H$) from the formulae

\[
P_H \approx \frac{sr}{c(s-kq)}
\]

\[
H = \min \left[ \frac{s\tilde{Q}(r)}{(s-kq)qc}, \frac{\tilde{Q}(r)+gq}{qc} \right]
\]

where \( \tilde{Q}(r) = \max[q_r, q\left(\frac{r}{2} + a\right)] \)

The method used to derive the formula for $P_H$ ignored the possibility of a non-zero overflow and supposed that each vehicle which is delayed at all has to stop. The formula to estimate $H$ takes into account the overflow and thus allows for the possibility that some vehicles have to stop more than once.

Akçelik (1980) presented a discussion of these and other similar formulae. He derived the estimate

\[
H = \frac{sr}{c(s-kq)} + \frac{E(Q)}{qc}
\]

again using the assumption that every vehicle which is delayed actually stops. This assumption was then relaxed to allow for those vehicles which incur small delays by reducing speed without stopping and for those which stop more than once without accelerating to their free speed between stops. Both of these effects were considered to contribute only partial stops to the total number experienced. The amended estimate

\[
H = f\left[\frac{sr}{c(s-kq)} + \frac{E(Q)}{qc}\right]
\]
was given together with a procedure to find appropriate values for $f$. The value $f = 0.9$ was recommended as being suitable for practical use in all cases.

Various authors have made numerical comparisons of the available delay formulae. In particular, Hutchinson (1972) compared both of Webster's (1958) formulae with those of Miller (1963, 1968) and one due to Newell (1965). Although there were great differences between the asymptotic behaviour of the formulae, they all gave similar estimates of delay in normal circumstances. Hutchinson concluded that the main criterion for selecting any particular delay formula should be one of computational convenience. McNeil and Weiss (1974) drew similar conclusions from a more limited comparison between estimates of delay calculated from McNeil's formula (5.21) using each of the upper bounds (5.31) and (5.32) and Darroch's method (5.28) and (5.29) to estimate the mean overflow.

Ohno (1978) reviewed critically the methods used by Darroch (1964), Newell (1965) and McNeil (1968) to derive their estimates of delay. He showed that the differences between these formulae (for a given value of $E(Q)$) are due to differences between the supposed behaviour of the queue between successive departures. In any case, these differences are of the order of the service time of a single vehicle so might be considered to be of minor importance. Due to the inconvenience of Darroch's (1964) method (5.28) and (5.29) to estimate the mean overflow, Ohno recommended the use of Miller's (1969) heuristic formula (5.26) which was shown to give rise only to small errors in the resulting estimates of delay.

If the value associated with the time spent by vehicles at a junction is regarded as being proportional to its duration and no other variable costs are considered, then the operating cost of a junction is directly proportional to the total rate of delay incurred there. Several authors
have considered the problem of finding signal-settings which minimise the total rate of delay. Most of the methods which have been devised are manual rules which result from consideration of simple junctions and are generalisation of approximate analyses.

Webster (1958) found that if only one stream of traffic receives right of way during each stage, then the total rate of delay estimated from (5.17) is minimised by signal-settings which give rise to approximately equal degrees of saturation in the streams of traffic. When the signals are set so that the degrees of saturation in the streams are equal, the total rate of delay was found to be minimised at a cycle time close to that given by

\[
\lambda = \frac{1.5L + 5}{1 - \frac{\nu}{\sum_{j=1}^{m} (\kappa_j q_j / s_j)} (1 - L/c)}
\]  

(5.36)

where \( L \) is the sum of the effective durations of all the transition periods in the cycle.

These observations lead to the signal-setting procedure known as Webster's method. For each stage, the stream which has the highest ratio \( \kappa_j q_j / s_j \) and which has right of way during that stage is identified. These are known as the representative streams and, renumbering if necessary, each one is taken to have the same index as the stage which it represents. The cycle time is then calculated from (5.36) and the stage durations are calculated from the formula

\[
\lambda_i = \frac{K_i q_i}{s_i} \left( \sum_{j=1}^{m} \kappa_j q_j / s_j \right)^{-1} (1 - L/c) 
\]  

(1 \leq i \leq m)

(5.37)

This procedure has the effect of minimising the maximum degree of saturation at the junction subject to the fixed cycle time given by (5.36).
Miller (1963) gave detailed consideration to the problem of minimising the total rate of delay at a junction where two streams have right of way in each of two stages. Applying an analysis similar to that of Webster (1958) to his own delay formula, Miller (1963) arrived at approximate formulae for the required signal-settings. These were later simplified (Miller, 1968) to

$$c = \frac{L + 2.2 \sqrt{L/\min_{1 \leq j \leq M} (s_j/k_j)}}{1 - \sum_{j=1}^{M} (\kappa_j q_j/s_j)}$$  \hspace{1cm} (5.38)$$

with stage durations then calculated from (5.37).

Allsop (1970, 1971) considered the problem of minimising the total rate of delay at a junction, allowing for weighting factors \( w_j \) \((1 \leq j \leq M)\) for the rate of delay in respect of the relative importance of delay to traffic in each stream. The problem can be stated as

$$\text{Minimise} \sum_{i=1}^{M} q_i w_i d_i$$
\[ \lambda_i \text{ (0 \leq i \leq m)} \] \hspace{1cm} (5.39)

subject to the constraints of (4.27). Allsop (1970) showed that when Webster's simplified delay formula (5.17) is used, this problem is equivalent to one of convex minimisation and an appropriate algorithm was given to find the solution. Unlike the methods due to Webster (1958) and Miller (1963, 1968), this method takes into consideration the delay incurred by vehicles in all streams rather than just the representative ones.

Allsop (1971) showed that the signal-settings calculated according to (5.39) are usually similar to those given by (5.36) and (5.37), but generally give shorter cycle times. A particular advantage of Allsop's (1970) method is that it does not require the representative streams to be identified: this process can be difficult at complicated junctions, especially when some streams have right of way during more than one stage.
Once a method has been adopted to calculate signal-settings, the sensitivity of the results to changes in the input data can be investigated. When the signal-setting method is an explicit calculation, such as those due to Webster (1958) and Miller (1963, 1968), this sensitivity analysis is quite straightforward. However, when the signal-settings are calculated so as to minimise the total rate of weighted delay, more intricate analysis is required. Allsop (1972c) has given expressions for the sensitivity in the latter case to small changes in arrival and saturation departure rates and (Allsop, 1973) to changes in the utilisation of the transition periods.

5.2.3 Mean delay.

In this sub-section, a formula is derived to estimate the mean delay incurred by vehicles at traffic signals where the durations of the effective red and green periods are random variables which are not correlated with arrivals of the vehicles which they control. This formula is appropriate to estimate the mean delay incurred by the non-priority vehicles in each stream of traffic at signal-controlled junctions where responsive priority is provided for buses.

The derivation given here follows that of McNeil (1968) but requires rather weaker hypotheses and uses a slightly different definition of delay. In particular, here the durations of the effective red and green periods are treated as random variables and the times taken for vehicles to depart from the queue during effective green periods are treated as independent and identically distributed random variables. The definition of delay used here excludes the time taken by vehicles to depart on the grounds that this would form part of the travel time of each vehicle in any case.
Lemma 5.6

The mean value of \( D_r \), the total delay caused to vehicles during a single effective red period, is given by

\[
E(D_r) = \frac{q}{2}E(r^2) + E(Q_n)E(r) + \text{Cov}(Q_n, r_n) \tag{5.40}
\]

Proof

For a single effective red period of duration \( r_n \), \( D_r \) is given by

\[
D_r = \int_{t=0}^{r_n} \tilde{Q}(t)dt = \int_{t=0}^{r_n} \tilde{A}(t) + \tilde{Q}_n dt \tag{5.41}
\]

Taking expectations gives

\[
E(D_r) = E[E(D_r | Q_n, r_n)] = E[\int_{t=0}^{r_n} (qt+Q_n)dt] = E[\frac{q}{2}r_n^2 + Q_n r_n] = \frac{q}{2}E(r^2) + E(Q_n)E(r) + \text{Cov}(Q_n, r_n) \tag{5.40}\]

Lemma 5.7

The mean value of \( \tilde{D}^g(Q) \), the total delay incurred by vehicles during a single effective green period of unlimited duration which commences with an initial queue of \( Q \) vehicles, is given by

\[
E[\tilde{D}^g(Q)] = \frac{k}{2s} \left( \frac{s}{s-kq} \right)^2 \left\{ \left[ \left( \frac{kq}{s} \right) (c^2 + 1) - \left( \frac{s+kq}{s} \right) \right] Q + \left( \frac{s-kq}{s} \right) Q^2 \right\} \tag{5.42}
\]

Proof

Consider the sequences of random variables \( \{A_n\} \) and \( \{Z_n\} \) defined by

\[
A_n = \begin{cases} \Omega, & (n=0) \\ \tilde{A}[\tilde{S}(A_{n-1})] & (n \geq 1) \end{cases} \tag{5.43}
\]

and

\[
Z_n = \begin{cases} r, & (n=0) \\ Z_{n-1} + \tilde{S}(A_{n-1}) & (n \geq 1) \end{cases} \tag{5.44}
\]

where \( \tilde{S}(N) \) is the time taken for \( N \) vehicles to depart from the queue during an effective green period. Thus \( A_{n+1} \) represents the number of
vehicles which join the queue while \( A_n \) vehicles depart and \( Z_n \)
represents the time, measured from the start of the preceding effective red
period at which \( \sum_{i=0}^{n-1} A_i \) vehicles have departed.

The instants \( t=Z_n \) \((n \geq 0)\) provide convenient points at which to
divide the time axis. This gives rise to an expansion of \( \bar{D}(Q) \) as
follows:

\[
\bar{D}^* = \int_{t=0}^{\infty} \tilde{Q}(t) dt = \sum_{n=0}^{\infty} \int_{t=Z_n}^{Z_{n+1}} \tilde{Q}(t) dt
\]

The contribution to \( \bar{D}(Q) \) arising during the time interval
\([Z_n, Z_{n+1}) \) \((n \geq 0)\) can be found by considering separately the delay
accruing to the \( A_n \) vehicles which were in the queue at time \( t=Z_n \) and
that accruing to the \( A_{n+1} \) vehicles which arrive during that time
interval. Taking expectations of the delay arising during the time
interval \([Z_n, Z_{n+1}) \) \((n \geq 0)\)

\[
E[\int_{t=Z_n}^{Z_{n+1}} \tilde{Q}(t) dt] = E\left[ E\left[ h(A_n - 1) \tilde{S}(A_n) \right] + E\left[ \int_{t=Z_n}^{Z_{n+1}} \tilde{Q}(t) dt \right] \left| \tilde{S}(A_n) \right| A_n \right]
\]

\[
= E\left[ hA_n \tilde{S}(A_n) - h \tilde{S}(A_n) + \frac{q}{2} \tilde{S}(A_n)^2 \left| A_n \right. \right]
\]

\[
= \frac{K}{2s} E\left[ \frac{Kg}{s} A_n \left( s^2 - 1 \right) A_n + \frac{Kg}{s} A_n \left( s^2 - s + 1 \right) A_n \right] \quad (n \geq 0) \quad \text{from (5.8)}
\]

Now \( E[A_n | A_{n-1}] = \left( \frac{Kg}{s} \right) A_{n-1} \quad \text{(n \geq 1)} \) \( \quad (5.47) \)

and \( E[A_n^2 | A_{n-1}] = E[E[A_n^2 | \tilde{S}(A_{n-1})] | A_{n-1}] \)

\[
= E[q^2 \tilde{S}(A_{n-1})^2 + q \tilde{S}(A_{n-1}) | A_{n-1}] \]

\[
= \left( \frac{Kg}{s} \right)^2 A_{n-1}^2 + \left( \frac{Kg}{s} \right)^2 C_s^2 + \left( \frac{Kg}{s} \right) I_{a} A_{n-1} \quad (n \geq 1) \quad \text{from (5.8)} \]

Repeated substitution of (5.47) and (5.48) into (5.46) gives the geometric
series

\[
E[\int_{t=Z_n}^{Z_{n+1}} \tilde{Q}(t) dt] = \frac{K}{2s} E\left[ \left( \frac{Kg}{s} \right)^n \left[ 1 + \left( \frac{s+Kg}{s} \right)^n - 1 \right] \left[ C_s^2 + \sum_{i=0}^{n-1} \left( \frac{Kg}{s} \right) I_{a} \right] \tilde{Q} + \left( \frac{Kg}{s} \right)^n \left( \frac{s+Kg}{s} \right)^n \tilde{Q} \right]
\]
which can be summed to give
\[
E\left\{ \int_{t=n}^{n+1} Q(t) \, dt \right\} = \frac{K}{2s} \left\{ \left( \frac{Kq}{s} \right)^n \left[ \frac{Kq}{s} \left( 1 + \frac{s+Kq}{s-Kq} \right) \right] r^2 + \left( \frac{s+Kq}{s-Kq} \right)^n I_{a-1} \right\} Q + \left( \frac{Kq}{s} \left( \frac{s+Kq}{s} \right)^2 \right)^n Q^3. \tag{5.49}
\]

Taking expectations in (5.45) and using (5.49) gives another geometric series. Summing this over \( n \) and simplifying gives (5.42). [ ]

The problem investigated in lemma 5.7 is that of ascertaining the mean delay caused in a queue during a busy period which commences with an initial queue of \( Q \) customers. Kendall (1951) referred to the method used in the proof given here as the branching process view of a queue. An alternative method of proof has been used by Daley and Jacobs (1969).

The results of lemmas 5.6 and 5.7 can now be combined to give the main result of this sub-section.

**Theorem 5.8**
The mean delay incurred by vehicles in a stream of traffic is given by
\[
d = \left( \frac{s}{s-Kq} \right) \frac{\left\{ E(r^2) + \frac{2}{q} \left[ \text{Cov}(Q_n, r_n) + E(Q_n) E(r) \right] + \left( \frac{Kq}{s-Kq} \right) \left[ (I_{a-1}) + \frac{Kq}{s} (C^2-1) \right] E(r) \right\}}{2[E(r) + E(q)]} \tag{5.50}
\]

**Proof**
The expected total delay \( E[\tilde{D}_{\tilde{Q}}(Q)] \) exceeds the expected total delay \( E[\tilde{D}_Q(Q)] \) by an amount corresponding to the delay which occurs after the end of the effective green period. Thus
\[
E[\tilde{D}_{\tilde{Q}}(r_n)] = E[\tilde{D}_{\tilde{Q}}(r_n)] - \tilde{D}_{\tilde{Q}}(Q_n + g_n)] \tag{5.51}
\]

Now \( E[\tilde{Q}(r_n)] = E(Q_n) + qE(r_n) \) \( (5.52) \)

and \( E[\tilde{Q}(r_n)^2] = E[\tilde{Q}(r_n) + \tilde{A}(r_n)^2]|Q_n, r_n] \)

\[
= E[Q_n^2 + 2qQ_n r_n + q^2 r_n^2 + I_{a} q r_n^2] + q^2 E(r^2) + I_{a} q E(r) \tag{5.53}
\]
If there is a stationary distribution for the queue size at the end of the effective green periods, then the moments of $Q_n$ are independent of $n$, so

$$
\begin{align*}
\mathbb{E}[\bar{Q}(r_n+g_n)] &= \mathbb{E}[Q_{n+1}] \\
&= \mathbb{E}[Q_n] \\
\end{align*}
$$

and

$$
\mathbb{E}[[\bar{Q}(r_n+g_n)]^2] = \mathbb{E}[\bar{Q}_n^2]
$$

Using (5.42) and (5.52-4) in (5.51) gives

$$
\mathbb{E}(D_n) = q \left( \frac{s}{2} \right) \left( \begin{array}{c}
\left( \frac{Kq}{s} \right) \mathbb{E}(x^2) + \left( \frac{2K}{s} \right) \text{Cov}(Q_n, r_n) + \left( \frac{K}{s-Kq} \right) \left[ (I_a-1) + \left( \frac{Kq}{s} \right) \left( \begin{array}{c}
0 \\text{Cov}(Q_n, r_n) + \mathbb{E}(Q_n) \mathbb{E}(r) \end{array} \right) \right] \\
\end{array} \right) + \left( \frac{2K}{s} \right) \mathbb{E}(Q_n) \mathbb{E}(r)
$$

Adding (5.40) to (5.55) gives

$$
\mathbb{E}(D) = q \left( \frac{s}{2} \right) \left( \begin{array}{c}
\left( \frac{s}{s-Kq} \right) \left[ \mathbb{E}(x^2) + \frac{2}{q} \text{Cov}(Q_n, r_n) + \mathbb{E}(Q_n) \mathbb{E}(r) \right] + \left( \frac{K}{s-Kq} \right) \left[ (I_a-1) + \\
\left( \frac{K}{s} \right) \left( \begin{array}{c}
0 \\text{Cov}(Q_n, r_n) + \mathbb{E}(Q_n) \mathbb{E}(r) \end{array} \right) \right] \\
\end{array} \right)
$$

which, on dividing by $\mathbb{E}[\bar{A}(r_n+g_n)] = q[\mathbb{E}(r) + \mathbb{E}(g)]$ and using (5.18), gives (5.50).

The method of analysis leading to the formula (5.50) for the mean delay can be used to investigate other quantities. Formulae are now derived for the mean number of stops per vehicle and the proportion of vehicles that have to stop. Here, any vehicle which is delayed at all is supposed to stop. A correction could be made after that devised by Akqelik (1980) which lead to (5.35) to obviate this assumption.
Theorem 5.9

The mean number of times each vehicle in a stream of traffic stops is given by

\[ H = \frac{E(r)}{[E(r)+E(g)]} \left( \frac{s}{s-Kq} \right) + \frac{E(Q_n)}{[E(r)+E(g)]q} \]  

(5.57)

Proof

Using the notation (5.43) which was introduced in the proof of lemma 5.7, \( \tilde{H}_g(Q) \), the total number of stops incurred during an effective green period of unlimited duration which commences with an initial queue of \( Q \) vehicles is given by

\[ \tilde{H}_g(Q) = \sum_{n=1}^{\infty} A_n \]  

(5.58)

Taking expectations and using (5.47) gives

\[ E[\tilde{H}_g(Q)] = \sum_{n=1}^{\infty} E(A_n) \]
\[ = \sum_{n=1}^{\infty} \left( \frac{Kq}{s-Kq} \right)^n Q \]
\[ = \left( \frac{Kq}{s-Kq} \right)Q \]  

(5.59)

Using an argument similar to the one which lead to (5.55) in the proof of theorem 5.8, the expected total number of stops incurred during a single effective green period is given by

\[ E[\tilde{H}_g(Q(r_n)) = \left( \frac{Kq}{s-Kq} \right) \{E[\tilde{Q}(r_n)] - E[\tilde{Q}(r_{n+1})] \} \]  

(5.60)

Since each of the \( Q(r_n) \) vehicles in the queue at the start of an effective green period has stopped during the preceding effective red period,

\[ H = \frac{\left( \frac{Kq}{s-Kq} \right) \{E[\tilde{Q}(r_n)] - E[\tilde{Q}(r_{n+1})] \} + E[\tilde{Q}(r_n)]}{[E(r)+E(g)]q} \]
\[ = \frac{E(r)}{[E(r)+E(g)]} \left( \frac{s}{s-Kq} \right) + \frac{E(Q_n)}{[E(r)+E(g)]q} \]  

(5.57)

from (5.52) and (5.54).
Corollary 5.10

The proportion of vehicles in a stream of traffic that stop at least once is given by

\[ P_H = \frac{E(r)}{[E(r)+E(g)]} \left( \frac{S}{s-Kq} \right) \]  \hspace{1cm} (5.61)

Proof

The value of \( P_H \) can be found by proceeding as in the proof of theorem 5.9 but counting only the first stop made by each vehicle. The \( H_g \) vehicles which stop during an effective green period have all arrived during that period and thus all these correspond to first stops, so (5.60) remains appropriate. All the vehicles that arrive during an effective red period stop at least once, but any vehicle which has to stop because it forms part of the overflow has stopped at least once before. Thus

\[ P_H = \frac{(Kq)}{(s-Kq)} \left[ \frac{E[\bar{Q}(r_n)]-E[\bar{Q}(r_n+q_n)]+E[\bar{A}(r_n)]}{[E(r)+E(g)]q} \right] = \frac{E(r)}{E(r)+E(g)} \left( \frac{S}{s-Kq} \right) \]  \hspace{1cm} (5.61)

from (5.52) and (5.54).

The estimate of mean delay given by (5.50) at fixed-time traffic signals is less than McNeil's (1968) estimate (5.21) by an amount \( \frac{Ks}{c(s-Kq)^2} \). This difference, which is of the order of one departure time, is due entirely to the difference between the definitions of delay used in the analyses.

When applied to fixed-time traffic signals, the estimate of the proportion of vehicles that stop given by (5.61) reduces to Webster's (1958) formula (5.33). Since the argument which lead to (5.61) assumed that vehicles stop instantaneously as they join the queue, the estimate of \( P_H \) given by this formula can be used as an estimate of the proportion of vehicles which are delayed at all. When (5.57) is applied to fixed-time traffic signals, it reduces to Akçelik's estimate (5.34).
5.2.4 Mean overflow.

In this sub-section, an approximate formula is derived for the mean overflow in a signal-controlled stream of traffic. Here the durations of the effective red and green periods are treated as random variables which are not correlated with arrivals of the vehicles which they control. This formula is appropriate for use in conjunction with those derived in sub-section 5.2.3 to investigate the consequences for non-priority traffic of providing priority for buses.

The derivation given here follows that of Miller (1963) but requires weaker hypotheses than were used there. The argument used requires that the mean rates at which vehicles arrive during consecutive red and green periods are identical. This requirement is certainly satisfied by the compound Poisson model of arrivals which is adopted here.

Use is made in the analysis of the approximate statistics (5.9) of \( \tilde{N}(g) \). Since this analysis starts from the bulk-service approximation and then uses further approximations, the use of exact statistics of \( \tilde{N}(g) \), were they available, would not yield an exact result. The analysis commences with two preliminary results.

\textbf{Lemma 5.11}

The variance of the difference between the number of possible vehicular departures during an effective green period and the number of vehicular arrivals during that effective green period and the preceding effective red period is given approximately by

\[
\text{Var}[\tilde{N}(g_n) - \tilde{A}(r_n, g_n)] = \left( \frac{S}{K} \right)^2 \text{Var}(g) + \left( \frac{S}{K} \right) \text{E}(g) \text{C}^2 - 2q \left( \frac{S}{K} \right) \text{Cov}(r_n, g_n) + q^2 \text{Var}(r) + q \text{E}(r_n, g_n) I_a \quad (5.62)
\]
Proof
Consider the statistics of $\tilde{\mathcal{A}}(r_n + g_n)$:

$$E\{\tilde{\mathcal{A}}(r_n + g_n)\} = E\{E[\tilde{\mathcal{A}}(r_n + g_n) | r_n, g_n]\}$$

$$= qE(r_n + g_n)$$ \hspace{1cm} (5.63)

and

$$\text{Var}\{\tilde{\mathcal{A}}(r_n + g_n)\} = E\{E\{[\tilde{\mathcal{A}}(r_n + g_n)]^2 | r_n, g_n\} - E\{E[\tilde{\mathcal{A}}(r_n + g_n) | r_n, g_n]\}^2\}$$

$$= E\{q^2 (r_n + g_n)^2 + q(r_n + g_n)I_a - q^2 (E(r_n + g_n))^2\} \text{ from (5.63)}$$

$$= q^2 [\text{Var}(r) + 2Cov(r_n, g_n) + \text{Var}(g)] + qE(r_n + g_n) I_a$$ \hspace{1cm} (5.64)

Expanding the left-hand side of (5.62) gives

$$\text{Var}[\tilde{\mathcal{N}}(g_n) - \tilde{\mathcal{A}}(r_n + g_n)] = \text{Var}[\tilde{\mathcal{N}}(g_n)] - 2Cov[\tilde{\mathcal{N}}(g_n), \tilde{\mathcal{A}}(r_n + g_n)] + \text{Var}[\tilde{\mathcal{A}}(r_n + g_n)]$$

$$= (\frac{s}{k})^2 \text{Var}(g) + (\frac{s}{k})E(g)C_s^2 - 2(\frac{s}{k}) q[\text{Var}(g) + \text{Cov}(r_n, g_n)] +$$

$$q^2 [\text{Var}(g) + 2Cov(r_n, g_n) + \text{Var}(r)] + qE(r_n + g_n) I_a \text{ from (5.9) and (5.63-4)}$$

$$= (\frac{s}{k} - q)^2 \text{Var}(g) + (\frac{s}{k}) E(g) C_s^2 - 2q(\frac{s}{k} - q) \text{Cov}(r_n, g_n) + q^2 \text{Var}(r) + qE(r_n + g_n) I_a$$ \hspace{1cm} (5.62)


Lemma 5.12

Let $\delta_n = \text{Max}\{0, \tilde{\mathcal{N}}(g_n) - [Q + \tilde{\mathcal{A}}(r_n + g_n)]\}$

and $g^*_n = \text{Min}\{g_n, \tilde{\mathcal{S}}(Q + \tilde{\mathcal{A}}(r_n + g_n))\}$

then

$$\text{Var}(\delta_n) = (\frac{s}{k})^2 \text{Var}(g_n - g^*_n) + E(\delta) C_s^2$$ \hspace{1cm} (5.65)

Proof

Because of the inverse relationship between the functions $\tilde{\mathcal{N}}(g)$ and $\tilde{\mathcal{S}}(n)$, the quantity $g^*_n$ defined above can be used to give the simple approximation

$$\delta_n = \tilde{\mathcal{N}}(g_n - g^*_n)$$

Using (5.9) together with this approximation gives (5.65). \hspace{1cm} []
No attempt is made here to evaluate $\text{Var}(g_n - g_n^*)$ but instead the approximation

$$\text{Var}(q_n - q_n^*) = \left[ I_a + \left( \frac{Kq}{s} \right) q_n^2 \right] \text{E}(\delta) + \left[ \frac{(s - Kq) \text{E}(g) - Kq \text{E}(r)}{s \text{E}(g)} \right] \text{Var}(q_n)$$  \hspace{1cm} (5.66)

is adopted. Using (5.66) in (5.65) gives the estimate

$$\text{Var}(\delta_n) = \left[ I_a + \left( \frac{s + Kq}{s} \right) q_n^2 \right] \text{E}(\delta) + \left( \frac{s}{K} \right)^2 \left[ \frac{(s - Kq) \text{E}(g) - Kq \text{E}(r)}{s \text{E}(g)} \right] \text{Var}(q_n)$$  \hspace{1cm} (5.67)

The approximate formula for the mean overflow can now be derived as follows.

**Theorem 5.13**

If the overflow in a stream of traffic has a stationary distribution, then the mean value is given approximately by

$$\text{E}(Q_n) \approx \text{Max}\left[ 0, \left\{ \left[ \frac{q}{K} \left( \frac{\text{E}(r) - \text{E}(g)}{\text{E}(g)} \right) \right] + q^2 \right\} \text{Var}(g) - 2q \left( \frac{Kg}{K} \text{Cov}(r_n, g_n) + \text{Cov}(Q_n, g_n) \right) + 2q \text{Cov}(Q_n, r_n) + q^2 \text{Var}(r) + \left[ 2q \text{E}(r) + \left( 2q - \frac{s}{K} \right) \text{E}(g) \right] I_a q \left[ \left( \frac{s + Kg}{s} \right) \text{E}(r) + \left( \frac{Kg}{s} \text{E}(g) \right) q_n^2 \right] \right] + 2\left( \frac{Kg}{K} \text{E}(g) - q \text{E}(r) \right) \right]$$  \hspace{1cm} (5.68)

**Proof**

Consider the random variable $\delta_n$ defined in lemma 5.12. This can be used to simplify the bulk-service recurrence relation (5.23) as follows

$$Q_{n+1} = \text{Max}[0, Q_n + \Delta(r_n + g_n) - \bar{N}(g_n)] \quad \text{from (5.23)}$$

$$= Q_n + \Delta(r_n + g_n) - \bar{N}(g_n) + \delta_n$$  \hspace{1cm} (5.69)

From the definition given above, $\delta_n$ takes non-zero values only when $Q_{n+1}$ is 0. Thus

$$Q_{n+1} \delta_n = 0$$  \hspace{1cm} (5.70)
Taking expectations over (5.69) and using the consequences (5.54) of the stationarity of the distribution of the queue size gives

\[ \mathbb{E}(\delta) = \mathbb{E}(\tilde{N}(g_n) - \bar{K}(r_n + g_n)) \]  
\[ (5.71) \]

Adding (5.71) to (5.69) and rearranging gives

\[ Q_{n+1} = Q_n - [\tilde{N}(q_n) - \bar{K}(r_n + q_n) - \mathbb{E}(\tilde{N}(g_n) - \bar{K}(r_n + q_n))] \]  
\[ (5.72) \]

After squaring both sides of (5.72) and using (5.70), taking expectations, using the estimate (5.9) for \( \mathbb{E}(\tilde{N}(g_n)) \) and rearranging gives

\[ \mathbb{E}(Q_{n+1}^2) + 2\mathbb{E}(Q)\mathbb{E}(\delta) + \text{Var}(\delta) = \mathbb{E}(Q_n^2) - 2\left(\frac{S}{\kappa} - q\right) \text{Cov}(Q_n, q_n) + 2q \text{Cov}(Q_n, r_n) + \]
\[ + \text{Var}(\tilde{N}(g_n) - \bar{K}(r_n + q_n)) \]  
\[ (5.73) \]

Using (5.54) to eliminate the second moments of \( Q_n \) and \( Q_{n+1} \) in (5.73) and rearranging gives

\[ \mathbb{E}(Q_n) = \frac{\text{Var}(\tilde{N}(g_n) - \bar{K}(r_n + g_n)) - 2\left(\frac{S}{\kappa} - q\right) \text{Cov}(Q_n, q_n) + 2q \text{Cov}(Q_n, r_n) - \text{Var}(\delta)}{2\mathbb{E}(\delta)} \]  
\[ (5.74) \]

Substituting in the right-hand side of (5.74) for \( \text{Var}(\delta) \) from (5.67), \( \mathbb{E}(\delta) \) from (5.71) and \( \text{Var}(\tilde{N}(g_n) - \bar{K}(r_n + g_n)) \) from (5.62), rearranging and using (5.9) yields the second argument of the maximum function in (5.68). Since \( Q_n \) is a positive quantity, it has a positive mean value. Thus if a negative value is given by this expression, the value 0 certainly gives a better estimate of \( \mathbb{E}(Q_n) \). Thus (5.68) is established. [ ]
Corollary 5.14

Under the hypotheses of theorem 5.13, an approximate upper bound on the mean overflow is given by

\[ E(Q_n) \leq \left[ \left( \frac{s}{k} - q \right)^2 \text{Var}(g) - 2 \left( \frac{s}{k} - q \right) \{ q \text{Cov}(x_n, g_n) + \text{Cov}(Q_n, g_n) \} + \right. \]

\[ + 2q \text{Cov}(Q_n, x_n) + q^2 \text{Var}(x) + q \mathbb{E}(x_n + g_n) I_a + \left( \frac{s}{k} \right) E(g) C^2 \]

\[ \leq 2 \left[ \left( \frac{s}{k} - q \right) E(g) - q \mathbb{E}(x) \right] \]

(5.75)

Proof

Since \( \frac{\text{Var}(\delta)}{2E(\delta)} \) is a positive quantity, adding it to the right-hand side of (5.74) gives an approximate upper bound for \( E(Q_n) \). Substituting for \( E(\delta) \) from (5.71) and for \( \text{Var}[\tilde{N}(g_n) - \tilde{A}(r_n + g_n)] \) from (5.62) in the resulting expression, rearranging and using (5.9) gives (5.75). [ ]

The approximation (5.67) for \( \text{Var}(\delta) \) is the weakest part of the analysis leading to the estimate (5.68) for \( E(Q_n) \). When applied to fixed-time traffic signals with constant departure times during the effective green period, (5.67) reduces to

\[ \text{Var}(\delta) \approx I_a E(\delta) \]

(5.76)

which is the estimate adopted by Miller (1963). As the traffic intensity tends to 1, \( E(\delta) \) given by (5.71) tends to 0 so \( \text{Var}(\delta) \) given by (5.67) also tends to 0. However, as the traffic intensity tends to 0, the estimate of \( \text{Var}(\delta) \) given by (5.67) tends to \( \left( \frac{s}{k} \right) \left[ I_a + \left( \frac{s+K_g}{s} \right) C^2 \right] E(g) + \left( \frac{s}{k} \right)^2 \text{Var}(g) \), an overestimate of the correct limiting value which is equal to \( \left( \frac{s}{k} \right)^2 \text{Var}(g) \). Thus like Miller's (1963) estimate (5.76) for \( \text{Var}(\delta) \), (5.67) overestimates this quantity when the degree of saturation is small.

When applied to fixed-time traffic signals with constant departure times during the effective green periods, (5.68) reduces to Miller's (1963) estimate (5.30) for \( E(Q_n) \). In the same circumstances, (5.75) reduces to McNeil's (1968) strict upper bound (5.31).
5.2.5 Covariation of the overflow.

The formulae for the mean delay (5.50) and the mean overflow (5.68) derived in the previous two sub-sections make use of the statistics $\text{Cov}(Q_n, r_n)$ and $\text{Cov}(Q_n, g_n)$. Expressions are derived in this sub-section which relate these statistics to those of the durations of the effective red and green periods.

Since the durations of the effective red and green periods are here assumed not to be correlated with arrivals of vehicles, the size of the overflow $Q_n$ cannot in itself affect the durations $r_n$ and $g_n$. However, the durations $r_i$ and $g_i$ ($i < n$) certainly have some effect on $Q_n$ and may themselves be correlated with $r_n$ and $g_n$ as a consequence of the control policy of the traffic signals. The first result given here relates $Q_n$ to events during the previous effective red and green periods.

**Lemma 5.15**

Suppose that the sizes of the overflow, effective red and effective green periods during cycle number $n-1$ are given by

$$
\begin{align*}
Q_{n-1} & = E(Q_{n-1}) + \delta Q_{n-1} \\
r_{n-1} & = E(r_{n-1}) + \delta r_{n-1} \\
g_{n-1} & = E(g_{n-1}) + \delta g_{n-1}
\end{align*}
$$

(5.77)

Then the expected value of $Q_n$ is given approximately by

$$
E[\bar{Q}_n(Q_{n-1}, r_{n-1}, g_{n-1})] = E(\bar{Q}_n) + [E(Q_{n-1}) + q_0 r_{n-1} - \frac{S}{k-q} \delta g_{n-1}] P(Q_n \neq 0)
$$

(5.78)

**Proof**

Let $\bar{Q}_n$ be the size of the overflow at the end of a realisation of the queueing process in which $Q_{n-1}$, $r_{n-1}$ and $g_{n-1}$ take their mean values. Then

$$
\bar{Q}_n = \bar{Q}_n[E(Q_{n-1}), E(r_{n-1}), E(g_{n-1})]
$$

(5.79)
Now suppose that the value of $Q_{n-1}$ is perturbed to $E(Q_{n-1}) + \delta Q_{n-1}$. For sufficiently small values of $\delta Q_{n-1}$, $Q_n$ will be changed by $\delta Q_{n-1}$ if $\hat{Q}_n \neq 0$ and will remain unchanged otherwise. Similarly, if $r_{n-1}$ is perturbed to $E(r_{n-1}) + \delta r_{n-1}$, then for sufficiently small values of $\delta r_{n-1}$, $Q_n$ will be changed by $\beta(\delta r_{n-1})$ if $\hat{Q}_n \neq 0$ and will remain unchanged otherwise. Finally, if $g_{n-1}$ is perturbed to $E(g_{n-1}) + \delta g_{n-1}$, then for sufficiently small values of $\delta g_{n-1}$, $Q_n$ will be changed by $\tilde{A}(\delta g_{n-1}) - N(\delta g_{n-1})$ if $\hat{Q}_n \neq 0$ and will remain unchanged otherwise.

Combining these observations gives

$$\tilde{Q}_n(Q_{n-1}, r_{n-1}, g_{n-1}) = \hat{Q}_n + [\delta Q_{n-1} + \beta(\delta r_{n-1}) + \tilde{A}(\delta g_{n-1}) - N(\delta g_{n-1})] \chi_n + o(\delta Q_{n-1}) + o(\delta r_{n-1}) + o(\delta g_{n-1})$$

(5.80)

where $\chi_n = \begin{cases} 1 & \text{if } \hat{Q}_n \neq 0 \\ 0 & \text{otherwise} \end{cases}$

Taking expectations in (5.80) over the arrivals and departures, using (5.9) and that $E(\chi_n) = P(\hat{Q}_n \neq 0)$ gives

$$E[\tilde{Q}_n(Q_{n-1}, r_{n-1}, g_{n-1})] = E(\hat{Q}_n) + [\delta Q_{n-1} + q \delta r_{n-1} - (\delta g_{n-1})] P(\hat{Q}_n \neq 0)$$

Approximating $E(\hat{Q}_n)$ and $P(\hat{Q}_n = 0)$ by $E(Q_n)$ and $P(Q_n = 0)$ respectively gives (5.78).

**Theorem 5.16**

The covariances of the size of the overflow with the durations of the next effective red and green periods are given approximately by

$$\text{Cov}(Q_n, r_n) \approx \sum_{i=1}^{\infty} [q \text{Cov}(r_{n-i}, r_n) - (\delta g_{n-1}) \text{Cov}(g_{n-1}, r_n)] [P(Q_n = 0)]^i$$

(5.81)

and

$$\text{Cov}(Q_n, g_n) \approx \sum_{i=1}^{\infty} [q \text{Cov}(r_{n-i}, g_n) - (\delta g_{n-1}) \text{Cov}(g_{n-1}, g_n)] [P(Q_n = 0)]^i$$

(5.82)

**Proof**

Consider the covariance of $Q_{n-1}$ with $r_n$ for some $i \geq 0$. By definition,
this is given by
\[ \text{Cov}(Q_{n-1}, r_n) = E \{ Q_{n-1} (Q_{n-1-1} r_{n-1-1} q_{n-1-1} r_n) - E(Q_{n-1})E(r_n) \}
= E \{ E(Q_{n-1})r_n + [\delta Q_{n-1} + q \delta r_{n-1} - (z-q) \delta q_{n-1}]P(Q_{n-1} \neq 0) r_n \} - E(Q_{n-1})E(r_n) \text{ from (5.78)}
= [\text{Cov}(Q_{n-1-1}, r_n) + q \text{Cov}(r_{n-1-1}, r_n) - (z-q) \text{Cov}(q_{n-1-1}, r_n)]P(Q_{n-1} \neq 0)
\quad (i \geq 0) \quad (5.83) \]

(since \( E(\delta Q_{n-1}) = E(\delta r_{n-1}) = E(\delta q_{n-1}) = 0 \)).

Treating \( P(Q_{n-1} \neq 0) \) as being independent of \( i \), multiplying both sides of (5.83) by \( [P(Q_n \neq 0)]^i \), summing over all positive values of \( i \) and rearranging gives (5.81). A similar argument where \( r_n \) is replaced by \( g_n \) leads to (5.82).

There remains the problem of finding the probability that the overflow is not zero, i.e. \( P(Q_n \neq 0) \). Miller's (1969) empirical formula (5.27) for the complementary probability \( P(Q_n = 0) \) can be used to provide an estimate. Although this formula gives remarkably good estimates when applied to fixed-time traffic signals, it was derived in an empirical manner which prevents its extension to more general signal-control policies. Because stochastic variations in the durations of the effective red and green periods will tend to make non-zero overflows more common, this approximation will underestimate \( P(Q_n \neq 0) \). However, this quantity is used here as a surrogate for the conditional probability \( P(\hat{Q}_n \neq 0) \). Here again stochastic variations will tend to make non-zero overflows more common than those occurring after cycles which start with a mean overflow and which have effective red and green periods of mean duration, so the value of \( P(Q_n \neq 0) \) will exceed that of \( P(\hat{Q}_n \neq 0) \).

These two effects will cause errors of opposite sign and so will tend to compensate for each other. In any case, the use of an exact expression for this quantity would not in itself lead to exact expressions for the covariances required.
5.3 Statistics of service

5.3.1 Introduction.

In the last section, various formulae were given to estimate the mean delay and other related quantities for a stream of traffic which is subject to effective red and green periods of variable duration. In order to apply these formulae, some statistics of the durations of the effective red and green periods are required. Sufficient information for this is provided by the quantities

\[
\begin{align*}
E(r) \\
E(r^2) \\
E(g) \\
E(g^2) \\
\text{Cov}(r_{n-i}, g_n) \\
\text{Cov}(r_{n-i}, r_n) \\
\text{Cov}(g_{n-i}, g_n)
\end{align*}
\]  

(5.84)

In this section, explicit formulae are derived to estimate these quantities for any stream of traffic at a junction where there is responsive priority for buses. Two combinations of priority rules are investigated: these are priority by extension and recall and priority by extension and recall with inhibition. The analysis presented here is appropriate to junctions where there are any number of streams of traffic and stages in the sequence provided that each stream receives right of way during a single interval in each uninterrupted cycle. As the analysis is presented here, each stream is supposed to receive a red signal indication throughout at least one stage in the sequence. This does not restrict the applicability of the analysis since if a stream receives right of way at all times except during some part of a transition period, then an
additional stage of zero duration can be inserted into that transition period.

The duration of each stage is taken to be fixed except when it is altered in order to grant priority to a bus. As in Chapter 4, priority vehicles are supposed to arrive according to the Poisson law with mean rate $\beta/c$ and to receive right of way only during the first $k$ of the $m$ stages of the sequence.

Before proceeding with the analysis, some preliminary results are given. These are then used to determine the extent of the analysis which is necessary in order to exhaust all of the possibilities which are within the restrictions described above. The first of these results identifies seven distinct classes of streams at a junction and shows that each stream belongs to exactly one of them. These classes are considered separately when formulae are derived for the statistics (5.84).

The second result shows that each stream at a junction is complementary to a notional stream which necessarily belongs to a different class. As a result of this complementary association between pairs of streams, if the statistics (5.84) are known for one of them, then they can be used to deduce the corresponding statistics for the other one. This observation is used to reduce the number of classes of streams for which a direct analysis is required from the original 7 to 3. If a stream belongs to one of the other 4 classes, then the statistics (5.84) for this stream can be deduced from those for a notional complementary stream which is shown to belong to one of the 3 classes for which a direct analysis is provided.

For each stream, the stages in a cycle during which it has right of way can be specified by the numbers of the stages which, in the absence of priority for buses, form the first and last parts of the period during
which that stream has right of way. Thus let

\( \bar{n}_1(j) \) be the number of the stage during which stream \( j \) first has right of way in the absence of priority for buses.

\( \bar{n}_2(j) \) be the number of the stage during which stream \( j \) last has right of way in the absence of priority for buses.

**Lemma 5.17 (Classification of streams)**

Each stream of traffic at a signal-controlled junction belongs to exactly one of the seven classes defined as follows.

- **Class 1**: \( 1 \leq n_1 \leq n_2 < k \)
- **Class 2**: \( 1 \leq n_1 < k, n_2 \leq m \)
- **Class 3**: \( 1 \leq n_1 < n_2 \leq k \)
- **Class 4**: \( k < n_1 \leq n_2 \leq m \)
- **Class 5**: \( 1 \leq n_1 \leq k < n_2 \leq m \)
- **Class 6**: \( k < n_1 < n_2 \leq m \)
- **Class 7**: \( 1 = n_1 \leq k < n_2 < m \).

**Proof**

Suppose initially that \( n_1 = 1 \). Either \( 1 \leq n_2 \leq k \), in which case the stream is in Class 1 or \( k < n_2 \leq m \), in which case the stream is in Class 7. There is no possibility that for this stream \( n_2 = m \) since then it would have right of way throughout the cycle.

Now suppose that \( 1 < n_1 \leq k \). Here there are three possibilities. Either \( k < n_2 < n_1 \), in which case the stream is in Class 3, or \( 1 \leq n_2 \leq k \), in which case the stream is in Class 1, or \( k < n_2 \leq m \), in which case the stream is in Class 2.

Finally suppose that \( k \leq n_1 \leq m \). Again there are three possibilities to consider. Either \( 1 \leq n_2 \leq k \), in which case the stream is in Class 5, or
k \leq n_2 < n_1$, in which case the stream is in Class 6, or $n_1 \leq n_2 \leq m$, in which case the stream is in Class 4.

Since one of the three cases $n_1 = 1$, $1 < n_1 \leq k$, and $k < n_1 \leq m$ must hold, this exhausts all the possible combinations of $n_1$ and $n_2$. A stream cannot belong to more than one class since the definitions are mutually exclusive.

Lemma 5.18

The quantities (5.84) for stream $j$ can be found by considering a notional stream $j'$ for which

\[
\begin{align*}
\tilde{n}_1(j') &= \begin{cases} 
1 & \text{if } \tilde{n}_2(j) = m \\
\tilde{n}_2(j) + 1 & \text{otherwise}
\end{cases} \\
\tilde{n}_2(j') &= \begin{cases} 
m & \text{if } \tilde{n}_1(j) = 1 \\
\tilde{n}_1(j) - 1 & \text{otherwise}
\end{cases}
\end{align*}
\]

(5.85)

\[a_{i0j'} = 1 - a_{i0j} \quad (1 \leq i \leq m)\]

and \[a_{i1j'} = 1 - a_{i1j} \quad (k < i \leq m)\]

where $a_{i0j}$ and $a_{i1j}$ are defined in sub-section 4.2.3. Furthermore, the class of stream $j'$ is different to the class of stream $j$.

Proof

If stream $j'$ satisfies the prescription (5.85), then it has right of way at exactly those times when stream $j$ does not. Thus the effective red and green periods for stream $j$ correspond to the effective green and red periods respectively for stream $j'$. The following relationships follow immediately from this observation.
\[ \begin{align*}
E\{\bar{r}(j)\} &= E\{\bar{g}(j')\} \\
E\{[\bar{r}(j)]^2\} &= E\{[\bar{g}(j')]^2\} \\
E\{\bar{g}(j)\} &= E\{\bar{r}(j')\} \\
E\{[\bar{g}(j)]^2\} &= E\{[\bar{r}(j')]^2\} \\
\text{Cov}[\bar{r}_{n-1}(j), \bar{r}_n(j)] &= \text{Cov}[\bar{g}_{n-1}(j'), \bar{r}_n(j')] \\
\text{Cov}[\bar{r}_{n-1}(j), \bar{g}_n(j)] &= \text{Cov}[\bar{g}_{n-1}(j'), \bar{g}_n(j')] \\
\text{Cov}[\bar{g}_{n-1}(j), \bar{g}_n(j)] &= \text{Cov}[\bar{r}_{n-1}(j'), \bar{r}_n(j')] \\
\end{align*} \] (5.86)

According to the prescription (5.85) and the classification scheme of lemma 5.17, the class of stream \( \tilde{j}'(j) \) depends on \( \bar{n}_1(j) \) and \( \bar{n}_2(j) \) as follows:

\[
\tilde{c}(j') = \begin{cases} 
3 & n_1 > 1 \\
2 & n_1 = 1 \\
5 & n_2 < k \\
4 & n_2 = k \\
5 & n_2 < m \\
4 & n_2 = m \\
1 & \tilde{c}(j) = 1 \\
1 & \tilde{c}(j) = 2 \\
1 & \tilde{c}(j) = 3 \\
6 & n_1 > k+1 \\
5 & n_1 = k+1 \\
7 & n_1 > k+1 \\
1 & n_2 < m \\
2 & n_2 = m \\
1 & n_1 = k+1 \\
4 & n_1 > k+1 \\
4 & n_2 < k \\
4 & n_2 = k \\
\end{cases} 
\] (5.87)

Thus \( \tilde{c}[\tilde{j}'(j)] \neq \tilde{c}(j) \) for any \( \bar{n}_1(j), \bar{n}_2(j) \).

[ ]

**Theorem 5.19**

If formulae are available for the statistics (5.84) for arbitrary streams in Classes 1, 2 and 4, then these statistics can be found for a stream in any class.

**Proof**

If the stream is in Class 1, 2 or 4, then by hypothesis the statistics (5.84) can be calculated directly. If the stream is in Class 3, then according to the result of lemma 5.18, the required statistics can be
deduced from (5.86) and the corresponding statistics for a stream which, according to (5.87), is in Class 1, for which formulae are available. Similarly, if the stream is in Class 5, then the required statistics can be found by considering a stream which, according to (5.87), is in Class 1, 2 or 4. Finally, if the stream is in Class 6 or 7, then the required statistics can be deduced by considering a stream which is in Class 4, for which formulae are available.

5.3.2 Priority by extension and recall.

In this sub-section, explicit formulae are derived for the statistics (5.84) for streams in Classes 1, 2 and 4 when priority by extension and recall is implemented. The result of theorem 5.19 shows that these formulae provide sufficient information to find the statistics (5.84) for an arbitrary stream in any class. The formulae are derived for each of these classes of stream in turn.

Theorem 5.20

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of $\beta/c$. Then the statistics (5.84) can be calculated for any stream $j$ in Class 1 from the formulae

\[
E(r) = c \sum_{n=k+1}^{m} \frac{\rho_n \cdot n!}{n \cdot n!} \\
E(r^2) = c^2 \sum_{n=k+1}^{m} \frac{\rho_n^2 + (1-\delta_{kn_2}) \cdot n_1 + 2 \rho_n \cdot n_1 + n_2 n_1}{n!} \\
\text{where } \rho_n = \frac{\sum_{i=1}^{n_2} \lambda_i + (1-\delta_{kn_2}) (\xi_i - \lambda_i) + \sum_{i=n_2}^{n_1} (1-a_{i0j}) \lambda_i + (1-a_{n1j}) \lambda_{n1}}{n_1-1} + \sum_{i=1}^{n_1-1} [\lambda_i + (1-a_{i0j}) \lambda_{i0}] \\
\text{(k<n≤m)}
\]

\[
E(g) = c \sum_{n=k+1}^{m} \frac{\sigma_n \cdot n!}{n \cdot n!} \\
E(g^2) = c^2 \sum_{n=k+1}^{m} \frac{\sigma_n^2 + \delta_{kn_2} \cdot n_2}{n!} \\
\text{where } \sigma_n = a_{n1j} \lambda_{n1} + \sum_{i=n_1}^{n_2} \lambda_i + \sum_{i=1}^{n_2} a_{i10j} \lambda_{i0} + \delta_{kn_2} (\xi_i - \lambda_i) \text{ (k<n≤m)}.
\]
\[ \text{Cov}(r_{\ell-1}, g_{\ell}) = \begin{cases} E(x_{\ell} g_{\ell}) - E(x) E(g) & (i=0) \\ 0 & \text{otherwise} \end{cases} \]  

where \( E(x_{\ell} g_{\ell}) = c^2 \sum_{n=k+1}^{m} P_n (\eta_n + \eta_1) \sigma_n \)  

\[ \text{Cov}(x_{\ell-1}, x_{\ell}) = \text{Cov}(g_{\ell-1}, g_{\ell}) = 0 \quad (i \geq 1) \]  

where \( \varepsilon_k \) and \( \upsilon_k \) are given by (4.52), \( P_n \) \((k \leq m)\) are given by (4.62) and \( \eta_n, \eta_{2n} \) \((k \leq m)\) are given by (4.65-6).

**Proof**

Consider an effective red period for stream \( j \) in Class 1. This is initiated by a call of stage \( n_2 + 1 \) and is terminated by the next call of stage \( n_1 \). During the intervening period, stage 1 is called exactly once: the number of the stage which is running when this happens provides a convenient means to condition the moments of the duration of the effective red period. Suppose that this number is \( n \) \((k \leq m)\): if \( n_2 < k \), then the duration of the effective red period up to the start of stage \( n \) varies as a result of any priority extensions which are granted during stage \( k \). The first two moments of the duration of this period are

\[
c_{[1 \sum_{n\geq n_2+1} 1_{i=n} \lambda_1 + (1-\delta_{kn}) (\varepsilon_k - \lambda_1) + \sum_{i=n_2}^{n-1} (1-a_{10j}) \lambda_10]}
\]

\[
c^2 \left[ [1 \sum_{n\geq n_2+1} 1_{i=n} \lambda_1 + (1-\delta_{kn}) (\varepsilon_k - \lambda_1) + \sum_{i=n_2}^{n-1} (1-a_{10j}) \lambda_10] \right]^2 + (1-\delta_{kn}) \upsilon_k
\]

respectively, where according to the result of corollary 4.20, \( \varepsilon_k \) and \( \upsilon_k \) are given by (4.52). The conditional moments of the time for which stage \( n \) runs are given in the result of lemma 4.25 as \( c_{\eta_n} \) and \( c^2 \eta_{2n} \) (4.65-6). The remaining part of the effective red period has a duration which is conditionally constant at \( c_{[(1-a_{n1j}) \lambda_n + \sum_{i=n_1}^{n} 1 \lambda_1 + (1-a_{10j}) \lambda_10]} \). Combining these parts gives the conditional moments

\[
E(r | n) = c_{[\rho_n + \eta_n]} \quad (k \leq m)
\]

\[
E(r^2 | n) = c^2 \left[ c_{[\rho_n + \delta_{kn}]}, \upsilon_k + 2 \rho_n \eta_n + \eta_{2n} \right]
\]  

(5.91)
where \( \rho_n \) is as given in (5.88). According to the result of lemma 4.24, the probabilities of the conditioning events are given by \( P_n \) of (4.62). Multiplying each of the conditional moments (5.91) by the appropriate conditioning probability and summing over \( n \) gives (5.88).

Now consider an effective green period for stream \( j \). Since \( 1 \leq n_1, n_2 \leq k \), there are two possible sources of variation in the duration of this period.

(a) If \( n_1 = 1 \), then depending on the stage which is running when stage 1 is called, stream \( j \) may receive different amounts of effective green time before the start of stage 1. If the transition is from stage \( n \) (\( k \leq n \leq m \)) to stage 1, then this is given by \( a_{n_1}^{1-j-n_1} c \).

(b) If \( n_2 = k \), then any priority extension to stage \( k \) will add to the effective green time for stream \( j \).

Thus, the conditional moments of the duration of the effective green period given that stage 1 is called while stage \( n \) (\( k \leq n \leq m \)) is running are given by \( c \sigma_n \) and \( c^2 (\sigma_n^2 + \delta_{n_1} \nu_k) \) (\( k \leq n \leq m \)). Arguing as for the moments of the duration of the effective red period leads to (5.89).

Finally, the durations of all distinct effective red and green periods are mutually independent except possibly for an effective red period and the effective green period which follows on immediately. This exception arises because the source of variation in the duration of the effective green period described above as case (a) is correlated with the duration of the preceding effective red period. The distributions of the durations of these periods are conditionally independent given the number of the stage which is running when stage 1 is called. Thus
\[ E(r_{\ell+1}|g_{\ell}) = \sum_{n=k+1}^{m} P\left[r \mid n\right] E(g \mid n) \]

\[ = c^2 \sum_{n=k+1}^{m} P(n + \eta_n) \sigma_n \]

so the expressions (5.90) for \( \text{Cov}(r_{\ell-1}, g_{\ell}) \) are established.

Theorem 5.21

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of \( \beta/c \). Then the statistics (5.84) can be calculated for any stream \( j \) in Class 2 from the formulae

\[
E(r) = c \left[ \sum_{n=k+1}^{m} P(n + \eta_n) \right] \sum_{n=2}^{n_2} \frac{\lambda_{n+1}}{\lambda_n} \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_n}{\lambda_{n+1}} \right] \sum_{i=1}^{n} \frac{\lambda_{n+1}}{\lambda_n} \]

\[
E(r^2) = \frac{c^2}{\lambda_1} \left[ \sum_{n=k+1}^{m} P(n + \eta_n) \right] \sum_{n=2}^{n_2} \frac{\lambda_{n+1}}{\lambda_n} \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_n}{\lambda_{n+1}} \right] \sum_{i=1}^{n} \frac{\lambda_{n+1}}{\lambda_n} \]

where \( \rho_n = \left\{ \begin{array}{ll}
(1-a_{n1j})\lambda_{n-1} + \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_i}{\lambda_{i+1}} \right] & \text{for } n < n_2 \\
(1-a_{n20j})\lambda_{n-2} + \sum_{i=1}^{n-2} \left[ \frac{(1-a_{10j})\lambda_i}{\lambda_{i+1}} \right] & \text{for } n_2 < n \leq m
\end{array} \right. \}

\[
E(g) = c \left[ \sum_{n=k+1}^{m} P(n + \eta_n) \right] \sum_{n=2}^{n_2} \frac{\lambda_{n+1}}{\lambda_n} \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_n}{\lambda_{n+1}} \right] \sum_{i=1}^{n} \frac{\lambda_{n+1}}{\lambda_n} \]

\[
E(g^2) = \frac{c^2}{\lambda_1} \left[ \sum_{n=k+1}^{m} P(n + \eta_n) \right] \sum_{n=2}^{n_2} \frac{\lambda_{n+1}}{\lambda_n} \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_n}{\lambda_{n+1}} \right] \sum_{i=1}^{n} \frac{\lambda_{n+1}}{\lambda_n} \]

where \( \sigma_n = \left\{ \begin{array}{ll}
\frac{a_{n1-1,0} \lambda_{n-1} + \sum_{i=1}^{n-1} \left[ \frac{(1-a_{10j})\lambda_i}{\lambda_{i+1}} \right]}{\lambda_{n+1}} & \text{for } n < n_2 \\
\frac{a_{n2-1,0} \lambda_{n-2} + \sum_{i=1}^{n-2} \left[ \frac{(1-a_{10j})\lambda_i}{\lambda_{i+1}} \right]}{\lambda_{n+1}} & \text{for } n_2 < n \leq m
\end{array} \right. \}

\[
\text{Cov}(r_{\ell-1}, g_{\ell}) = \left\{ \begin{array}{ll}
E(r_{\ell+1}g_{\ell}) - E(r)E(g) & \text{for } i=-1 \\
0 & \text{otherwise}
\end{array} \right. \]

where \( E(r_{\ell+1}g_{\ell}) = c^2 \left[ \sum_{n=k+1}^{m} P(n + \eta_n) \right] \sum_{n=2}^{n_2} \frac{\lambda_{n+1}}{\lambda_n} \sum_{i=1}^{n} \left[ \frac{(1-a_{10j})\lambda_i}{\lambda_{i+1}} \right] \sum_{i=1}^{n} \frac{\lambda_{n+1}}{\lambda_n} \]

\[
\text{Cov}(r_{\ell-1}, r_{\ell}) = \text{Cov}(g_{\ell-1}, g_{\ell}) = 0 \quad \text{(i\geq1)}
\]
where \( e_k \) and \( u_k \) are given by (4.52),

\[
P_n \quad (k \leq n \leq m)
\]

are given by (4.62)

and \( n_n, n_{2n} \quad (k \leq n \leq m) \) are given by (4.65-6).

Proof

According to the definition of Class 2 given in lemma 5.17, \( 1 \leq n_1 \leq k \leq n_2 \leq m \). Thus an effective red period for stream \( j \) can be initiated by a priority recall of stage 1 while stage \( n \quad (k \leq n \leq n_2) \) is running and will otherwise be initiated when stage \( n_{2} + 1 \) is called if \( n_2 < m \) and when stage 1 is called if \( n_2 = m \). In the first case, the duration of the effective red period is conditionally equal to \( c \rho_n \quad (k \leq n \leq n_2) \) and in the second case it is a random variable with conditional moments given by

\[
\begin{align*}
E(r_n | n) &= c(\rho_n + \eta_n) \\
E(r^2_n | n) &= c^2(\rho_n^2 + 2\rho_n \eta_n + \eta_{2n})
\end{align*}
\]

(5.96)

where \( \eta_n \) and \( \eta_{2n} \quad (k \leq n \leq m) \) are given by (4.65-6) and \( \rho_n \quad (k \leq n \leq m) \) are given in (5.93). Since the conditioning probabilities are \( P_n \quad (k \leq n \leq m) \) given in (4.62), (5.93) is established.

Now consider an effective green period for stream \( j \). This is always initiated when stage \( n_1 \) is called and will be terminated by a priority recall of stage 1 if this occurs before the end of stage \( n_2 \) when it will otherwise be terminated. In the first case, if stage 1 is recalled while stage \( n \quad (k \leq n \leq n_2) \) is running, then the conditional moments of the duration of the effective green period are given by

\[
\begin{align*}
E(g_n | n) &= c(\sigma_n + \eta_n) \\
E(g^2_n | n) &= c^2(\sigma_n^2 + \sigma_n \eta_n + \eta_{2n})
\end{align*}
\]

(5.97)
where \( \eta_n \) and \( \eta_{2n} \) (\( k n \leq \eta_n \)) are given by (4.65–6) and \( \sigma_n \) (\( k n \leq \eta_n \)) are given in (5.94). In the second case, the conditional moments of the duration of the effective green periods are \( c_\sigma_m \) and \( c(\sigma^2 + \mu_k) \), where \( \mu_k \) is given in (4.52) and \( \sigma_m \) is given in (5.94), irrespective of how many more stages there are before stage 1 is called. The appropriate conditioning probabilities for the first case are \( P_n \) (\( k n \leq \eta_n \)) and for the second case is \( \sum_{n=n_2+1}^m P_n \), so (5.94) is established.

Finally, the durations of all distinct effective red and green periods are mutually independent except for an effective green period and the effective red period which follows on immediately. This exception arises because in cases where an effective green period is terminated by a transition to stage 1, the next effective red period is initiated by the same transition. The distributions of the durations of these periods are conditionally independent given the number of the stage which is running when stage 1 is called. Thus

\[
E(g_r \ell_{n+1}) = c^2 \sum_{n=k+1}^m P_n E(g \| n) E(r \| n)
\]

\[
= \sigma_c \left[ \sum_{n=k+1}^{n_2} P_n (\sigma + \eta_n) \rho_n + \sigma \sum_{n=n_2+1}^m P_n (\rho + \eta_n) \right] \tag{5.98}
\]

from which the expressions (5.95) follow immediately.

Before proceeding to the derivation of formulae for streams in Class 4, some further notation is introduced for quantities which will be used in it. A preliminary result is then proved to establish expressions for these quantities. Let

\( \tilde{P}_n^+(N) \) be the conditional probability that there are exactly \( n \) stages in a cycle given that there are at least \( N \) (\( k < N < n \leq m \))

\( \mu_n \) be the mean time between the start of stage 1 and the next call of stage \( n \) (\( k < n \leq m \))
\( \mu_{2n} \) be the second moment of the time between the start of stage 1 and the next call of stage \( n \) (\( k \leq n \leq m \)).

**Lemma 5.22**

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of \( \beta/c \). Then the conditional probabilities \( \tilde{P}_{i}^{+}(N) \) are given by

\[
\tilde{P}_{i}^{+}(N) = \begin{cases} 
\tilde{P}_{0}(\sum_{j=N}^{i-1} \lambda_{j}) j_{0}^{1}(1-\tilde{P}_{0}(\lambda_{i-1},0,0,\lambda_{i1})) & (k < N \leq i < m) \\
\tilde{P}_{0}(\sum_{j=N}^{i-1} \lambda_{j}) j_{0}^{1} & (k < N \leq i = m)
\end{cases}
\]  

(5.99)

Furthermore, the first two moments of the time between the start of stage 1 and the next call of stage \( n \) (\( k \leq n \leq m \)) satisfy

\[
\begin{align*}
\mu_{n} &= \alpha_{n} + \frac{\sum_{j=k+1}^{n-1} P_{j}}{n-1} \nu_{n} \quad (k < n \leq m) \\
\mu_{2n} &= \frac{\nu_{k}}{2} + 2\mu_{n} \sigma_{n}^{2} + \frac{\nu_{n}^{2} + \nu_{2n}}{2} \quad (k < n \leq m)
\end{align*}
\]

(5.100)

where

\[
\begin{align*}
\nu_{n} &= \sum_{i=k+1}^{n-1} P_{i} (\alpha_{i} + \lambda_{i-1,0} + \eta_{1} + \lambda_{i1}) / \sum_{j=k+1}^{n-1} P_{j} \quad (k < n \leq m) \\
\nu_{2n} &= \sum_{i=k+1}^{n-1} P_{i} (\alpha_{i} + \lambda_{i-1,0} + \lambda_{i1})^{2} + \nu_{k} + 2(\alpha_{i} + \lambda_{i-1,0} + \lambda_{i1}) \eta_{1} + \eta_{2i} / \sum_{j=k+1}^{n-1} P_{j} \quad (k < n \leq m) \\
\alpha_{i} &= \sum_{j=1}^{i-1} \lambda_{j}^{1} + \sum_{j=1}^{i-2} \lambda_{j0}^{1} + (\epsilon_{i} - \lambda_{i1}) \quad (k < i \leq m)
\end{align*}
\]

(5.101)

and \( \epsilon_{k} \) and \( \nu_{k} \) are given in (4.52), \( P_{i} \) (\( k \leq i \leq m \)) are given by (4.62) and \( \eta_{1}, \eta_{2i} \) (\( k \leq i \leq m \)) are given by (4.65-6).
Proof

The conditioning event for the probabilities \( \bar{p}^+_n(N) \) \((N \leq n \leq m)\) is that there are at least \( N \) stages \((k < N \leq m)\) in a cycle. This event has probability \( \sum_{j=0}^{N-1} p_j \) where \( p_j \) \((k < j \leq m)\) are given by (4.62). Normalising the unconditional probabilities \( p_i \) \((N \leq i \leq m)\) by this probability gives

\[
\bar{p}^+_i(N) = \frac{p_i}{\sum_{j=N}^{m} p_j}
\]

\[
\text{for } (k < N \leq i < m) \quad \text{from (4.62)}
\]

\[
\bar{p}^+_0(N) = \frac{\left( \sum_{j=1}^{i-1} \lambda_j + \sum_{j=0}^{i-2} \lambda_j \right) \left[ 1 - \bar{p}_0(\lambda_{i-1} + \lambda_{i-1,0}) \right]}{\sum_{j=1}^{m-1} \sum_{j=0}^{m-2} N-1 \sum_{j=0}^{N-2} \bar{p}_0(j=0)}
\]

\[
\text{for } (k < N \leq i \leq m) \quad \text{from (5.99)}
\]

where the last step follows from the independence of bus arrivals during disjoint time intervals.

Using the terminology introduced in sub-section 4.4.2, consider the 2-state M-R system in which state 1 corresponds to the controller running any stage numbered \( i < n \) for some \( n \) \((k < n < m)\) or undergoing a transition between any two such stages. The trial period is the interval during which the arrival of a bus at the detector will cause a transition to stage 1, i.e., the period which starts with the call of stage \( k+1 \) and ends with the call of stage 1 (if a prolonging event occurs) or with the call of stage \( n \). A prolonging event corresponds to the arrival at the detector of a bus which is granted priority by a recall of stage 1: the period of prolongation is the time between the start of the trial period during which the prolonging event occurs and the next call of stage \( k+1 \). As in the case of priority by extension which was considered in corollary 4.19, the
independence of events in successive trial periods follows from the result of lemma 4.15.

The probability that a prolonging event occurs is the same as the probability that there are fewer than \( n \) stages in the cycle, or \( \sum_{j=k+1}^{N-1} P_j \) where \( P_j \) \((k \leq j \leq m)\) are given by (4.62). If a prolonging event occurs during the transition period immediately before stage \( i \) or during stage \( i \) itself \((k < i < n)\), the conditional moments of the duration of the exceeding period of prolongation are given by

\[
\bar{V}_2(i)c^2 = c^2[(\alpha + \lambda_i - 1, 0 + \lambda_i) + \nu_k + 2(\alpha_i + \lambda_i - 1, 0 + \lambda_i)\eta_i + \eta_{2i}]
\]

\[
\bar{V}_2(i)c^2 = c^2[\alpha_i + \lambda_i - 1, 0 + \lambda_i + \nu_k + 2(\alpha_i + \lambda_i - 1, 0 + \lambda_i)\eta_i + \eta_{2i}]
\]

where \( \alpha_i \) \((k < i < n)\) are as given in (5.100),

\( \nu_k \) and \( \nu_k \) are given in (4.52)

and \( \eta_i, \eta_{2i} \) \((k < i < n)\) are given by (4.65-6).

The conditional probability of the conditioning event, given the occurrence of a prolonging event, is \( P_i / \sum_{j=k+1}^{N-1} P_j \), so the quantities \( \nu_n \) and \( \nu_{2n} \) given in (5.100) correspond to the unconditional moments of the duration of the period of prolongation.

Now \( \alpha_n \) is the mean time from the start of stage 1 to the call of stage \( n \) when no bus is granted priority by recall. According to the result of theorem 4.17, the mean time in excess of this caused by priority recalls of stage 1 is given by \( \nu_n \sum_{j=k+1}^{N-1} P_j / (1 - \sum_{j=k+1}^{N-1} P_j) \) where \( \nu_n \) is given in (5.100). Thus the formula in (5.100) for \( \mu_n \) is established: the formula for \( \mu_{2n} \) follows from a second application of the result of theorem 4.17.
Theorem 5.23

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of $\beta/c$. Then the statistics (5.84) can be calculated for any stream in Class 4 from the formulae

\[
E(r) = c \left[ \sum_{n=1}^{n_2} \tilde{P}^+_{n}(n_1) \rho_n + \sum_{n=n_2+1}^{m} \tilde{P}^+_{n}(n_1)(\rho_n + \eta_n) + \mu_{n_1} \right]^{n_2}
\]

\[
E(r^2) = c^2 \left\{ \sum_{n=1}^{n_2} \tilde{P}^+_{n}(n_1)(\rho^2 + 2\rho \mu_n) + \sum_{n=n_2+1}^{m} \tilde{P}^+_{n}(n_1)(\rho^2 + 2\rho \eta_n + 2\eta_n \mu_{n_1} + \mu_{2n_1}) \right\}
\]

where

\[
\rho_n = \begin{cases} 
(1-a_{n1}) \lambda_n n_1 + (1-a_{n1-1,0}) j \lambda_n n_1 - 1,0 & (n_2 \leq n \leq n_2) \\
(1-a_{n2}) \lambda_n n_2 + 1 + \sum_{i=n_2+1}^{m} (\lambda_i + \lambda_n) + \lambda_n + 1 \lambda_n - 1,0 & (n_2 \leq n \leq m)
\end{cases}
\]

\[
E(g) = c \left[ \sum_{n=1}^{n_2} \tilde{P}^+_{n}(n_1)(\sigma_n + \eta_n) + \sigma \sum_{n=n_2+1}^{m} \tilde{P}^+_{n}(n_1) \right]^{n_2}
\]

\[
E(g^2) = c^2 \left[ \sum_{n=1}^{n_2} \tilde{P}^+_{n}(n_1)(\sigma^2 + 2\sigma \eta_n + 2\sigma \eta_n \mu_{n_1} + \sigma \sum_{n=n_2+1}^{m} \tilde{P}^+_{n}(n_1) \right]
\]

where

\[
\sigma_n = \begin{cases} 
(1-a_{n1-1,0}) \lambda_n n_1 - 0,0 + \sum_{i=n_1}^{n_2} (\lambda_i + \lambda_n) + a_{n1} \lambda n_1 & (n_2 \leq n \leq n_2) \\
(1-a_{n2}) \lambda_n n_2 - 0,0 + \sum_{i=n_2+1}^{m} (\lambda_i + a_{n1} \lambda_n) & (n_2 \leq n \leq m)
\end{cases}
\]

\[
Cov(r_{L-i}, g_L) = \begin{cases} 
E(r_{L+1} g_L) - E(r)E(g) & (i=-1) \\
0 & \text{otherwise}
\end{cases}
\]

where

\[
E(r_{L+1} g_L) = c \left[ \sum_{n=1}^{n_2} \tilde{P}^+_{n}(n_1)(\sigma_n + \eta_n)(\rho_n + \mu_n) + \sigma \sum_{n=n_2+1}^{m} \tilde{P}^+_{n}(n_1)(\rho_n + \eta_n + \mu_{n_1}) \right]
\]

\[
Cov(r_{L-i}, r_L) = Cov(g_{L-i}, g_L) = 0 & (i \geq 1)
\]

where $n_1$, $n_2$ ($\leq n \leq m$) are given by (4.65-6),

\[
\tilde{P}^+_{n}(n_1) (n_1 \leq n \leq m)
\]

and $\mu_{n_1}$, $\mu_{n_2}$ are given by (5.99).

Proof

According to the definition of Class 4 given in Lemma 5.17, $\leq n \leq m$.

Thus an effective red period for stream $j$ can be initiated by a recall of stage 1 while stage $n$ ($n_1 \leq n \leq n_2$) is running and will otherwise be
initiated at the end of stage $n_2$. The effective red period will end at a
time $c(1-a_{n_1-1,0,j}^n) \lambda_{n_1-1,0}$ after the next call of stage $n_1$. Accordingly,
in the first case, the conditional moments of the duration of the effective
red period are given by
\[
\begin{align*}
E(r|n) &= c(\rho_n + \mu_{n_1}^n) \\
E(r^2|n) &= c^2(\rho_n^2 + 2\rho_n\mu_{n_1}^n + \mu_{2n_1}^n)
\end{align*}
\]
\[ (n_1 \leq n \leq n_2) \] (5.105)

In the second case, the conditional moments are given by
\[
\begin{align*}
E(r|n) &= c(\rho_n + \eta_{2n}^n) \\
E(r^2|n) &= c^2[\rho_n^2 + 2\rho_n\eta_{2n}^n + 2(\rho_n + \eta_{n_2}^n)\mu_{n_1}^n + \mu_{2n_1}^n]
\end{align*}
\]
\[ (n_2 \leq n \leq m) \] (5.106)

where $\eta_{2n}^n$ are given by (4.65-6),
\[
\eta_n^m \text{ and } \mu_{2n}^m \text{ are given by (5.100)}
\]
and $\rho_n$ are given in (5.102).

Since there must be at least $n$ stages in a cycle in order for an
effective red period to be initiated during it, the conditioning
probabilities for (5.105-6) are $f_{n_1}^m(n_1)$ ($n_1 \leq n \leq m$) given by (5.99), so
(5.102) is established.

Now consider an effective green period for stream $j$. This is always
initiated when stage $n_1$ is called and will be terminated by a priority
recall of stage 1 if this occurs before the end of stage $n_2$ when it will
otherwise be terminated. In the first case, if stage 1 is recalled while
stage $n$ ($n_1 \leq n \leq n_2$) is running, then the conditional moments of the
duration of the effective green period are given by
\[
\begin{align*}
E(g|n) &= c(\sigma_n^m + \eta_n^m) \\
E(g^2|n) &= c^2(\sigma_n^m + 2\sigma_n^m \eta_n^m + \eta_{2n}^m)
\end{align*}
\]
\[ (n_1 \leq n \leq n_2) \]

where \( n_n \), \( n_{2n} \) \( (n_1 \leq n \leq n_2) \) are given by (4.65-6) and \( \sigma_n \) \( (n_1 \leq n \leq n_2) \) are given in (5.103). In the second case, the duration of the effective green period is conditionally equal to \( \sigma_m \). The appropriate conditioning probabilities are \( \tilde{P}_n^{\dagger}(n) \) \( (n_1 \leq n \leq n_2) \) and \( \sum_{n=n_2+1}^{m} \tilde{P}_n^{\dagger}(n) \), so (5.103) is established.

Finally, the durations of all distinct effective red and green periods are mutually independent except for an effective green period and the effective red period which follows on immediately. This exception arises because in cases where an effective green period is terminated by a transition to stage 1, the next effective red period is initiated by the same transition. Arguing as in the proof of theorem 5.21, (5.104) is established.

5.3.3 Priority by extension and recall with inhibition.

In this sub-section, explicit formulae are derived for the statistics (5.84) for streams in Classes 1, 2 and 4 when priority by extension and recall with inhibition is implemented. As before, the result of theorem 5.19 shows that these formulae provide sufficient information to find the statistics (5.84) for an arbitrary stream in any class.

The analysis presented in sub-section 5.3.2 was simplified by the independence of events during distinct cycles. This enabled the distributions of the durations of stages in each cycle to be treated as being independent of those in other cycles. When the inhibition rule is used, this property no longer holds: there are two distinct distributions for the duration of each stage depending on whether or not inhibition is in effect at the start of a cycle. Since a cycle during which inhibition is in effect is always followed by one during which it is not, some serial correlations are introduced by the inhibition rule. The first results
presented in this sub-section investigate the way in which information concerning the state of inhibition is propagated from one cycle to the next. Let

\[ \tilde{p}_{n}^{+a}(N) \] be the probability that stage \( n \) is interrupted by a priority recall of stage 1 given that there are at least \( N \) stages in the cycle and that inhibition is not in effect \((1 \leq N \leq n \leq m)\). 

I denote the state 'inhibition is in effect'

\[ \tilde{p}_{n}^{(n)}(N) \] denote the state 'inhibition is not in effect'

\[ \tilde{p}_{s_{1}s_{2}}^{(n)}(N) \] be the probability that state \( s_{2} \) obtains when stage \( N \) is called for the \( n \)th time after a call of that stage which occurred while state \( s_{1} \) obtained \((k < N \leq m; n \geq 1; s_{1}, s_{2} \in \{U, I\})\).

**Lemma 5.24**

Suppose that priority by extension and recall with inhibition is provided for buses arriving at a mean rate of \( \beta / c \). Then the conditional probabilities \( \tilde{p}_{s_{1}s_{2}}^{(n)}(N) \) can be calculated from the recursive formula

\[
\tilde{p}_{s_{1}s_{2}}^{(n + 1)}(N) = \tilde{p}_{s_{1}U}^{(1)}(N) \tilde{p}_{s_{1}s_{2}}^{(n)}(N) + \tilde{p}_{s_{1}I}^{(1)}(N) \tilde{p}_{s_{1}s_{2}}^{(n)}(N) \quad (k < N \leq m; n \geq 1; s_{1}, s_{2} \in \{U, I\})
\]

(5.107)

using the expressions

\[
\tilde{p}_{s_{1}s_{2}}^{(1)}(N) = \begin{cases}
\left[ \tilde{p}_{m}^{+}(N) - \tilde{p}_{m}^{+a}(N) \right] \sum_{i=1}^{m} p_{i}^{+} & (s_{1} = U, s_{2} = U) \\
\sum_{i=1}^{m} \tilde{p}_{m}^{+}(a)(N) + \left[ \tilde{p}_{m}^{+}(N) - \tilde{p}_{m}^{+a}(N) \right] \sum_{i=k+1}^{N-1} p_{i}^{+} & (s_{1} = U, s_{2} = I) \\
\sum_{i=1}^{m} p_{i}^{+} & (s_{1} = I, s_{2} = U) \\
\sum_{i=k+1}^{N-1} p_{i}^{+} & (s_{1} = I, s_{2} = I)
\end{cases}
\]

(5.108)

where

\[ \tilde{p}_{n}^{+a}(N) = \tilde{p}_{0}^{+a}(N) \] \((k < N \leq m)\) are given by (4.77)

and \( \tilde{p}_{i}^{+}(N) \) \((k < i \leq m)\) are given by (5.99).
Proof
The conditioning event for the probabilities $\tilde{P}_{+}^{a}(N)$ is that there are at least $N$ stages in a cycle during which inhibition is not in effect. Dividing the probabilities $\tilde{P}_{+}^{a}(N;i \leq m)$ given in (4.67) by the conditional probability that there are at least $N$ stages in a cycle given that inhibition is not in effect and proceeding as in the proof of lemma 5.22 gives (5.109).

Suppose that inhibition is not in effect when stage $N$ ($k < n < m$) is called. Inhibition will not be in effect at the next call of that stage if there are $m$ stages in the current cycle with no recall being granted and there are at least $N$ stages in the next cycle. The first of these events has probability $\tilde{P}_{+}^{a}(N)$ and the conditional probability that the second event occurs given that the first has is $\sum_{i=N}^{m} P_{+}^{b}$ where $P_{+}^{b}$ ($k \leq i \leq m$) are given by (4.77). Multiplying these two probabilities gives the probability that they both occur, so the formula for $\tilde{P}_{+}^{(1)}(N)$ is established. Inhibition will be in effect at the next call of stage $N$ if either the next occurrence of stage 1 results from a recall or a recall is granted during the next cycle before stage $N$ is called. These events are mutually exclusive and have probability $\sum_{i=N}^{m} \tilde{P}_{+}^{a}(N)$ and $\sum_{i=k+1}^{m} P_{+}^{b}$ respectively so the formula for $\tilde{P}_{+}^{(1)}(N)$ is established.

Now suppose that inhibition is in effect when stage $N$ ($k \leq n \leq m$) is called. Inhibition will not be in effect at the next call of stage $N$ provided that there are at least $N$ stages in the next cycle. Thus the formula for $\tilde{P}_{+}^{(1)}(N)$ is established. Finally, inhibition will be in effect at the next call of stage $N$ if there are fewer than $N$ stages in the next cycle as a result of a recall being granted, so the formula for $\tilde{P}_{+}^{(1)}(N)$ is established and the proof of (5.108) is complete.
Consider the state at the next call of stage $N$ ($k < n < m$) after one at which state $S_1$ obtained: state $S$ obtains with probability \( \tilde{p}^{(1)}_{S_1S}(N) \) ($S_1, S \in \{U, I\}$). The probability of a transition from state $S$ at this intermediate call of stage $N$ to state $S_2$ in calls later is conditionally independent of $S_1$ and is given by \( \tilde{p}^{(n)}_{SS_2}(N) \) ($S, S_2 \in \{U, I\}$). Since the intermediate state $S$ is not specified, the probability \( \tilde{p}^{(n+1)}_{S_1S_2}(N) \) is found by summing the probabilities \( \tilde{p}^{(1)}_{S_1S}(N) \tilde{p}^{(n)}_{SS_2}(N) \) ($S \in \{U, I\}$) over both possible values of the intermediate state $S$. Thus (5.107) is established.

The formula given for the conditional probabilities \( \tilde{p}^{+a}_{m}(N) \) ($k \leq n < m$) in (5.109) are identical to those given for \( \tilde{p}^{+}_{n}(N) \) ($k \leq n < m$) in (5.99). This is because the events represented are identical - if $N_k \leq m$, then there can be $n$ stages in the cycle only if stage $n$ is interrupted by a priority recall of stage 1. The formula for \( \tilde{p}^{+a}_{m}(N) \) differs from that for \( \tilde{p}^{+}_{m}(N) \) because the former represents only one of the two events represented by the latter.

The probabilities \( \tilde{p}^{(n)}_{S_1S_2}(N) \) correspond to the transition probabilities of a Markov chain with state space \( \{U, I\} \). This observation would permit the deduction of (5.107) directly but would not in itself be particularly useful in the remainder of the analysis.

The next result shows how these transition probabilities can be used to find the statistics (5.84) from conditional moments of the durations of effective red and green times for any stream in Classes 1, 2 or 4. The conditioning events used are the states of the controller during the stage which is running when the call is made which initiates the period under consideration. These events are denoted by a pair \( S_1S_2 \) to indicate that $S_1$ obtained immediately before the start of the current period and state
obtained immediately before the start of the next such period. Thus, for example, \( E(r|UI) \) denotes the conditional mean duration of an effective red period which is initiated by the call of a stage when inhibition is not in effect given that inhibition is in effect when the next such effective red period is initiated. When an expectation is conditioned by a single event, as in \( E(X|U) \), it represents the expected value of the random variable over all occurrences which are initiated by a call of a stage when that conditioning state obtains.

Lemma 5.25

Suppose that priority by extension and recall with inhibition is provided for buses arriving at a mean rate of \( \beta/c \). Then for any stream in Class 1, 2 or 4, the statistics (5.84) can be calculated from the conditional expectations

\[
E(X|S_{1,2}) (X \in \{r, r^2, g, g^2, r_n, g_n, r_{n-1}, g_{n-1}, r_1\}; S_1, S_2 \in \{U, I\})
\]  

(5.110)

using the formulae

\[
E(X|S) = E(1)_{SU} (n^*) E(X|SU) + E(1)_{SI} (n^*) E(X|SI) \quad \{X \in \{r, r^2, g, g^2, r_n, g_n, r_{n-1}, g_{n-1}, r_1\}; S \in \{U, I\})
\]  

(5.111)

\[
E(X) = \frac{(1-P^0_1) \sum_{j=n}^{m} P^+_j E(X|U) + P^0_1 E(X|I)}{(1-P^0_1) \sum_{j=n}^{m} P^+_j + P^0_1} \quad \{X \in \{r, r^2, g, g^2, r_n, g_n, r_{n-1}, g_{n-1}, r_1\}
\]  

(5.112)

\[
Cov(X_{n}, Y) = E(X_{n} Y) - E(X)E(Y) \quad \{X, Y \in \{r, g\}
\]  

(5.113)

\[\begin{align*}
\text{where} & \\
E(X_{n-n \rightarrow n}) = & \left[ (1-P^0_1) \sum_{j=n}^{m} P^+_j \right] E(X|UU) + P^0_1 E(X|IU) \times \left[ \hat{P}^{(n-1)}_{UU} (n^*) E(X|U) + \hat{P}^{(n-1)}_{UI} (n^*) E(X|I) \right] \\
& + \left[ (1-P^0_1) \sum_{j=n}^{m} P^+_j \right] UU \times \left[ \hat{P}^{(n-1)}_{UU} (n^*) E(X|U) + \hat{P}^{(n-1)}_{UI} (n^*) E(X|I) \right] \\
& \times \left[ \frac{(1-P^0_1) \sum_{j=n}^{m} P^+_j + P^0_1}{(1-P^0_1) \sum_{j=n}^{m} P^+_j} \right] \quad \{n \geq 1, X \in \{r, g\}\}
\end{align*}
\]
\[
\begin{align*}
\text{Cov}(r_{n-n}, g_{n}) &= E(r_{n-n}g_{n}) - E(r)E(g) \quad (n \geq 1) \\
\text{Cov}(g_{n-n}r_{n}) &= E(g_{n-n}r_{n}) - E(r)E(g) \quad (n \geq 2)
\end{align*}
\]

where

\[
E(X_{n-n}Y_{n}) = \left\{ \begin{array}{ll}
\left[ (1-P_{I}) \sum_{j=n^{*}}^{m} \tilde{P}_{j}^{(1)}(n^{*})E(X|U) + P_{I} \tilde{P}_{I}^{(1)}(n^{*})E(X|I) \right] \times \\
\left[ \tilde{P}_{I}^{(n-1)}(n^{*})E(Y|U) + P_{I} \tilde{P}_{I}^{(n-1)}(n^{*})E(Y|I) \right] + \\
\left[ (1-P_{I}) \sum_{j=n^{*}}^{m} \tilde{P}_{j}^{(n)}(n^{*})E(X|U) + P_{I} \tilde{P}_{I}^{(n)}(n^{*})E(X|I) \right] \times \\
\left[ \tilde{P}_{I}^{(n-1)}(n^{*})E(Y|U) + P_{I} \tilde{P}_{I}^{(n-1)}(n^{*})E(Y|I) \right]/
\end{array} \right.
\]

\[
\begin{align*}
\begin{cases}
\tilde{P}_{j}^{(n)} & \text{if the stream is in Class 1 or Class 2} \\
\tilde{P}_{j}^{(n)} & \text{if the stream is in Class 4}
\end{cases}
\end{align*}
\]

\[
P_{I} \quad \text{is given by } (4.68), \\
\tilde{P}_{j}^{(n)} \quad (k<j<m) \quad \text{are given by } (4.77)
\]

and \( \tilde{P}_{S_{1}S_{2}}^{(n)} \) \( (k<n^{*}; n \geq 1; S_{1}, S_{2} \in \{U, I\}) \) are given by (5.107-8).

**Proof**

The first part of this proof establishes that for streams in each of the classes to be considered, if conditional distributions are available for the durations of the effective red and green periods given the state of the controller when stage \( n_{1} \) is called, then the appropriate distribution to use can be determined from the state of the controller when stage \( n^{*} \) is called. Each of the three classes of streams is considered in turn.

Consider a stream in Class 1 for which \( n_{1} = 1 \). The state of the controller remains unchanged from the call of stage 2 until after the next call of stage 1. Thus the state will be the same at the time stage \( k+1 \) is called as it was when stage \( n_{2} + 1 \) was called, initiating an effective red period. Furthermore, the state will still be the same when stage 1 is next called, initiating the next effective green period. If the stream is in Class 1 and \( n_{1} > 1 \), then the duration of the effective green period is independent of the state of the controller.
Now consider a stream in Class 2. The state of the controller will remain unchanged from the call of stage \( n \) to the next call of stage 1. Thus the state will be the same at the time stage \( k+1 \) is called as it was when stage \( n_1 \) was called and as it will be when whichever of stages \( n_2+1 \) and 1 is called initiating the next effective red period. Thus for streams in Classes 1 and 2 the appropriate conditional distributions for the durations of the effective red and green periods can be determined from the state of the controller when stage \( k+1 \) is called.

Now consider a stream in Class 4. An effective green period will be initiated in a cycle only if there are at least \( n_1 \) stages in that cycle. If an effective green period is initiated, then it will be terminated by the call of whichever of stages \( n_2+1 \) and 1 occurs first. The state of the controller when this occurs will be the same as it was when stage \( n_1 \) was called, so the appropriate conditional distributions for the durations of the effective red and green periods for any stream in Class 1, 2 or 4 can be determined from the state of the controller when stage \( n^* \) is called where \( n^* \) is given by (5.115).

If the controller is in state \( S_1 \) (\( S_1 \in (U,I) \)) when an effective red or green period is initiated, then it will be in state \( S_2 \) (\( S_2 \in (U,I) \)) with probability \( \widetilde{P}_{S_1,S_2}^{(1)}(n^*) \) when the next such period is initiated. The appropriate conditional expectation of the random variable \( X \) (\( X \in \{r,r^2,g,g^2,r_{N-1}/g_{N-1},r_N/g_N, r_{N-1}/g_N\} \)) is \( E(X|S_1,S_2) \); multiplying by the probability of the conditioning event and summing over both possibilities for \( S_2 \) yields (5.111).

Now consider the start of an arbitrary cycle. With probability \( P_I \), given by (4.68), inhibition is in effect so an effective red and green period will be initiated while the controller is in that state. With probability \( (1-P_I) \) inhibition is not in effect and with conditional
probability \( \sum_{j=1}^{m} P_j^+ \), where \( P_j^+ \) \((k,j,m)\) are given by (4.77) and \( n^k \) is given by (5.115), an effective red and an effective green period will be initiated while the controller is in this state. Thus the probability that inhibition is in effect when an effective red or green period is initiated is \( P / [P + (1 - P) \sum_{j=1}^{m} P_j^+] \) and the probability that it is not in effect is \( (1 - P) / [P + (1 - P) \sum_{j=1}^{m} P_j^+] \). Multiplying the conditional expectations \( E(X|S) \) \((S \in \{U,I\}, X \in \{r,g, g^2, r, g, g^2, r, g, g^2, r\})\) by the appropriate probabilities and summing over all possibilities for the state \( S \) gives (5.112).

Now consider the expected value of the product \( X_{n-n} X_n \) \((X \in \{r,g\}, n \geq 1)\). Suppose that the period numbered \( N-n \) is initiated while state \( S_1 \) obtains, that the next such period is initiated while state \( S_2 \) obtains and that the period numbers \( N \) is initiated while state \( S_3 \) obtains \((S_1, S_2, S_3 \in \{U,I\})\). The conditional expectation of the product is then \( E(X|S) \) \((S \in \{U,I\})\). The probability that \( S_1 = U \) is \( (1 - P) / [P + (1 - P) \sum_{j=1}^{m} P_j^+] \), the probability that \( S_1 = I \) is \( P / [P + (1 - P) \sum_{j=1}^{m} P_j^+] \), the conditional probabilities for state \( S_2 \) are given by \( \tilde{P}^{(1)} \) \((n^k)\) of (5.108) and the conditional probabilities for state \( S_3 \) are given by \( \tilde{P}^{(n-1)} \) \((n^k)\) of (5.107) with the additional values

\[
\tilde{P}^{(0)} \frac{n^k}{S_2 S_3} = \begin{cases} 1 & \text{if } S_2 = S_3 \\ 0 & \text{otherwise} \end{cases} \quad (k < n^k \leq m)
\]

Multiplying the conditional expectation of the product \( X_{n-n} X_n \) by the product of these probabilities, summing over all possible states \( S_1, S_2 \) and \( S_3 \), rearranging and using the standard formula for covariances gives (5.113).

The durations of the periods \( r_{N-n} \) and \( g_n \) \((n \geq 1)\) and \( g_{N-n} \) and \( r_n \) \((n \geq 2)\) are conditionally independent given the state of the controller at the appropriate times. Thus (5.114) follows from an argument similar to that used to establish (5.113).
In some cases the relationship between the durations of sequent effective red and green periods differs from that specified by the transition probabilities \( p^{(1)}_{S_1 S_2} \) and the corresponding conditional expectations. This difference can arise because of the way in which the first period is terminated and the second initiated. Thus the argument which was used to establish (5.113-4) cannot always be extended to find the expected value of the products \( E(g_{N-1} r_N) \) and \( E(r_N g_N) \). Instead these statistics are derived from (5.111) using conditional values to be established in the next results.

In some cases not all of the conditional expectations (5.110) are well-defined since some of the transitions are not possible. This is of no consequence since the associated transition probabilities are 0 so the conditional expectations contribute nothing to the unconditional expectations.

**Theorem 5.26**

Suppose that priority by extension and recall with inhibition is provided for buses arriving at a mean rate of \( \beta/c \). Then, by means of lemma 5.25, the statistics (5.04) can be deduced for any stream \( j \) in Class 1 from the formulae

\[
E(r|UU) = c \left[ \sum_{i=1}^{M_2} (1-a_{i0j})/\lambda_i + \sum_{i=1}^{M_2} (1-a_{10i})/\lambda_i + \sum_{i=1}^{M_2} (1-\delta_{ik2})/\lambda_i \right]
\]

\[
E(r|UI) = c \sum_{n=k+1}^{m} a_n/\sum_{i=k+1}^{M_2}
\]

\[
E(r|IU) = E(r|UU)
\]

\[
E(r^2|UU) = [E(r|UU)]^2 + c^2 (1-\delta_{kn2}) u_k
\]

\[
E(r^2|UI) = c^2 \sum_{n=k+1}^{m} a_n/\sum_{n=1}^{M_2} + 2c^2 a_n/\sum_{n=1}^{M_2} + (1-\delta_{kn2})/\sum_{i=k+1}^{M_2}
\]

\[
E(r^2|IU) = E(r^2|UU)
\]

(5.116)
\[
E(g | \{\{U \cup U\}|} = c \left[ \sum_{i=1}^{m} \lambda_{i} \lambda_{i} + \sum_{j=1}^{n_{2}} \lambda_{j} + \delta_{k-n_{2}} (\epsilon_{k} - \lambda_{k}) \right] \\
E(g | \{U \cup I\}) = c \sum_{n=k+1}^{m} \frac{p_{n}^{a}}{\sum_{i=k+1}^{m} p_{i}^{a}} \\
E(g | \{I \cup U\}) = E(g | \{U \cup U\}) \\
E(g^{2} | \{U \cup U\}) = \left[ E(g | \{U \cup U\}) \right]^{2} + c^{2} \delta_{kn_{2}} u_{k} \\
E(g^{2} | \{U \cup I\}) = c^{2} \sum_{n=k+1}^{m} \frac{p_{n}^{a}(\sigma_{n} + \delta_{kn_{2}} u_{k})}{\sum_{i=k+1}^{m} p_{i}^{a}} \\
E(g^{2} | \{I \cup U\}) = E(g^{2} | \{U \cup U\}) \\
\]

(5.117)

\[
E(g_{n-1}^{(1)} \{U \cup U\}) = E(g | \{U \cup U\}) \left[ \underline{E}^{(1)}(U \cup U) + \underline{E}^{(1)}(I \cup I) \right] \\
E(g_{n-1}^{(1)} \{U \cup I\}) = E(g | \{U \cup I\}) E(r | \{U \cup I\}) \\
E(g_{n-1}^{(1)} \{I \cup U\}) = E(g_{n-1}^{(1)} \{U \cup I\}) \\
E(g_{n-1}^{(1)} \{U \cup I\} | S_{1}, S_{2}) = E(g | \{U \cup U\}) \left[ \underline{E}^{(1)}(U \cup U) + \underline{E}^{(1)}(S_{1}, S_{2}) \right] \\
E(r_{n} g_{n} | \{U \cup U\}) = E(r | \{U \cup U\}) E(g | \{U \cup U\}) \\
E(r_{n} g_{n} | \{U \cup I\}) = c^{2} \sum_{n=k+1}^{m} \frac{p_{n}^{a}(\rho_{n} + \sigma_{n})}{\sum_{i=k+1}^{m} p_{i}^{a}} \\
E(r_{n} g_{n} | \{I \cup U\}) = E(r_{n} g_{n} | \{U \cup U\}) \\
E(r_{n} g_{n} | S_{1}, S_{2}) = E(r | S_{1}, S_{2}) E(g | \{U \cup U\}) \\
\]

(5.118)

(5.119)

where

\[
\begin{align*}
\rho_{n} &= \sum_{i=n, j+1}^{n-1} \lambda_{i} - (1-\delta_{kn_{2}}) (\epsilon_{k} - \lambda_{k}) + \sum_{i=n}^{n-1} (1-a_{i}) \lambda_{i} + (1-a_{j}) \lambda_{j} + (1-a_{n} \lambda_{n} + (1-a_{10} \lambda_{10}) \\
\sigma_{n} &= a_{n} \lambda_{n} + \sum_{i=1}^{n-1} \lambda_{i} + \sum_{j=1}^{n-1} a_{j} \lambda_{j} + \delta_{k-n_{2}} (\epsilon_{k} - \lambda_{k}) \\
\end{align*}
\]

(5.120)

\(k < n \leq m\)

\(\epsilon_{k}\) and \(u_{k}\) are given by (4.52),

\(n_{1}^{a}, n_{2i}^{a} (k < i \leq m)\) are given by (4.64)

and \(p_{n}^{a} (k < n \leq m)\) are given by (4.67).
Proof

As mentioned in the proof of lemma 5.25, for streams in Class 1 the state of the controller at the start of an effective red and an effective green period is the same as it is at the call of stage \( n^* = k + 1 \). Thus from (5.108), \( \tilde{P}_{II}^{(1)}(n^*) = 0 \). From the result of lemma 5.25 together with this observation, the conditional expectations (5.116-9) provide sufficient information to find the statistics (5.84). All that remains is to establish the formulae.

Consider the duration of an effective red period for stream \( j \) in Class 1. If the transition is from state \( U \) to state \( U \), then no recall is granted before the next effective red period is initiated. Thus the effective red period starts after the call of stage \( n_2 + 1 \), runs throughout stage \( m \) and then continues until after the next call of stage \( n \). The only possible source of variability here is from priority extensions to stage \( k \): this arises only if \( n_2 < k \). Thus the formulae for \( E(r|UU) \) and \( E(r^2|UU) \) are established. The duration of an effective red period during a transition from state \( I \) to state \( U \) has an identical distribution since, again, no recall can be granted. Finally, a transition from state \( U \) to state \( I \) occurs when a recall is granted during the effective red period. The conditional probability that this occurs while stage \( n \) is running is \( P_{III}^{m} \frac{m}{n} \frac{P_{III}^{1} \frac{1}{i=k+1}}{n} \) in which case the first two moments of the duration of the effective red period are \( c(n + n^*) \) and 

\[
c [\rho^n + 2\rho \eta^n + (1 - \delta_n) \eta^2] k_n \] 

where \( \delta_n \) is given in (4.52), \( \eta^n \), \( \eta_{2n} \), \( \rho^n \) and \( \rho_n \) \( (k\eta_n) \) are given by (4.64), (4.67) and (5.120) respectively. The formulae for \( E(r|UI) \) and \( E(r^2|UI) \) are derived by multiplying each of these conditional expectations by the appropriate conditioning probability and summing over \( n \) \( (k\eta_n) \). Thus (5.116) is established.
Now consider the duration of an effective green period for stream \( j \) in Class 1 for which \( n_1 = 1 \). If the transition is from state \( U \) to state \( U \), then stage 1 is called when the time allocated for stage \( m \) has elapsed. The only possible source of variability here is from priority extensions to stage \( k \); this arises only if \( n_2 = k \). Thus the formulae for \( E(g^1|UU) \) and \( E(g^2|UU) \) are established for this case. The duration of an effective green period during a transition from state \( I \) to state \( U \) has an identical distribution since again it cannot be initiated by a priority recall of stage 1. Finally, if the transition is from state \( U \) to state \( I \), then the period must be initiated by a priority recall of stage 1. The conditional probability that this occurs during stage \( n \) is again \( \prod_{n_1=2}^{m} P_{n_1}^{a} (k \leq m) \) in which case the first two moments of the duration of the effective green period are \( \sigma_n^{2} \) and \( \sigma_n^{2} + \delta_{kn_2}^{a} k \) where \( \sigma_n^{2} \) (\( k \leq m \)) are given by (5.120). Proceeding as for the moments of the effective red periods yields the expressions for \( E(g^1|UI) \) and \( E(g^2|UI) \). If \( n_1 > 1 \), then the only possible source of variation in the duration of the effective green periods is from priority extensions to stage \( k \) irrespective of the type of the transition. Each of the formulae for \( E(g^1|S_1S_2) \) and \( E(g^2|S_1S_2) \) \( (S_1, S_2 \in \{U,I\}) \) reduces to the same form which is appropriate for this case. Thus (5.117) is established for all streams in Class 1.

Since the transition from an effective green to an effective red period occurs before the start of stage \( k+1 \), it is identical in all cases. Thus the conditional expectations of the product \( E_{n-1} r_n \) can be found from the conditional expectations of the durations of the effective red and green periods. Consider a stream in Class 1 for which \( n_1 = 1 \). If an effective green period is initiated while state \( S_1 \) obtains and the next effective green period is initiated while state \( S_2 \) obtains \( (S_1, S_2 \in \{U,I\}) \), then the intervening effective red period will be
initiated while state $S_2$ obtains. The conditional expectation of the duration of that effective red period is then given by $E(r|S_2)$ in (5.111). Since $\hat{F}^{(1)}_{II}(k+1)=0$, (5.118) is established for this case. Now consider a stream in Class 1 for which $n > 1$. The distribution of the duration of the effective green period is independent of the state of the controller and of the duration of the next effective red period, so $\text{Cov}(g_{N-1}, r_N)=0$. Since $E(g|S_1, S_2)$ has the value $E(g)$ for all $S_1$ and $S_2$, (5.118), when substituted into (5.113), gives $E(g_{N-1} r_N)=E(g)E(r)$ and $\text{Cov}(g_{N-1}, r_N)=0$ as required. Thus (5.118) is established for all streams in Class 1.

The transition from an effective red period to the next effective green period occurs when stage 1 is called if $n_1=1$ and before the call of stage $k+1$ if $n_1>1$. Consider a stream in Class 1 for which $n_1=1$. If an effective red period is initiated while state $S_1$ obtains, then the next effective green period will also be initiated while state $S$ obtains. If the next effective red period is initiated while state $U$ obtains then the intervening effective green period is initiated by a transition from stage $m$ to stage 1 without a priority recall being granted. Thus $E(r g, SU)=E(r|SU)E(g|SU)$ ($n_1=1; S \in \{U, I\}$). However, if an effective red period is initiated while state $U$ obtains and the next effective red period is initiated while state $I$ obtains, then the intervening effective green period is initiated by a priority recall of stage 1. The conditional probability that this occurs while stage $n$ is running ($k<n=m$) is $\frac{p_a n}{\sum_{1}^{m} p_a i}$ and the conditional expectations of the durations of the effective red and green periods are $c(n_1 + n_a)$ and $c_1 n$ respectively. Multiplying the conditional expectations by the appropriate conditioning probabilities and summing over $n$ gives (5.119) for $n_1=1$. If $n_1>1$, then the durations of the effective green periods are independent of the state of the controller and of the duration of the preceding effective
green period. Following the argument used to establish the corresponding
formula for \( E(g_{N-1}|r_N) \) yields the formula (5.119) for this case. Thus
(5.119) is established for all streams in Class I.

Theorem 5.27
Suppose that priority by extension and recall with inhibition is provided
for buses arriving at a mean rate of \( \beta/c \). Then the statistics (5.84) can
be deduced for any stream \( j \) in Class 2 from the formulae

\[
E(x|SU) = c\left[ \sum_{i=1}^{m} (1-a_{10j})^{\lambda_{10}^i} \sum_{n=0}^{\infty} \frac{\lambda_{1}^i}{n!} \prod_{i=1}^{n-1} \left[ \frac{1}{1-a_{10j}} \right] \right] \quad (S \in \{U,I\})
\]

\[
E(x|UI) = c\left[ \sum_{n=k+1}^{m} \frac{P_a^{n}}{n^n} \sum_{n=2}^{n_2} \frac{P_a^{n} (\rho_n + \gamma_n)}{n^n} \right] \quad (S \in \{U,I\})
\]

\[
E(x^2|SU) = [E(x|SU)]^2 \quad (S \in \{U,I\})
\]

\[
E(x^2|UI) = c^2 \left[ \sum_{n=k+1}^{m} \frac{P_a^{n}}{n^n} \sum_{n=2}^{n_2} \frac{P_a^{n} (\rho_n + \gamma_n)}{n^n} \right] \quad (S \in \{U,I\})
\]

\[
E(g|SU) = c\left[ \sum_{i=1}^{m} (1-a_{10j})^{\lambda_{10}^i} \sum_{n=0}^{\infty} \frac{\lambda_{i}^i}{n!} \prod_{i=1}^{n-1} \left[ \frac{1}{1-a_{10j}} \right] \right] \quad (S \in \{U,I\})
\]

\[
E(g|UI) = c\left[ \sum_{n=k+1}^{m} \frac{P_a^{n} (\sigma_n + \gamma_n)}{n^n} \sum_{n=2}^{n_2} \frac{P_a^{n} \rho_n}{n^n} \right] \quad (S \in \{U,I\})
\]

\[
E(g^2|SU) = [E(g|SU)]^2 + c^2 \nu_k \quad (S \in \{U,I\})
\]

\[
E(g^2|UI) = c^2 \left[ \sum_{n=k+1}^{m} \frac{P_a^{n} (\sigma_n + \gamma_n)}{n^n} \sum_{n=2}^{n_2} \frac{P_a^{n} \rho_n}{n^n} \right] \quad (S \in \{U,I\})
\]

\[
E(q_{N-1}|r_N|SU) = E(q|SU)E(x|SU) \quad (S \in \{U,I\})
\]

\[
E(q_{N-1}|r_N|UI) = c^2 \left[ \sum_{n=k+1}^{m} \frac{P_a^{n} (\sigma_n + \gamma_n)}{n^n} \sum_{n=2}^{n_2} \frac{P_a^{n} \rho_n}{n^n} \right] \quad (S \in \{U,I\})
\]

\[
E(x|a_N|S_1S_2) = E(x|S_1S_2) \left[ \bar{b}_{1}^{(1)} E(g|S_2U) + \bar{b}_{2}^{(1)} E(g|S_2I) \right] \quad (S_1, S_2 \in \{U,I\})
\]

(5.124)
where
\[ \rho_n = \begin{cases} 
(1-a_{1j}) \lambda_{n1} + \sum_{i=1}^{n-1} [\lambda_i + (1-a_{i0j}) \lambda_{10}] & (k<n\leq n_2) \\
(1-a_{2j}) \lambda_{n2} + \sum_{i=n_2+1}^{n-1} [\lambda_i + (1-a_{i0j}) \lambda_{10}] & (n_2<n\leq m) 
\end{cases} \] (5.125)

and
\[ \sigma_n = \begin{cases} 
a_{n1-1,0} \lambda_{n1-1,0} + \sum_{i=n1}^{n1-1} (\lambda_i + \lambda_{10}) + a_{1j} \lambda_{n1} & (k<n\leq n_2) \\
a_{n1-1,0} \lambda_{n1-1,0} + \sum_{i=n1}^{n1-1} (\lambda_i + a_{i0j} \lambda_{10}) & (n_2<n\leq m) 
\end{cases} \] (5.126)

\( \varepsilon_k \) and \( \psi_k \) are given by (4.52),
\( \eta_{ii}^a \), \( \eta_{2i}^a \) (\( k<i\leq m \)) are given by (4.64),
and \( \rho_n^a \) (\( k<n\leq m \)) are given by (4.67).

**Proof**
As with streams in Class 1, \( n^s=k+1 \) so \( \tilde{p}^{(1)}_{II}(n^s)=0 \) for streams in Class 2. From the result of lemma 5.25, the conditional expectations (5.121-4) provide sufficient information to find the statistics (5.84).

Consider the duration of an effective red period for stream \( j \) which is initiated while state \( S \) obtains (\( S\in\{U,I\} \)) given that state \( U \) obtains when the next such period is initiated. In this case no recall can be granted before stage 1 next occurs since that would cause the next effective red period to be initiated while state \( I \) obtains. Thus the duration of the effective red period is exactly
\[ \sum_{i=n1}^{n1-1} (1-a_{i0j}) \lambda_{i0} \sum_{i=n2}^{n1} \lambda_i + \sum_{i=1}^{n1-1} [(1-a_{i0j}) \lambda_{10}] \] so the formulae for \( \mathbb{E}(r|SU) \) and \( \mathbb{E}(r^2|SU) \) are established. Now consider the duration of an effective red period which is initiated while state \( U \) obtains given that the next such period is initiated while state \( I \) obtains. For this to occur, the intervening occurrence of stage 1 must result from a priority recall. If this recall is granted while stage \( n \) (\( n<n_2 \)) is running, then the duration of the effective red period is exactly \( \sigma_n^a \) where \( \rho_n \) is given by (5.125). If the recall is granted while stage \( n \) (\( n_2<n<m \)) is
running, then the first two conditional moments of the duration of the
effective red period are $c(\rho_{n+1}^a)$ and $c^2(\rho_{n+2}\eta_{n+1}^a)$ where $\eta_{n+1}^a$, $\eta_{n+2}^a$
(knsm) are given by (4.64). The conditioning probabilities are
$p_{n_i}^a/\sum_{i=0}^n p_{n_i}^a$ (knsm) so (5.121) is established.

Now consider the duration of an effective green period for stream $j$.
If the next effective green period is initiated while state $U$ obtains,
then the current effective green period cannot be curtailed by a priority
recall of stage 1. Thus the first two conditional moments are
$c[\sum_{i=1}^{\eta_{n+1}^a} a_{10j}^i 10 + \sum_{i=1}^{\eta_{n+2}^a} \lambda_i (1-c_{1-k}^i)]$ and $c^2\{[\sum_{i=1}^{\eta_{n+1}^a} a_{10j}^i 10 + \sum_{i=1}^{\eta_{n+2}^a} \lambda_i (1-c_{1-k}^i)]^2 + u_{n+1}^a\}.$
Now suppose that the current effective green period is initiated while
state $U$ obtains and that the next one is initiated while state $I$
obtains. As before, this means that the intervening occurrence of stage 1
results from a priority recall. If this is granted while stage $n$ ($k<n\leq n_{m+1}^a$)
is running, then the conditional moments of the duration of the effective
green period are $c(\sigma_{n+1}^a)$ and $c^2(\sigma_{n+2}^2 + 2\sigma_{n+2}^a \eta_{n+2}^a + u_{n+2}^a)$ where $u_{n+2}^a$ is
given in (4.52) and $\sigma_{n}$ (knsm) are given by (5.126). If this is granted while
stage $n$ is running ($n_2^m<n\leq n_{m+1}^a$), then the effective green period is not
curtailed so the conditional moments of its duration are the same as for
the transitions from $S$ to $U$ ($S\epsilon\{U,I\}$). Multiplying the conditional
expectations by the appropriate conditioning probabilities and summing over
$n$ (knsm) gives (5.122).

If an effective green period is initiated while state $S_1$ obtains and
the next one is initiated while state $S_2$ obtains, then the intervening
effective red period will be initiated while state $S_1$ obtains and the
next one will be initiated while state $S_2$ obtains. In the case where
$S_2=U$, the durations of the effective red and green periods are
conditionally independent so $E(g_{n-1}^{r_m}|SU)=E(g|SU)E(r|SU)$ ($S\epsilon\{U,I\}$). If
$S_1=U$ and $S_2=I$ then stage 1 must be recalled in the interim. If the
recall is granted while stage $n$ is running (knsm), then the
conditional expectation of the product \( g_{N-1}^r N \) is \( c^2 (\rho_n + \eta_n)^{n_1} \rho_n \); if the
recall is granted while stage \( n \) is running \( (n_2 < n < m) \), then the
conditional expectation is \( c \sigma_m (\rho_n + \eta_n)^{n_1} \). Multiplying the conditional
expectations by the appropriate conditioning probabilities and summing over
\( n \) \((k \leq n \leq m)\) gives (5.123).

If an effective red period is initiated while state \( S_1 \) obtains and
the next one is initiated while state \( S_2 \) obtains, then the intervening
effective green period will be initiated while state \( S_2 \) obtains. Once
these states have been specified, since the transition from an effective
green period to the next effective red period always occurs when stage
\( n_2 + 1 \leq k + 1 \) is called, the duration of the effective red and green periods
are conditionally independent. Thus (5.124) is established. []

**Theorem 5.28**

Suppose that priority by extension and recall with inhibition is provided
for buses arriving at a mean rate of \( \beta / \alpha \). Then the statistics (5.84) can
be deduced for any stream in Class 4 from the formulae

\[
\begin{align*}
E(r \mid SU) &= c \rho_0^m (S \in \{U, I\}) \\
E(r \mid UI) &= c \left( \sum_{n_1} p_1^m \rho_0^m + \sum_{n_2} p_1^m (\rho_0 + \eta_0)^{n_1} \right) / \\
&\quad \left[ \sum_{n_1} p_1^m (\rho_0 + \eta_0)^{n_1} \right] \\
E(r \mid II) &= c \sum_{n_1} p_1^m (\rho_0 + \eta_0)^{n_1} / \\
&\quad \sum_{n_1} p_1^m (\rho_0 + \eta_0)^{n_1} \\
E(r^2 \mid SU) &= c^2 (\rho_0^2 + \eta_0) \quad (S \in \{U, I\}) \\
E(r^2 \mid UI) &= c^2 \left( \sum_{n_1} p_1^m (\rho_0^2 + \eta_0) + \sum_{n_2} p_1^m (\rho_0^2 + 2 \rho_0 \eta_0 + \eta_0^2) \right) / \\
&\quad \left[ \sum_{n_1} p_1^m (\rho_0 + \eta_0)^{n_1} \right] \\
E(r^2 \mid II) &= c^2 \sum_{n_1} p_1^m (\rho_0^2 + 2 \rho_0 \eta_0 + \eta_0^2 + 2 \eta_0^2) / \\
&\quad \sum_{n_1} p_1^m (\rho_0 + \eta_0)^{n_1}
\end{align*}
\]

(5.127)
\[
E(g|UU) = c \sigma_n^{n_1-1} \\
E(g|UI) = c \left\{ \sum_{n=1}^{n_2} p_n^a (n+1)^2 + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} / \left\{ \sum_{i=1}^{n_1-1} p_i^a + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} \\
E(g|IS) = c \sigma_n^{n_1-1} \\
E(g^2|UU) = c^2 \sigma_n^{n_1-1} \\
E(g^2|UI) = c^2 \left\{ \sum_{n=1}^{n_2} p_n^a (n+1)^2 + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} / \left\{ \sum_{i=1}^{n_1-1} p_i^a + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} \\
E(g^2|IS) = c^2 \sigma_n^{n_1-1} \\
E(g_{N-1}|U) = c \sigma_n^{n_1-1} \\
E(g_{N-1}|U) = c \left\{ \sum_{n=1}^{n_2} p_n^a (n+1)^2 + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} / \left\{ \sum_{i=1}^{n_1-1} p_i^a + \sum_{m=n+2}^{n_1} p_m^a (1-a_{10j}) \right\} \\
E(g_{N-1}|I) = E(g|IS)E(r|IS) \\
E(r, q, S_1S_2) = E(r|S_1S_2)E(g|S_1S_2) + E(g|S_1S_2) \\
E(g|S_1S_2) \\
\text{where} \\
\sigma_n = \left\{ \begin{array}{l}
(1-a_{10j}) \lambda_n + \sum_{i=1}^{n_1-1} \left[ \lambda_i + (1-a_{10j}) \lambda_i \right] + (\varepsilon_k - \lambda_k) \\
(1-\varepsilon_k - \lambda_k)
\end{array} \right\} \\
\varepsilon_k \text{ and } \mu_k \text{ are given by (4.52),} \\
P_m \text{ is given by (4.62),} \\
\eta_n, \eta_n^a \text{ (}k \leq n \leq m) \text{ are given by (4.64) and } \\
P_n \text{ (}k \leq n \leq m) \text{ are given by (4.67).}
Proof

From the result of lemma 5.24, the conditional expectations (5.127-30) provide sufficient information to find the statistics (5.84).

Consider the duration of an effective red period which is initiated while state $S$ obtains ($S \in \{U, I\}$) given that state $U$ obtains when the next such period is initiated. In order for this to occur, no recall can be granted during the cycle in which the effective red period is initiated or before the call of stage $\gamma_1$ in the next cycle. Thus the first two conditional moments of the duration of the effective red period are

$$c_0 \rho_0^2$$ and $$c_0^2 (\rho_0^2 + \eta_k)$$ where $\rho_0$ and $\eta_k$ are given by (5.131) and (4.52) respectively. Thus the formulae for $E(r|SU)$ and $E(r^2|SU)$ ($S \in \{U, I\}$) are established.

Now consider the duration of an effective red period which is initiated while state $U$ obtains given that state $I$ obtains when the next such period is initiated. In order for this to occur, a recall must be granted either during the cycle in which the effective red period is initiated or before stage $n$ is called in the next cycle. With conditional probability $P_n^a/[\sum_{i=n_1}^{n_n} P_i^a + (P^a - P^0) \sum_{i=m}^{n_1-1} P_i^a]$ stage 1 is recalled while stage $n$ ($n_1 \leq n \leq n_2$) is running in the cycle during which the effective red period is initiated. If $n_1 \leq n < n_2$, then the first two conditional moments of the duration of the effective red period are $c_0 \rho_n^2$ and $c_0^2 (\rho_n^2 + \eta_k)$ where $\rho_n$ ($n_1 \leq n \leq n_2$) are given in (5.131). If $n_2 < n \leq m$, then these moments are $c_0 (\rho_n^2 + \eta_n^2)$ and $c_0^2 (\rho_n^2 + 2\rho_n \eta_n^2 + \eta_n^2)$ where $\eta_n^2$, $\eta_{2n}^2$ and $\rho_n$ ($n_2 \leq n \leq m$) are given by (4.64) and (5.131) respectively. With conditional probability $P_n^a/[\sum_{i=n_1}^{n_n} P_i^a + (P^a - P^0) \sum_{i=m}^{n_1-1} P_i^a]$, stage 1 is not recalled during the cycle in which the effective red period was initiated but is recalled during stage $n$ ($k < n < n_1$) of the next cycle. In this case the conditional moments are $c_0 (\rho_n^2 + \eta_n^2)$ and $c_0^2 (\rho_n^2 + 2\rho_n \eta_n^2 + \eta_n^2)$ where $\eta_n^2$, $\eta_{2n}^2$ and $\rho_n$ ($k < n < n_1$) are given by (4.64) and (5.131) respectively. Thus the formulae for $E(r|UI)$ and $E(r^2|UI)$ are established.
Finally, consider the duration of an effective red period which is initiated while state I obtains given that state I obtains when the next such period is initiated. This can occur only if a recall is granted before stage \( n \) is called during the cycle following the one in which the effective red period is initiated. With conditional probability \( \frac{P_a}{\sum_{i=k+1}^{n-1} P_a} \) the recall is granted while stage \( n \) (\( k \leq n \)) is running. In this case the conditional moments of the effective red period are \( c(\rho_n + \eta_n^a) \) and \( c^2(\rho_n^2 + 2\rho_n \eta_n^a + \eta_n^a + \eta_{2n}^a) \) where \( \eta_n^a, \eta_{2n}^a, P_n \) and \( \rho_n \) \( (k \leq n \leq n_1) \) are given by (4.64), (4.67) and (5.131) respectively. Thus the formulae for \( E(r|II) \) and \( E(r^2|II) \) are established and the proof of (5.127) is complete.

Now consider the duration of an effective green period. Suppose that state U obtains when each of this and the next such period is initiated. This can occur only if no recall is granted before the next call of stage \( n_1 \) so the effective green period will certainly be uninterrupted. Thus it will have a duration of exactly \( \sigma_m \) where \( \sigma_m \) is given by (5.132), so the formulae for \( E(g|UU) \) and \( E(g^2|UU) \) are established. Similarly, if state I obtains when an effective green period is initiated, then it will be uninterrupted and will also have conditional moments given by these same formulae. Now suppose that an effective green period is initiated while state U obtains and that state I obtains when the next such period is initiated. As before, with conditional probability \( \frac{P_a}{\sum_{i=m}^{n_1} P_a} \) stage 1 is recalled while stage \( n \) is running \( (n_1 \leq n \leq n_2) \). In this case the first two conditional moments of the duration of the effective green period are \( c(\sigma_n + \eta_n^a) \) and \( c^2(\sigma_n^2 + 2\sigma_n \eta_n^a + \eta_{2n}^a) \) where \( \sigma_n \) \( (n_1 \leq n \leq n_2) \) are given by (5.132). In all other cases the effective green period is uninterrupted so has the same conditional moments as for the transition from state U to state U. Multiplying the conditional
moments by the appropriate conditioning probabilities and summing gives the formulae for \( E(g \mid U) \) and \( E(g^2 \mid U) \). Thus (5.128) is established.

If an effective green period is initiated while state \( S \) obtains \((S \in \{U, I\})\), then the next effective red period will also be initiated while state \( S \) obtains. Since the duration of an effective green period is exactly \( c_n \) during any transition from state \( I \) and during transitions from state \( U \) to state \( U \), the formulae for \( E(g_{N-1} r_n \mid IS) (S \in \{U, I\}) \) and for \( E(g_{N-1} r_n \mid UU) \) given in (5.129) follow immediately. The only remaining case is that of a transition from state \( U \) to state \( I \). If a recall is granted while stage \( n \) is running \((n_1 \leq n \leq n_2)\), then the durations of the effective green and red periods are conditionally independent with means \( c(n_2 - n_1) \) and \( c_0 n \) respectively. Otherwise the duration of the effective green period is exactly \( c_n \) and the conditional mean of the duration of the effective red period can be calculated as before to be

$$
\left[ \sum_{n=0}^{m} P_n (\epsilon_n + n)^a \right] + \left( \sum_{n=0}^{m} P_n (\epsilon_n + n)^a \right) / \left[ \sum_{i=0}^{m} P_i + \left( \sum_{i=0}^{m} P_i \right) \right] .
$$

Multiplying the conditional expectations of the product of the durations of the effective green and red periods by the appropriate conditioning probabilities and summing gives the formula for \( E(g_{N-1} r_n \mid UI) \) in (5.129). Thus (5.129) is established.

Finally, if an effective red period is initiated while state \( S_1 \) obtains and state \( S_2 \) obtains when the next such period is initiated \((S_1, S_2 \in \{U, I\})\), then the intervening effective green period is initiated while state \( S_2 \) obtains. The duration of the effective green period is conditionally independent of the duration of the preceding effective red period and has mean given by

$$
\tilde{g}(1)_{S_1} E(g \mid S_2 U) + \tilde{g}(1)_{S_2} E(g \mid S_2 I) .
$$

The expression (5.130) follows immediately.
5.4 Delay to priority vehicles

5.4.1 Introduction.

The objectives of providing priority for buses include the reduction of delay incurred by them. Thus there is a natural interest in methods to estimate the size of any reduction in delay which may arise from the implementation of a priority scheme. In this section an approximate method is developed to estimate the mean delay incurred by priority vehicles. This method is applied to an arbitrary junction in the absence of priority and when there is priority by extension, extension and recall and extension and recall with inhibition. Comparison of these estimates yields an estimate of the reduction in the mean delay which results from implementation of the priority scheme.

Previous investigations of this problem have led only to crude estimates of the changes in mean delay. The Department of the Environment (1973b) described a method which actually estimates the mean time between the arrival of a bus and the next green indication. This method supposes that the time at which the first bus arrives is distributed uniformly over the duration of an uninterrupted cycle, that priority is never inhibited when a bus arrives at the detector and that the passage of priority vehicles is not impeded by other vehicles. Jacobson and Sheffi (1981) improved on this method to the extent that some account was taken of congestion caused by non-priority vehicles. However, their analysis remains approximate because the other two assumptions used by the Department of the Environment (1973b) were not relaxed. Furthermore, the analysis presented can be applied only to junctions controlled by traffic signals with 2 stages in the sequence.

The method adopted here supposes that buses arrive at the detector according to the Poisson law, accounts for cycles during which priority is not available due to the action of the inhibition rule and gives an
improved treatment of congestion caused by non-priority vehicles. The formula derived to estimate the mean delay incurred in the absence of priority is similar to that derived in Section 5.2 for non-priority vehicles under the same circumstances. The reason for the difference is shown to be due to the relatively simple treatment of the decay of the queue during effective green periods. While the more elegant analysis of Section 5.2 could be applied to this case, it could not be extended readily to cases where priority is available. When comparing two different circumstances such as the absence and presence of priority, similar methods should be used for the analysis in each case.

5.4.2 A method to estimate delay in the absence of priority.

In this sub-section a method is developed which leads to a formula to estimate the mean delay that would be incurred by priority vehicles in the absence of priority. The first results of this sub-section establish formulae which will be used throughout this section. Some further notation is required. Let

\[ w \] be the delay incurred by a priority vehicle

\[ W \] be the total delay incurred by priority vehicles during a single pair of effective red and green intervals

\[ V \] be the total number of priority vehicles which arrive at the detector during a single pair of effective red and green intervals.

**Lemma 5.29**

Suppose that at time \( t=0 \) there are \( Q \) vehicles queueing in the stream which contains priority vehicles and that the signals which control that stream are effectively red until time \( t=r \) and are effectively green from
then onwards. Then in the absence of priority the mean delay incurred by a
priority vehicle which arrives at the detector at time \( t = t_a \) is given
approximately by

\[
E(\tilde{w}(Q,r,t_a)) = \begin{cases} 
\frac{r + \frac{K^O}{s} - \left(\frac{s - Kq}{s}\right)(t_a + \tau c)}{s} \quad &\text{for } -\tau c \leq t_a \leq \left(\frac{s}{s - Kq}\right)r + \left(\frac{K^O}{s - Kq}\right)c \\
0 &\text{otherwise}
\end{cases}
\] (5.133)

where \( \tau \) was introduced in sub-section 4.4.1 and \( K \) was introduced in
sub-section 4.2.1.

**Proof**

The number of vehicles which will have to cross the stop-line before the
priority vehicle can be \( Q + A(t_a + \tau c) \). Thus the priority vehicle will
be able to leave the queue at time \( t = r + \tilde{S}[Q + A(t_a + \tau c)] \): since the priority
vehicle would arrive at the stop-line at time \( t_a + \tau c \) if there were no
queue, the delay it incurs is given by

\[
\tilde{w}(Q,r,t_a) = \max\{0, r + \tilde{S}[Q + A(t_a + \tau c)] - (t_a + \tau c)\}
\] (5.134)

Because of the non-linearity of the maximum function, there is no
convenient method of finding the expected value of (5.134). However, taking
expectations over the arrivals and departures before taking the maximum in
(5.134) yields the estimate (5.133).

**Lemma 5.30**

Suppose that at time \( t = 0 \) there are \( Q \) vehicles queuing in the stream
which contains priority vehicles and that the signals which control that
stream are effectively red until time \( t = r \) and are effectively green from
then onwards. Then in the absence of priority the expected total delay
incurred by priority vehicles which arrive at the detector after time
\( t = t_0 \) is given approximately by
\[
E(W^\# | Q, r, t_0) = \begin{cases} 
\frac{1}{2} \left( \frac{s}{s-Kq} \right) \left[ r + \frac{Kq}{s} - \left( \frac{s-Kq}{s} \right) (t_0 + rc) \right]^2 \frac{\beta}{c} \cdot \left[ t_0 \leq \left( \frac{s}{s-Kq} \right) r + \left( \frac{Kq}{s-Kq} \right) - rc \right] \\
0 \quad \text{otherwise}
\end{cases}
\] (5.135)

**Proof**

The total delay can be found by integrating with respect to time the product of the delay which would be incurred by a priority vehicle and the arrival density. Thus

\[
\tilde{W}^\#(Q, r, t_0) = \int_{t=t_0}^{t_0+c} \tilde{w}(Q, r, t) \beta/c \, dt
\] (5.136)

Taking expectations in (5.136) with respect to arrivals and departures of non-priority vehicles and using the result of lemma 5.29 gives

\[
E(\tilde{W}^\# | Q, r, t_0) = \begin{cases} 
\int_{t=t_0}^{t_0+c} \left( \frac{s}{s-Kq} \right) r + \left( \frac{Kq}{s-Kq} \right) - rc \\
\cdot \left[ t + \frac{Kq}{s} - \left( \frac{s-Kq}{s} \right) (t + rc) \right] \frac{\beta}{c} \, dt \quad \text{if } t_0 \leq \left( \frac{s}{s-Kq} \right) r + \left( \frac{Kq}{s-Kq} \right) - rc \\
0 \quad \text{otherwise}
\end{cases}
\]

which, on integration, gives (5.135).

The main result of this subsection is given next.

**Theorem 5.31**

Suppose that the stream which contains priority vehicles has an effective red period of duration \( r \) in each cycle of duration \( c \). Then in the absence of priority, the mean delay incurred by priority vehicles is given approximately by

\[
E(w) = \frac{1}{2c} \left( \frac{s}{s-Kq} \right) \left[ r^2 + \frac{2rE(Q)}{q} \right]
\] (5.137)

**Proof**

Consider the total delay incurred by those priority vehicles which would, in the absence of any impediment, arrive at the stop-line during an
effective red period and the subsequent effective green period. Since these vehicles are just the ones which would arrive at the detector during the interval \([-\tau c, (1-\tau)c]\), this delay is given by \(E(W^p|Q, r, -\tau c) - E(W^p|Q', 0, -\tau c)\) where \(Q\) (\(Q'\)) is the number of vehicles in the queue at the beginning (respectively, end) of the cycle. An additional total delay of \(Q[r+\tilde{S}(Q)]\) will be incurred by vehicles in the overflow queue: a proportion \(\beta/cq\) of this will be incurred by priority vehicles. Finally, by a similar argument, a total delay of \(\frac{\beta}{2cq} Q'\tilde{S}(Q')\) would be incurred after the end of the cycle to priority vehicles which arrive at the detector during the interval \([-\tau c, (1-\tau)c]\). Thus the expected total delay incurred by priority vehicles during the cycle is given by

\[
E(W|Q, Q') = E(W^p|Q, r, -\tau c) - E(W^p|Q', 0, -\tau c) + \frac{\beta}{2cq} E(Q[2r+\tilde{S}(Q)]-Q'\tilde{S}(Q'))
\]

(5.138)

Using the result of lemma 5.30, taking expectations over the overflow size and rearranging gives

\[
E(W) = \frac{1}{2c} \left( \frac{s}{s-kq} \right)[r^2 + \frac{2rE(Q)}{q}]\beta
\]

(5.139)

Since the expected number of priority vehicles to arrive during the cycle is just \(\beta\), \(E(W)=E(W)/\beta\) so (5.137) is established.

The expression (5.137) for the mean delay incurred by a priority vehicle in the absence of priority corresponds to the first two terms of (5.50) when this is applied to fixed-time traffic signals. The difference of \(\frac{r}{2c} \frac{Ks}{(s-kq)^2} [(I-1)+\left(\frac{C^p}{s}\right)(C^2-1)]\) arises as a result of the approximate analysis used in this sub-section. In particular, it can be attributed to the step where the expectation operator was passed inside the maximum function in the right-hand side of (5.134). This error, which is of the order of the service time of one vehicle, will generally be small by
comparison with the mean delay. In any case, when investigating the change in mean delay due to the implementation of a priority scheme, this error might be expected to cancel out, or nearly to do so, with the corresponding error in the estimate of delay when there is priority.

5.4.3 Priority by extension.

In this sub-section, the method which was used to derive the formula (5.137) is used to estimate the mean delay incurred by priority vehicles when there is priority by extension. The resulting formula is of interest in its own right and will also be used in sub-section 5.4.5 to estimate the total delay incurred priority vehicles during cycles in which inhibition is in effect.

The derivation given here depends on two additional assumptions which merit some discussion. The first is that the extension period is sufficiently long to enable every priority vehicle which is granted priority by extension to pass the stop-line without incurring any delay. When applying the resulting formula, care must be taken to ensure that this assumption is satisfied: if it is not, then the duration of the extension period is not long enough for the extension priority method to be effective. If the formula derived here is used in an optimisation procedure, then a constraint may be required to ensure that this assumption is not violated.

The second assumption is that non-priority vehicles do not prevent priority vehicles from reaching the detector. As a consequence of these assumptions, priority vehicles can never be included in the overflow queue. This results in some simplification in the analysis.
Lemma 5.32

Suppose that priority by extension is provided for buses arriving at a mean rate of $\beta/c$. Then the expected total delay incurred by priority vehicles during an effective red period and the subsequent effective green period is given approximately by

$$
E[W^{(e)}] = \begin{cases} 
\frac{1}{2} \left( \frac{s}{s-K} \right) \left\{ [r-(s-Kq)/s] + (2K/s) [r-(s-Kq)/s] \right\} B(Q) \left( \frac{E(Q)}{s-Kq} \right) \\
\text{otherwise}
\end{cases}
$$

(5.140)

where $r = \sum_{i=k}^{m} (1-a_{i0}) \lambda_{i0} + \sum_{i=k+1}^{m} \lambda_{i}$

(5.141)

Proof

Since any bus which arrives at the detector during the time interval $[(1-t)c, c]$ will be granted an extension and will thus, by assumption, incur no delay, the quantity under investigation is just the expected total delay incurred by priority vehicles which arrive at the detector during the time interval $[0, (1-t)c]$. The duration of an effective red period for the stream which contains priority vehicles is given by (5.141), so

$$
E[W^{(e)}|Q,Q'] = E(W^{(e)}, r, 0) - E(W^{(e)}, Q', 0, c)
$$

(5.142)

Using the result of lemma 5.30, replacing powers of $Q$ by their expected values and rearranging gives (5.140).

Lemma 5.33

Suppose that priority by extension is provided for buses arriving at a mean rate of $\beta/c$. Then the mean number of buses which arrive during an effective red period and the subsequent effective green period is given by

$$
E[V^{(e)}] = \beta + e^{\beta t} - (1+\beta t)
$$

(5.143)
Proof
The mean number of buses to arrive at the detector during the interval 
[0, (1-τ)C] is just \((1-τ)C\). Any bus which arrives after this interval 
but before the start of the next cycle will be granted an extension, so the 
mean number of extensions granted during a cycle must be added to this. A 
bus arrives at the detector during a period of duration \(τC\) with probability 
\(P = 1 - p_0(τ)\): since this causes stage \(k\) to be extended by a 
period of duration \(τC\), the process is re-initiated. Thus the mean number 
of extensions is given by \(\sum_{n=1}^{\infty} e^{βτ} - 1\) (from (4.36)), so (5.143) is 
established.

Theorem 5.34
Suppose that priority by extension is provided for buses arriving at a mean 
rate of \(β/C\). Then the mean delay incurred by these buses is given 
approximately by

\[
E[w^{(e)}] = \begin{cases} 
\frac{s[(r-(s-Kq)/s)τC]^2 + \frac{2K}{s}[r-(s-Kq)/s]τC]E(Q)}{2c(s-Kq)(1+[e^{βτ}-(1+βτ)]/β)} & \text{otherwise} \\
0 & \text{otherwise} 
\end{cases}
\]

where \(r\) is given by (5.141).

Proof
According to the result of lemma 5.32, the expected total delay incurred 
during a cycle is given approximately by (5.140). The mean number of buses 
to arrive during a cycle is given by (5.143): using these expressions in 
the relationship \(E[w^{(e)}] = E[W^{(e)}]/E[V^{(e)}]\) gives (5.144).

In the proof of theorem 5.34 a further approximation was introduced in 
passing from (5.142) to the unconditional estimate of the total delay 
inherited during a cycle. This will result in an underestimate of delay. In 
particular, consider the rather unlikely case where \(τC > [(sK + KEB(Q))/(s-Kq)]\).
This corresponds to the detector being placed so far in front of the stop-line that either priority vehicles are given extensions or they arrive at the stop-line after the mean time at which a queue of mean size has dissipated. Here (5.144) yields an estimate of 0 whereas buses which arrive during a cycle which commences with large initial queues may indeed incur some delay.

5.4.4 Priority by extension and recall.

In this sub-section the method which was used in the last two sub-sections is used to estimate the mean delay incurred by priority vehicles when there is priority by extension and recall. The derivation given here uses the conditional expectations of the total delay incurred by priority vehicles during an effective red period and the subsequent effective green period given the number of the stage which is running when stage 1 is called. In effect, this conditions the expected total delay on the time of the first arrival of a priority vehicle at the detector after the call of stage \( k=1 \).

Lemma 5.35

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of \( \beta/c \). Then the expected total delay incurred by those priority vehicles which arrive at the detector during an effective red period in which stage 1 is called while stage \( n \) is running (\( k\leq n\leq m \)) and the subsequent effective green period is given approximately by

\[
E[w^{(r)}|n] = \begin{cases} 
\left\{U_n^{(1)} + [1-P_0(\lambda_{n-1},0+y_n)]U_n^{(2)} + U_n^{(3)} + U_n^{(4)} + U_n^{(6)} \right\} / [1-P_0(\lambda_{n-1},0+y_n)] \\
U_n^{(1)} + [1-P_0(\lambda_{n-1},0+y_n)]U_n^{(2)} + U_n^{(3)} + U_n^{(4)} + P_0(\lambda_{n-1},0+y_n)U_n^{(5)} - U_n^{(6)} 
\end{cases} 
\]

(5.145)
where

\[
U_n^{(1)} = \begin{cases}
0 & \text{if } t_n \leq 0 \\
\frac{\beta}{2c} \left( \frac{s}{s-kq} \right) \left[ \lambda_{n-1,0} + \lambda_{n1} \right] c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c \\
[\text{otherwise}]
\end{cases}
\]  

\[
U_n^{(2)} = \begin{cases}
0 & \text{if } t_n \leq 0 \\
\frac{\beta}{2c} \left( \frac{s}{s-kq} \right) \left[ \lambda_{n-1,0} + \lambda_{n1} \right] c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c \\
[\text{otherwise}]
\end{cases}
\]  

\[
U_n^{(3)} = \begin{cases}
0 & \text{if } t_n \leq 0 \\
\frac{\beta}{2c} \left( \frac{s}{s-kq} \right) \left[ \lambda_{n-1,0} + \lambda_{n1} \right] c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c \\
[\text{otherwise}]
\end{cases}
\]  

\[
U_n^{(4)} = \begin{cases}
0 & \text{if } t_n \leq 0 \\
\frac{\beta}{2c} \left( \frac{s}{s-kq} \right) \left[ \lambda_{n-1,0} + \lambda_{n1} \right] c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c \\
[\text{otherwise}]
\end{cases}
\]  

\[
U_n^{(5)} = \begin{cases}
0 & \text{if } t_m \leq 0 \\
\frac{\beta}{2c} \left( \frac{s}{s-kq} \right) t_m \tau c \\
[\text{otherwise}]
\end{cases}
\]  

\[
U_n^{(6)} = \left( \frac{s}{s-kq} \right) ^2 \left[ E(Q) \right] ^2
\]  

\[
t_n = \left( \frac{s}{s-kq} \right) \left[ \lambda_{n-1,0} + \lambda_{n1} \right] c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c
\]  

\[
t_n^+ = \left( \frac{s-kq}{kq} \right) \tau c - \left( \frac{s}{kq} \right) \left[ \lambda_{n-1,0} \right] c - \frac{E(Q)}{s} - \frac{\psi}{s}
\]  

\[
t_m^+ = \left( 1-a_m \right) \lambda_{m1} c + \frac{K}{s} [E(Q) + q\psi c] - \frac{s-kq}{s} \tau c
\]  

and

\[
\psi = \sum_{i=k}^{n} \lambda_i 0^+ + \sum_{i=k+1}^{n-1} \lambda_i
\]
Proof

Suppose initially that \( k \leq n \leq m \) : there are two cases to consider, depending on when the first bus arrives at the detector. With conditional probability \( [1 - \tilde{p}_0(\lambda_{n-1}, 0^+1_n)]/[1 - \tilde{p}_0(\lambda_{n-1}, 0^+1_n)] \) this occurs between the call of stage \( n \) and the end of the minimum green time for that stage. The conditional probability density of bus arrivals in this interval is \( \beta/(c[1 - \tilde{p}_0(\lambda_{n-1}, 0^+1_n)]) \) and the expected delay incurred by one which arrives at time \( t \in [0, \gamma_n, \gamma_n] \) is

\[
E\{w^*|Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \}.
\]

If a bus arrives at the detector during this interval and is granted priority by recall, then the expected total delay incurred by all buses which arrive at the detector after stage \( t \) is called

\[
E\{w^*|Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, [\psi_n+\gamma_n]_c \} - E\{w^*|Q', 0, -tc \}.
\]

The conditional probability density for the arrival of the first bus at the detector during the remainder of stage \( n \), i.e. the time interval \( [0,\gamma_n, \gamma_n] \), is

\[
\beta_0\frac{(t_c)}{c[\gamma_n, \gamma_n]}\frac{1-\tilde{p}_0(\lambda_{n-1}, 0^+1_n)}{[1 - \tilde{p}_0(\lambda_{n-1}, 0^+1_n)]}.
\]

In this case, the expected delay incurred by the first bus is

\[
E\{w^*|Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \}.
\]

The expected total delay incurred by buses which arrive after that time is

\[
E\{w^*|Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \} - E\{w^*|Q', 0, -tc \}.
\]

Integrating the delays with respect to the conditional probability density functions for the time of arrival of the first bus at the detector gives

\[
E\{w^*[Q, \psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \} = \int_{t=0}^{(\lambda_{n-1}, 0^+1_n)} E\{w^*[Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \} dt + \int_{t=(\lambda_{n-1}, 0^+1_n)}^{(1-\tilde{p}_0(\lambda_{n-1}, 0^+1_n))} E\{w^*[Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \} dt + \int_{t=(\lambda_{n-1}, 0^+1_n)}^{(1-\tilde{p}_0(\lambda_{n-1}, 0^+1_n))} E\{w^*[Q, [\psi_n+\gamma_n + (1-a_{n1j})\lambda_n]_c, \psi_n+c+t \} dt - E\{w^*[Q', 0, -tc \} \} / [1 - \tilde{p}_0(\lambda_{n-1}, 0^+1_n)] \] (5.154)
Substituting the forms (5.133) and (5.135) for $E(w|Q,r,t_a)$ and $E(w^g|Q,r,t_\theta^g)$ respectively in (5.154) and substituting appropriate moments for powers of the overflow gives an expression which can be integrated using (4.36) to give (5.145) for $k<n<m$.

Now suppose that $n=m$; here there is an additional case to consider. This case is that no priority recall is granted and occurs with conditional probability $\tilde{p}_0(\lambda_{m-1,0}^m)$. The conditional expected total delay incurred by priority vehicles is then

$$E[w^*|Q,\lambda_m^{1},\lambda_0^{m-1} c, (1-a_{0j}^m) \lambda_0^{m-1} c] - \tilde{E}[w^*|Q',0,-\tilde{c}]$$

The conditional probability density functions for the other two cases differ from those given above for $(k<n,m)$ only by a multiplicative factor of $[1-\tilde{p}_0(\lambda_{m-1,0}^m)]$: proceeding as before leads to (5.145) when $n=m$.

**Lemma 5.36**

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of $\beta/c$. Then the expected total number of buses to arrive at the detector during an effective red period in which stage 1 is called while stage $n$ is running $(k<n=m)$ and the subsequent effective green period is given by

$$E[V(x)|n] = \frac{\beta(\lambda_{n-1,0}^n + \tilde{p}_0(\lambda_{n-1,0}^n)) - \tilde{p}_0(\lambda_{n-1,0}^n) + \beta \lambda_n^1 + \sum_{i=1}^{k-1} \lambda_n^i + \sum_{i=1}^{k-1} \lambda_0^i)}{[1-\tilde{p}_0(\lambda_{n-1,0}^n)]} \left\{ \begin{array}{ll}
\frac{\beta(\lambda_{n-1,0}^n)}{[1-\tilde{p}_0(\lambda_{n-1,0}^n)]} & (k<n=m) \\
\beta(\lambda_{n-1,0}^n) + \tilde{p}_0(\lambda_{n-1,0}^n) + \beta \lambda_n^1 + \sum_{i=1}^{k-1} \lambda_n^i + \beta \lambda_0^i + \sum_{i=1}^{k-1} \lambda_0^i & (n=m) \end{array} \right\}$$

**(5.155)**

**Proof**

Suppose that $k<n<m$. The conditional probability that the first bus arrives during the transition period preceding stage $n$ or during the minimum green period for that stage is
The conditional probability density of bus arrivals during that interval is then \( \beta / \{ c[1 - \tilde{P}_0(\lambda_{n-1}, 0^{+} \gamma_{n})] \} \). Integrating this conditional probability density over the interval \([0, \lambda_{n-1}, 0^{+} \gamma_{n}) \) and multiplying by the conditioning probability gives a conditional expected number \( \beta (\lambda_{n-1}, 0^{+} \gamma_{n}) / [1 - \tilde{P}_0(\lambda_{n-1}, 0^{+} \gamma_{n})] \) of arrivals during this interval.

With conditional probability

\[
\tilde{P}_0(\lambda_{n-1}, 0^{+} \gamma_{n}) [1 - \tilde{P}_0(\lambda_{n}, \gamma_{n})] / [1 - \tilde{P}_0(\lambda_{n-1}, 0^{+} \gamma_{n})]
\]

the first bus arrives after the minimum green time for stage \( n \) has elapsed: in this case exactly one bus arrives before stage 1 is called.

Arguing as in the proof of lemma 5.33, the mean number of buses to arrive between the call of stage 1 and the end of the effective green period for \( k = \ldots, k-1 \) buses is \( \beta \left[ \frac{\lambda_{n1}^{+} \gamma_{1}^{+} \gamma_{10}^{+} + \ldots + \beta_{n1}^{+} \gamma_{1}^{+} \gamma_{10}^{+}}{e^{r_{1}^{+}} - (1 + \beta_{1})} \right] \). Adding the three contributions gives (5.155) for \( k \leq n \leq m \).

Suppose now that \( n = m \). With conditional probability

\( 1 - \tilde{P}_0(\lambda_{m-1}, 0^{+} \gamma_{m}) \), a recall is granted during stage \( m \) in which case the formula derived above applies. Otherwise the expected total number of bus arrivals is

\[
\beta \left[ \frac{\lambda_{m0}^{+} \gamma_{1}^{+} \gamma_{10}^{+} + \ldots + \beta_{m1}^{+} \gamma_{1}^{+} \gamma_{10}^{+} - 1}{e^{r_{1}^{+}} - (1 + \beta_{1})} \right].
\]

Using \( \lambda_{m0} = \lambda_{m1} \), (5.155) is established for \( n = m \).

Theorem 5.37

Suppose that priority by extension and recall is provided for buses arriving at a mean rate of \( \beta / c \). Then the mean delay incurred by priority vehicles is given approximately by

\[
E[w(r)] = \frac{\sum_{n=k+1}^{m} P_n E[w(x)] | n]}{\sum_{n=k+1}^{m} P_n E[v(x)] | n]}, \quad (5.156)
\]

where \( P_n \) \( (k \leq n \leq m) \) are given by (4.62), \( E[w(x)] | n] \) \( (k \leq n \leq m) \) are given by (5.145), and \( E[v(x)] | n] \) \( (k \leq n \leq m) \) are given by (5.155).
Proof

Since \( P_n \) (\( k \leq n \leq m \)) are the conditioning probabilities for the conditional expectations \( E[W^{(x)}|n] \) (\( k \leq n \leq m \)) given approximately by (5.145), the expected total delay incurred by priority vehicles during an effective red period and the subsequent effective green period is given by

\[
E[W^{(x)}] = \sum_{n=k+1}^{m} P_n E[W^{(x)}|n]
\]

(5.157)

Similarly, the expected total number of priority vehicles to arrive during the same period is given by

\[
E[V^{(x)}] = \sum_{n=k+1}^{m} P_n E[V^{(x)}|n]
\]

(5.158)

Dividing (5.157) by (5.158) and using the results of lemmas 5.35 and 5.36 gives (5.156).

5.4.5 Priority by extension and recall with inhibition.

In this subsection the results of the last two subsections are used to provide an estimate of the mean delay incurred by priority vehicles when there is priority by extension and recall with inhibition. The method used here is to estimate the expected total delay and number of arrivals in an effective red period during which inhibition is in effect and the subsequent effective green period by the corresponding estimates when there is priority by extension alone. Since this method ignores any correlation between the size of the overflow and the duration of the next effective red period, it will tend to over-estimate delays, especially when the mean overflow is large. The error introduced by this approximation gives rise to an over-estimate of delays since the queue in the stream which contains buses will tend to be smaller at the start of an effective red period during which inhibition is in effect thus giving rise to a negative correlation between \( Q_n \) and \( r_n \).
Theorem 5.38
Suppose that priority by extension and recall with inhibition is provided for buses arriving at a mean rate of \( \beta / c \). Then the mean delay incurred by priority vehicles is given approximately by

\[
E[W(I)] \approx \frac{P_1 E[W(e)] + (1-P_1) \sum_{n=k+1}^{m} P_n E[W(x) | n]}{P_1 E[V(e)] + (1-P_1) \sum_{n=k+1}^{m} P_n E[V(x) | n]}. \tag{5.159}
\]

where \( P_n \ (k \leq n \leq m) \) are given by (4.62),

\( P_1 \) is given by (4.68),

\( E[W(e)] \) and \( E[V(e)] \) are given by (5.140) and (5.143) respectively and \( E[W(x) | n] \) and \( E[V(x) | n] \) (\( k \leq n \leq m \)) are given by (5.145) and (5.155) respectively.

Proof
According to the result of lemma 4.26, the probability that inhibition is in effect during all of an effective red period for priority vehicles is \( P_1 \) (4.68). During an effective red period in which inhibition is in effect and the subsequent effective green period the only form of priority which is available is extension. Thus the expected total delay and number of priority vehicles to arrive during such an interval are given approximately by (5.140) and (5.143) respectively. With probability \( (1-P_1) \), priority by both extension and recall is available during an effective red period and the subsequent effective green period so the expected total delay and number of priority vehicles to arrive are given approximately by (5.157) and (5.158) respectively. Multiplying each of these conditional expectations by the appropriate conditioning probability and adding gives
\[ E[W^{(1)}] = P_i E[W^{(e)}] + (1-P_i) E[W^{(x)}] \]  \hspace{1cm} (5.160)

and \[ E[V^{(1)}] = P_i E[V^{(e)}] + (1-P_i) E[V^{(x)}] \]  \hspace{1cm} (5.161)

Dividing (5.160) by (5.161) and substituting for \( E[W^{(x_i)}] \) and \( E[V^{(x_i)}] \) from (5.157) and (5.158) respectively gives (5.159).
5.5 Practical considerations for the estimation
of delay to non-priority vehicles

5.5.1 Introduction.

In Sections 5.2 and 5.3 of this chapter, formulae were derived which can be used to estimate the mean delay incurred by non-priority vehicles in each stream when priority is provided for buses. Because of the considerable number of cases which can arise, a large number of formulae were given. These formulae are to be used in various combinations depending on the priority method in use and the class to which the stream under consideration belongs. The calculations are made more intricate in some cases because indirect methods are used.

In order to clarify the way in which these formulae are to be used, an outline is given in sub-section 5.5.2 of a practical procedure to estimate the mean delay incurred by non-priority vehicles in each stream. This procedure is illustrated by two interrelated flow-charts which indicate the order in which the results of Sections 5.2 and 5.3 are to be used in any particular case. These flow-charts also serve as a guide to the main results of this chapter. Some aspects of the calculations are discussed with reference to these flow-charts.

Part of the intricacy of the formulae presented in Sections 5.2 and 5.3 can be attributed to the degree of generality permitted by the analysis. In sub-section 5.5.3 a discussion is given of some possible restrictions of the analysis which give rise to less intricate formulae. The extent to which these restrictions affect the results of the analysis depends on several considerations. These include the amount of detailed information which is available, the desired accuracy of the results and the complexity of the junction under investigation.
5.5.2 Procedures to estimate delay.

Because of the range of possibilities for which the analysis in Section 5.3 was presented, only a comparatively small number of the results given there are required to estimate the mean delay to vehicles in any one stream. Two flow-charts are given in this sub-section which show the results to be used in any particular case and the order in which they are used.

The procedure presented here shows how to estimate the mean delay incurred by non-priority vehicles in a single stream. Various other estimates, such as that of the mean overflow, arise from the procedure. Furthermore, the procedure can be augmented to estimate other quantities such as the mean number of times that a vehicle has to stop. This procedure can be used in turn for each stream of traffic at a junction.

The flow-chart in Figure 5.1 depicts the process of estimating the mean delay incurred by vehicles in a stream which is in any of the classes introduced in lemma 5.17. The last three steps of this flow-chart make use of the statistics (5.84); the rest of the flow-chart is concerned with the estimation of these statistics. If the stream under consideration belongs to class 1, 2 or 4, then the statistics (5.84) can be calculated directly from the procedure depicted by the flow-chart in Figure 5.2. Otherwise the statistics (5.84) are calculated for a notional stream which is complementary to the stream under consideration. The required statistics can then be deduced from the result of lemma 5.12. The result of theorem 5.13 guarantees that this procedure will always lead to the required estimates.

The flow-chart in Figure 5.2 depicts the process of estimating the statistics (5.84) for a single stream which is in class 1, 2 or 4. Both
Figure 5.1: Flow-chart for calculations to estimate the mean delay to non-priority vehicles in stream \( j \) when there is priority for buses.
Figure 5.2: Flow chart for calculation of statistics (5.84) for stream $j$ in class 1, 2 or 4 when there is priority by extension and recall with or without inhibition.
of the combinations of priority rules considered in Section 5.3, namely extension and recall and extension and recall with inhibition, are included in the flow-chart. In the latter case, an additional step appears in the flow-chart since the individual calculations for the streams in the different classes give only the conditional expectations (5.110). The additional step is to combine these quantities according to the method described in lemma 5.25 to yield the required statistics. The essence of this flow-chart is to indicate which one of the six theorems, 5.20, 5.21, 5.23, 5.26, 5.27 and 5.28 is appropriate to any particular case.

5.5.3 Some possible simplifications.

The formulae presented in Sections 5.2 and 5.3 allow for some possibilities which might arise only occasionally. Irrespective of how often these possibilities do arise, the formulae presented are complicated by their facility to allow for them. In this sub-section, some restrictions of the analysis are discussed. These lead to more manageable formulae at the expense of some loss of generality.

The three areas in which some simplification is possible are as follows. Perhaps the least important simplification is to restrict the generality of the arrival and departure processes of non-priority vehicles. The second possibility is to restrict the range of values of $m$ and $k$ to reduce the number of classes to which a stream can belong. Finally, the durations of the transition periods can be specified as all being equal and the start and end lags for each stream can be treated as being independent of the transition which initiates or terminates the effective green period.

The analysis in Section 5.2 allowed for compound Poisson arrivals and variable departure times for non-priority vehicles. If these processes are restricted to simple Poisson and constant departure times, then this has
Figure 5.3: Possible classes of streams depending on the number of stages in the cycle.
the effect of setting \( I_a = 1 \) and \( C_s = 0 \). Accordingly, the third term of the expression (5.50) for the mean delay reduces to
\[
\frac{-K^2 q}{2(s-Kq)} \frac{E(r)}{[E(r)+E(g)]}.
\]
Similarly, the expression (5.68) for the mean overflow will be simplified by the replacement of \( I_a \) by 1 and the removal of the last term from the numerator. While these two simplifications make the expressions (5.50) and (5.68) easier to evaluate, they do not reduce the number of statistics required for their use.

The analysis in Section 5.3 allows for an arbitrary number of stages in the cycle and for priority vehicles to receive right of way in any lesser number of stages. This gives rise to the seven different classes of streams. If the number of stages in the cycle is restricted, then some of these classes may be empty. Figure 5.3 shows schematically the different possibilities that arise from various such restrictions: higher positions in the figure represent greater restriction on the number of stages in the cycle. In the three cases shown where \( k=1 \), no stream can belong to class 2: furthermore, no stream can have a complementary stream which belongs to this class. Thus if \( k=1 \), then only the formulae for streams in classes 1 and 4 are required. In the three cases shown where \( m=k+1 \), there are \( m \) stages in every cycle, an effective red period is initiated for each stream in every cycle and the start and end lags for each stream will always be the same. This results in some considerable simplification of the analysis, especially for streams in class 4.

When priority by extension and recall is provided for buses, the statistics (5.84) can be calculated for any stream in class 4 from the formulae (5.102-4) given in theorem 5.23. However, if \( m=k+1 \), then these formulae can be reduced considerably to give
\[ E(r) = c \left[ \sum_{i=1}^{m} (1-a_{i0j}) \lambda_{i0} + \sum_{i=1}^{k} \lambda_i + (\varepsilon_k - \lambda_k) \right] \]

\[ E(r^2) = [E(r)]^2 + c^2 \upsilon_k \]

\[ E(g) = c (a_{k0j} \lambda_{i0} + \eta_m + a_{m1j} \lambda_{m1}) \]

\[ E(g^2) = [E(g)]^2 + c^2 (\eta_{2m} - \eta_m^2) \]

\[ \text{Cov}(r_{i0}, g_{j0}) = 0 \quad \text{if } i \geq 0 \]

\[ \text{Cov}(r_{i0}, g_{j0}) = 0 \quad \text{if } i < 0 \]

where \( \varepsilon_k \) and \( \upsilon_k \) are given in (4.52)

and \( \eta_m \), \( \eta_{2m} \) are given by (4.65-6).

When priority by extension and recall with inhibition is provided for buses, the conditional expectations (5.110) can be calculated for any stream in class 4 using the formulae (5.127-32) given in theorem 5.28. The statistics (5.84) can then be deduced for that stream using the formulae (5.111-5) of lemma 5.25. However, if \( m=k+1 \), then the formulae (5.127-32) can be reduced considerably. In this case, the transition II cannot occur so the corresponding conditional expectations are not required in the analysis. The other conditional moments of \( r \) and \( g \) are given by

\[ E(r|SU) = c \left[ \sum_{i=1}^{m} (1-a_{i0j}) \lambda_{i0} + \sum_{i=1}^{k} \lambda_i + (\varepsilon_k - \lambda_k) \right] \quad (S \in \{U, I\}) \]

\[ E(r|UI) = E(r|UU) \]

\[ E(r^2|SU) = [E(r|SU)]^2 + c^2 \upsilon_k \quad (S \in \{U, I\}) \]

\[ E(r^2|UI) = E(r^2|UU) \]

\[ E(g|SU) = c (a_{k0j} \lambda_{i0} + \lambda_m + a_{m1j} \lambda_{m1}) \quad (S \in \{U, I\}) \]

\[ E(g|UI) = c (a_{k0j} \lambda_{i0} + \lambda_m + a_{m1j} \lambda_{m1}) \]

\[ E(g^2|SU) = [E(g|SU)]^2 \quad (S \in \{U, I\}) \]

\[ E(g^2|UI) = [E(g|UI)]^2 + c^2 [\eta_{2m}^a - (\eta_m^a)^2] \]

where \( \varepsilon_k \) and \( \upsilon_k \) are given in (4.52)

and \( \eta_m^a \), \( \eta_{2m}^a \) are given by (4.64).

The conditional expectations (5.129-30) could be derived for this case in a similar manner. However, they are not required since the durations of
the effective red and green periods are mutually independent. Thus

$$\text{Cov}(r_{\ell-1}, s_{\ell}) = \begin{cases} 1 \geq 0 \\ \text{i<0} \end{cases}$$

Finally, suppose that the start and end lags for each stream are independent of the transitions which initiate and terminate the effective green period. In this case the additional effective green time gained during the initiating and terminating transition periods can be represented by the single quantity $\delta_j$ where

$$\delta_j = \begin{cases} a_{n-1, j} \lambda_{n-1, 0} + a_{n_0 j} \lambda_{n_0} \\ a_{n_0 j} \lambda_{n_0} + a_{n_2 0 j} \lambda_{n_2} \end{cases}$$

(5.164)

Using this new quantity, $a_{i0j}$ and $a_{i1j}$ can be eliminated from the expressions used to calculate the statistics (5.84). These quantities appear only in the formulae for $\rho_n$ and $\sigma_n$: simplified formulae are given here which are appropriate for the case described above.

The formulae for $\rho_n$ and $\sigma_n$ ($k<n=\infty$) in (5.88-9) are identical to those given in (5.120). Thus for streams in class 1, under the conditions described, the formulae for $\rho_n$ and $\sigma_n$ can be reduced using (5.164) to give

$$\rho_n = \sum_{i=n_2+1}^{n_1} \lambda_i + (1-\delta_{k,n_2}) (\lambda_k - \lambda_k) + \sum_{i=n_2}^{n_1-1} \lambda_i + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0} - \delta_j)$$

(5.165)

$$\sigma_n = \sum_{i=n_1}^{n_2} \lambda_i + \delta_{k,n_2} (\lambda_k - \lambda_k) + \sum_{i=n_1}^{n_2-1} \lambda_i + \delta_j$$

Similarly, for streams in class 2, the formulae in (5.93-4) and (5.125-6) for $\rho_n$ and $\sigma_n$ ($k<n=\infty$) can be reduced to give
\[
\begin{align*}
\rho_n &= \frac{\sum_{i=n_2+1}^{n-1} \lambda_i + \sum_{i=n_2}^{n-1} \lambda_0 + \lambda_n + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0}) - \delta_j}{\text{(k<n<m)}} \\
\sigma_n &= \begin{cases} \\
\frac{n_2}{\sum_{i=n_1}^{n_2-1} \lambda_i + \sum_{i=n_1}^{n_2-1} \lambda_0 + (\varepsilon_k - \lambda_k) + \delta_j} & \text{(n_2<n<m)}
\end{cases}
\end{align*}
\] (5.166)

For streams in class 4, different formulae are appropriate for \(\rho_n\) depending on whether or not the inhibition rule is in use. Thus if priority by extension and recall is provided, then the formulae (5.102) reduce to

\[
\rho_n = \frac{\sum_{i=n_2+1}^{n-1} \lambda_i + \sum_{i=n_2}^{n-1} \lambda_0 + \lambda_n - \delta_j}{\text{(n_1<n<m)}} \quad (5.167)
\]

and if priority by extension and recall with inhibition is provided, then the formulae (5.131) reduce to

\[
\rho_n = \begin{cases} \\
\frac{m}{\sum_{i=n_2}^{n-1} \lambda_i + \sum_{i=n_2}^{n-1} \lambda_0 + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0}) + (\varepsilon_k - \lambda_k) - \delta_j} & \text{(n=0)} \\
\frac{\sum_{i=n_2}^{n-1} \lambda_i + \sum_{i=n_2}^{n-1} \lambda_0 + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0}) + \lambda_n + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0}) + 2(\varepsilon_k - \lambda_k) - \delta_j}{\text{(n_1<n<n_2)}} \\
\frac{n_1-1}{\sum_{i=n_2}^{n-1} \lambda_i + \sum_{i=n_2}^{n-1} \lambda_0 + \lambda_n + \sum_{i=1}^{n_1-1} (\lambda_i + \lambda_{i0}) + (\varepsilon_k - \lambda_k) - \delta_j} & \text{(n_2<n<m)}
\end{cases} 
\] (5.168)

The formulae (5.103) and (5.132) for \(\sigma_n\) are identical and reduce to

\[
\begin{align*}
\sigma_n &= \begin{cases} \\
\sum_{i=n_1}^{n-1} (\lambda_i + \lambda_{i0}) + \delta_j & \text{(n_1<n<n_2)} \\
\frac{n_2}{\sum_{i=n_1}^{n_2-1} \lambda_i + \sum_{i=n_1}^{n_2-1} \lambda_0 + \delta_j} & \text{(n_2<n<m)}
\end{cases} 
\end{align*}
\] (5.169)

If the durations of all the transition periods are identical, then the formulae (5.165-9) can be simplified further. This can be achieved by replacing each occurrence of \(\lambda_{i0}\) or \(\lambda_{i1}\) by some single quantity. However, because of the convention concerning the value of summations over null ranges, not all of the summations can be eliminated without the introduction of some corresponding notation.
CHAPTER 6
WORKED EXAMPLES

6.1 Introduction

In order to illustrate the use of the results presented in Chapters 4 and 5, they are applied in this chapter to two example junctions. Both of these are closely related to real junctions which have been used elsewhere as examples. Various of the calculations introduced in Chapters 4 and 5 are performed for each junction in turn. The results of these calculations are then compared and discussed.

Signal-settings which, when implemented with responsive priority, would be capacity-equivalent to some given set in the absence of priority are calculated for each junction according to the methods described in Chapter 4. This is done for a range of mean bus arrival rates and for each of a variety of different combinations of priority rules. The mean delay incurred by non-priority vehicles in each stream at the junctions and by priority vehicles are estimated according to the methods of Chapter 5 for each of the two most commonly used combinations of priority rules.

Two sets of results are presented for each combination of priority rules: in the first, the signal-settings used remain fixed as the mean bus arrival rate varies whereas in the second the signal-settings used are calculated according to the methods of Chapter 4 to achieve capacity-equivalence. These results are then used to estimate the minimum ratio of the mean occupancy of priority to non-priority vehicles required for implementation of responsive priority to give a net saving in passenger delay. The results of these analyses are compared to illustrate some of the consequences of the use of capacity-equivalent signal-settings in conjunction with responsive priority.
Figure 6.1: The Derby junction
6.2 The junction at Derby

6.2.1 Introduction.

The first junction for which example calculations are presented corresponds to the site of an experimental priority implementation at Derby, England (Department of the Environment, 1973b). Figure 6.1 depicts the junction as it is used here. The buses for which priority is provided travel in stream 1 and have right of way during the first of the three stages in the sequence.

The numerical data used as the basis for the calculations are given in Tables 6.1. The sequence comprises three stages: the durations given in Table 6.1a are taken from the Department of the Environment (1973b). The data for the streams are given in Table 6.1b. The saturation departure rate for each stream is given in units of vehicles rather than pscu's per hour and this corresponds to the quality $\kappa_j/\mu_j$. No flow data were given by the Department of the Environment (1973b) in their report of the experiment: the data in Table 6.1b were devised specially for this example. In particular, the flows of non-priority vehicles were chosen so that the signal-settings given in Table 6.1a are similar to those which minimise the total rate of delay at the junction (Allsop, 1970, 1971) in the absence of priority vehicles.

The data for this example were chosen to be straightforward. The arrival process in each stream is taken to be simple Poisson, so the indexes of dispersion are all equal to unity. The effective green time for each stream is taken to exceed the indicated green time by 1 second and all green indications are supposed to start and stop exactly at the beginning.
Table 6.1a: Signal-settings used for the Derby junction example (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage durations</th>
<th>Transition period following</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>uninterrupted</td>
<td>minimum</td>
</tr>
<tr>
<td>1</td>
<td>16.0</td>
<td>7.0</td>
</tr>
<tr>
<td>2</td>
<td>24.0</td>
<td>7.0</td>
</tr>
<tr>
<td>3</td>
<td>28.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>

duration of an extension period: 8 s

Table 6.1b: Stream data used for the Derby junction example

<table>
<thead>
<tr>
<th></th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean arrival rate (veh/h)</td>
<td>600</td>
<td>1481</td>
<td>835</td>
<td>808</td>
<td>623</td>
<td>941</td>
</tr>
<tr>
<td>Index of dispersion of arrivals</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Saturation departure rate (veh/h)</td>
<td>4000</td>
<td>3462</td>
<td>2941</td>
<td>3462</td>
<td>2830</td>
<td>3529</td>
</tr>
<tr>
<td>Coefficient of variation of</td>
<td>.41</td>
<td>.35</td>
<td>.32</td>
<td>.37</td>
<td>.39</td>
<td>.32</td>
</tr>
<tr>
<td>departure times</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional green time (s)</td>
<td>1.0</td>
<td>6.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>First stage for green indication</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Last stage for green indication</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
and end of a stage. The durations of all the transition periods are taken to be equal to 5.0 seconds.

6.2.2 Capacity-equivalent signal-settings.

In this sub-section, the results of Chapter 4 are applied to the Derby junction described in sub-section 6.2.1. Calculations are performed for each of five different combinations of priority rules. These combinations are extension, extension and hurry-call, extension and recall, extension and hurry-call with inhibition, and extension and recall with inhibition. The calculations for extension are elementary and are not specific to the particular junction. The other calculations use the information given in Table 6.1a but not that given in Table 6.1b.

According to the result of lemma 4.35, when priority is implemented by extension, only the duration of stage k need be adjusted in order to maintain capacity-equivalence as the mean bus arrival rate varies. The correction to the duration of stage k required to achieve this is to subtract the mean time for which stage k is extended from its original duration. Thus the only information required to achieve capacity-equivalence when priority is implemented by extension is the mean extension to stage k. This can be calculated using Equation (4.50) from the mean bus arrival rate and the duration of a single extension period. Figure 6.2 shows how this quantity varies with the mean bus arrival rate for a range of priority extension periods.

When priority by extension and hurry-call is implemented, according to the result of theorem 4.38, capacity-equivalent signal-settings, if they exist, can be calculated directly from (4.102). The range of mean bus arrival rates for which solutions to (4.102) exist can be found from the result of corollary 4.42. In the case of the Derby junction, the least
upper bound on the range is 66.7 buses per hour. The factors by which the stage durations given in Table 6.1a must be multiplied in order to achieve capacity-equivalence as the mean bus arrival rate varies are shown in Figure 6.3.

When priority by extension and recall is implemented, according to the result of theorem 4.46, capacity-equivalent signal-settings, if they exist, satisfy (4.127). According to the result of lemma 4.50, the range of mean bus arrival rates for which solutions to (4.127) exist is bounded above by the solution to (4.136). In this case, the upper bound is 73.3 buses per hour. Figure 6.4 shows the factors by which the stage durations given in Table 6.1a must be multiplied in order to achieve capacity-equivalence as the mean bus arrival rate varies.

The effect of using the inhibition rule in conjunction with either of the last two combinations of priority rules is to extend considerably the range of mean bus arrival rates for which acceptable signal-settings exist. According to the result of theorem 4.55, capacity-equivalent signal-settings satisfy the appropriate form of (4.148). The most restrictive constraint upon the range of bus arrival rates for which acceptable solutions exist is determined by the minimum permissible duration of stage 1. When priority by extension and hurry-call with inhibition is implemented, the upper bound calculated according to the result of corollary 4.36 is 610 buses per hour. The corresponding value for priority by extension and recall with inhibition is 483 buses per hour. Figures 6.5 and 6.6 show the factors by which the stage durations given in Table 6.1a must be multiplied in order to achieve capacity-equivalence for each of these combinations of priority rules as the mean bus arrival rate varies.
Figure 6.2: Priority by extension

Mean extension to stage $k$
Figure 6.3: Priority by extension and hurry-call
capacity-equivalent signal-settings for Derby junction.
Figure 6.4: Priority by extension and recall
Capacity-equivalent signal-settings for Derby junction
Figure 6.5: Priority by extension and hurry-call with inhibition
Capacity-equivalent signal-settings for Derby junction
Figure 6.6: Priority by extension and recall with inhibition
Capacity-equivvalent signal-settings for Derby junction
6.2.3 Estimates of delay.

In this sub-section, the results of the calculations described in Chapter 5 are presented. These calculations are performed for each of priority by extension and recall and priority by extension and recall with inhibition for a range of mean bus arrival rates. Two sets of calculations are presented for each combination of priority rules. The first of these shows the effect of variations in the mean bus arrival rate while the signal-settings are held fixed. The second set of calculations shows the effect of using signal-settings calculated according to the methods described in Chapter 4 and illustrated in sub-section 6.2.2.

Suppose that priority by extension and recall is implemented for buses arriving at a mean rate of 20 per hour. The signal-settings given in Table 6.2a are calculated from the result of theorem 4.46 so that when they are implemented in these circumstances they are capacity-equivalent to those given in Table 6.1a in the absence of priority. The mean delays incurred by non-priority vehicles in each of the six streams at the junction can be estimated from the results of theorems 5.8, 5.13, 5.20, 5.21 and 5.23. The mean delay incurred by priority vehicles can be estimated from the result of theorem 5.37. The results of these calculations for a mean bus arrival rate of 20 buses per hour when the signal-settings in Table 6.2a are implemented are given in Table 6.2b.

In order to assess the consequences of providing priority, the mean delays incurred by vehicles in each of the various categories in the absence of priority are given in Table 6.2c. These figures were calculated using the same total flow in each stream as for the calculations which lead to the figures in Table 6.2b. Thus the total flow in stream 1 is 620 vehicles per hour, comprising 600 non-priority vehicles and 20 buses per hour; all other flows used are as given in Table 6.1b. The signal-settings used are those given in Table 6.1a.
Table 6.2a: Signal-settings which, when implemented with priority by extension and recall for buses arriving at a mean rate of 20 per hour, are capacity-equivalent to those given in Table 6.1a (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage duration (uninterrupted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.01</td>
</tr>
<tr>
<td>2</td>
<td>24.71</td>
</tr>
<tr>
<td>3</td>
<td>35.11</td>
</tr>
</tbody>
</table>

Table 6.2b: Estimates of mean overflow and delay when priority by extension and recall is provided for buses arriving at a mean rate of 20 per hour and the signal-settings in Table 6.2a are implemented

<table>
<thead>
<tr>
<th></th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>2.21</td>
<td>0.19</td>
<td>4.77</td>
<td>2.15</td>
<td>1.90</td>
<td>7.00</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>44.42</td>
<td>11.58</td>
<td>48.36</td>
<td>35.90</td>
<td>35.93</td>
<td>52.52</td>
<td>5.48</td>
</tr>
</tbody>
</table>

Table 6.2c: Estimates of mean overflow and delay in the absence of priority when the mean bus arrival rate is 20 per hour and the signal-settings in Table 6.1a are implemented

<table>
<thead>
<tr>
<th></th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>1.31</td>
<td>0.85</td>
<td>1.87</td>
<td>0.63</td>
<td>1.04</td>
<td>3.73</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>38.13</td>
<td>12.66</td>
<td>31.66</td>
<td>25.18</td>
<td>31.26</td>
<td>41.07</td>
<td>38.19</td>
</tr>
</tbody>
</table>
The estimate of the mean delay incurred by buses in the absence of priority calculated using (5.137) is 38.19 seconds; this differs from the 38.13 seconds estimated for non-priority vehicles using (5.50). The difference of 0.06 seconds arises from the simplifications used in the analysis of delay incurred by buses in the absence of priority. This was discussed in sub-section 5.4.2. The estimated reduction in the mean delay incurred by buses as a result of implementing priority is 32.71 seconds, corresponding to a reduction in the rate of delay by 0.1817 vehicles. This is achieved at the expense of an increase in the rate of delay incurred by non-priority traffic of 10.47 vehicles from 40.92 to 51.39 vehicles. Thus in order for the implementation of priority to result in a net saving in passenger delay, the ratio of the mean occupancy of priority to non-priority vehicles must exceed 57.62.

If the mean bus arrival rate differs from 20 per hour, then the estimates of delay will vary from the values given in Table 6.2b. Figure 6.7 shows the factors by which the delays given in Table 6.2b must be multiplied in order to estimate the mean delays as the mean bus arrival rate varies but the signal-settings remain fixed at the values given in Table 6.2a. This graph indicates the consequences of variations in the mean bus arrival rate from that for which the priority system was designed.

Suppose now that priority by extension and recall with inhibition is implemented for buses arriving at a mean rate of 20 per hour. The signal-settings given in Table 6.3a are those calculated according to the result of theorem 4.55 so that when implemented in these circumstances, they are capacity-equivalent to those given in Table 6.1a in absence of priority. Estimates of the corresponding mean delays, calculated from the results of theorems 5.8, 5.13, 5.27, 5.28 and 5.38 and lemma 5.25 are given in Table
Table 6.3a: Signal-settings which, when implemented with priority by extension and recall with inhibition for buses arriving at a mean rate of 20 per hour, are capacity-equivalent to those given in Table 6.1a (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage duration (uninterrupted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.20</td>
</tr>
<tr>
<td>2</td>
<td>24.53</td>
</tr>
<tr>
<td>3</td>
<td>32.86</td>
</tr>
</tbody>
</table>

Table 6.3b: Estimates of mean overflow and delay when priority by extension and recall with inhibition is provided for buses arriving at a mean rate of 20 per hour and the signal-settings in Table 6.3a are implemented

<table>
<thead>
<tr>
<th></th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>1.92</td>
<td>0.33</td>
<td>3.34</td>
<td>1.40</td>
<td>1.83</td>
<td>6.43</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>42.36</td>
<td>11.82</td>
<td>41.08</td>
<td>31.36</td>
<td>35.39</td>
<td>50.21</td>
<td>11.76</td>
</tr>
</tbody>
</table>
6.3b. Figure 6.8 shows the factors by which the delays given in Table 6.3b must be multiplied in order to estimate the mean delays as the mean bus arrival rate varies but the signal-settings remain fixed at the values given in Table 6.3a.

The next calculations presented here illustrate the effect of variations in the mean arrival rate of buses for which priority is provided when the signal-settings used are adjusted according to the methods developed in Chapter 4. The signal-settings appropriate for a mean arrival rate of 0 buses per hour are those given in Table 6.1a: the corresponding estimates of delay are given in Table 6.4. The value of 4.51 given for the mean delay incurred by priority vehicles is the limiting value as the mean bus arrival rate tends to 0 when priority is provided by either extension and recall or extension and recall with inhibition. The same values apply to both of these cases since in this limiting case there is no disruption to the normal operation of the traffic signals and the probability that inhibition is in effect 0.

If priority by extension and recall is provided and the stage durations used are adjusted from those given in Table 6.1a according to the factors indicated in Figure 6.4, then the estimates of delay differ from those given in Table 6.4 by the factors indicated by Figure 6.9. If the inhibition rule is also used and the stage durations used are adjusted according to the factors indicated in Figure 6.6, then the estimates of delay vary according to the factors indicated by Figure 6.10.

Figures 6.9 and 6.10 show how the mean delays incurred by vehicles in each of the categories vary with the mean bus arrival rate when the signal-settings are adjusted accordingly. These results can be used together with estimates of the mean delays incurred in the absence of priority to estimate the minimum ratio of mean occupancy of priority to non-priority
Table 6.4: Estimates of mean overflow and delay when the mean bus arrival rate tends to 0 per hour and the signal-settings in Table 6.1a are implemented

<table>
<thead>
<tr>
<th></th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>1.09</td>
<td>0.85</td>
<td>1.87</td>
<td>0.63</td>
<td>1.04</td>
<td>3.73</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>36.92</td>
<td>12.66</td>
<td>31.66</td>
<td>25.18</td>
<td>31.26</td>
<td>41.07</td>
<td>4.51</td>
</tr>
</tbody>
</table>

Mean delay incurred by priority vehicles in the absence of priority: 36.98 s
Key:

- Stream 1
- Stream 2
- Stream 3
- Stream 4
- Stream 5
- Stream 6
- Priority vehicles

Figure 6.7:
Priority by extension and recall
Delay estimates for Derby junction
with fixed signal-settings
Key:

- Stream 1
- Stream 2
- Stream 3
- Stream 4
- Stream 5
- Stream 6
- Priority vehicles

Figure 6.8:
Priority by extension and recall with inhibition
Delay estimates for Derby junction
with fixed signal-settings
Figure 6.9:

Priority by extension and recall

Delay estimates for Derby junction

with capacity-equivalent signal-settings
Figure 6.10:

Priority by extension and recall with inhibition

Delay estimates for Derby junction

with capacity-equivalent signal-settings
Figure 6.11: Minimum ratio of priority to non-priority vehicle occupancy to give a net reduction in passenger delay, Derby junction.
vehicles required for the implementation of responsive priority to give a net saving in passenger delay. The way in which this ratio varies with the mean bus arrival rate when the signal-settings are adjusted accordingly is shown in Figure 6.11 for both of the combinations of priority rules considered above.

6.3 The Chapel Hill junction

6.3.1 Introduction.

The second junction for which example calculations are given is based on an example used by Tully (1976). This represents part of the junction of Chapel Hill with the inner ring road in Huddersfield. Figure 6.12 depicts the junction as it is used here. The buses for which priority is provided travel in stream 2 and have right of way during the first two of the four stages in the sequence.

The numerical data used as the basis for the calculations are given in Tables 6.5. The sequence used is one found by Tully (1976) to be suitable for use at the junction and comprises 4 stages. Signal-settings were calculated according to the method described by Allsop (1972b) and discussed in sub-section 4.3.2 to maximise the reserve capacity at the junction when the cycle-time is specified as 100s. These signal-settings were then rounded to the nearest second to give those in Table 6.5a. The data given in Table 6.5b for the streams were devised for this example to illustrate several of the possibilities for which the analysis presented in Chapters 4 and 5 can cater. Various durations are used for the transition periods and the streams experience different start and end lags. Two sets of additional green times are given depending on whether there are 3 or 4 stages in a particular cycle. A variety of indexes of dispersion are used for the arrival patterns in the streams.
The results presented in this section follow the general pattern of those presented in Section 6.2. However, no results are given for priority by extension alone: this priority method is omitted here since Figure 6.2 is appropriate for any junction. Because of the similarity between the analyses of the two examples, fewer details of the calculations are given in this section. Each graph in this section uses axes which are identical to those of the corresponding graph in Section 6.2.

6.3.2 Capacity-equivalent signal-settings.

In this sub-section, the methods described in Chapter 4 are used to calculate signal-settings which are suitable for use with responsive priority. These calculations are performed for four combinations of priority rules, namely extension with each of hurry-call and recall both with and without inhibition. Use is made in this sub-section of the information given in Table 6.5a but not of that given in Table 6.5b.

When priority by extension and hurry-call is implemented, the results of sub-section 4.5.4 can be applied to calculate signal-settings which are capacity-equivalent to those given in Table 6.5a, where they exist. In this case, the least upper bound on the mean bus arrival rate for which these signal-settings exist is 70.4 per hour. The factors by which the stage durations given in Table 6.5a must be multiplied in order to achieve capacity-equivalence as the mean bus arrival rate varies are shown in Figure 6.13.

If priority by extension and recall is implemented, then the results of sub-section 4.5.5 are appropriate. In this case, the upper bound on the mean bus arrival rate for which capacity-equivalent signal-settings exist is 77.7 per hour. The factors by which the stage durations given in Table 6.5a must be multiplied in order to achieve capacity-equivalence as the mean bus arrival rate varies are shown in Figure 6.14.
Table 6.5a: Signal-settings used for the Chapel Hill junction example (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage durations uninterrupted</th>
<th>Stage durations minimum</th>
<th>Transition periods following 4-stage cycle</th>
<th>Transition periods following 3-stage cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.0</td>
<td>5.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>2</td>
<td>10.0</td>
<td>0.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>18.0</td>
<td>6.5</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>32.0</td>
<td>7.0</td>
<td>4.0</td>
<td>--</td>
</tr>
</tbody>
</table>

Duration of an extension period: 10 s.

Table 6.5b: Stream data used for the Chapel Hill junction example

<table>
<thead>
<tr>
<th>Mean arrival rate (veh/h)</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>1000</td>
<td>600</td>
<td>900</td>
<td>1100</td>
<td>600</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Index of dispersion of arrivals</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>1.0</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saturation departure rate (veh/h)</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3600</td>
<td>4000</td>
<td>3600</td>
<td>3000</td>
<td>4000</td>
<td>3000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficient of variation of departure times</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33</td>
<td>0.22</td>
<td>0.39</td>
<td>0.45</td>
<td>0.22</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional green time in 4-stage cycle (s)</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>5.0</td>
<td>1.5</td>
<td>3.0</td>
<td>4.0</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional green time in 3-stage cycle (s)</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
<td>--</td>
<td>5.0</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First stage for green indication</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last stage for green indication</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.13: Priority by extension and hurry-call
Capacity-equivalent signal-settings for Chapel Hill junction
Figure 6.14: Priority by extension and recall

Capacity-equivalent signal-settings for Chapel Hill junction
Figure 6.15: Priority by extension and hurry-call with inhibition
Capacity-equivalent signal-settings for Chapel Hill junction
Figure 6.16: Priority by extension and recall with inhibition
Capacity-equivalent signal-settings for Chapel Hill junction
The effect of using the inhibition rule with each of these combinations or priority rules is illustrated in Figures 6.15 and 6.16. The data for these graphs were calculated according to the methods developed in sub-section 4.5.6. Again, these graphs show the factors by which the stage durations given in Table 6.5a must be multiplied in order to achieve capacity-equivalence as the mean bus arrival rate varies.

6.3.3 Estimates of delay.

Suppose that priority by extension and recall is implemented for buses arriving at a mean rate of 20 per hour. The signal-settings given in Table 6.6a are calculated from the result of theorem 4.46 so that when they are implemented in these circumstances they are capacity-equivalent to those given in Table 6.5a. The resulting estimates of delay for the various categories of vehicles are given in Table 6.6b. These can be compared with the estimates given in Table 6.6c for the mean delay incurred by vehicles in each of the various categories with the same flows but in the absence of priority and when the signal-settings given in Table 6.5a are implemented. Figure 6.17 shows the effect of variations in the mean bus arrival rate on these estimates of delay when the signal-settings remain fixed at the values given in Table 6.6a.

Suppose now that priority by extension and recall with inhibition is provided for buses arriving at a mean rate of 20 per hour. Signal-settings which are appropriate for these circumstances can be calculated from the results given in sub-section 4.5.6 and are given in Table 6.7a. The resulting estimated delay for the various categories of vehicles are given in Table 6.7b. Figure 6.18 shows the effect of variations in the mean bus arrival rate on these estimates of delay when the signal-settings remain fixed at the values given in Table 6.7a.
Table 6.6a: Signal-settings which, when implemented with priority by extension and recall for buses arriving at a mean rate of 20 per hour, are capacity-equivalent to those given in Table 6.5a (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage duration (uninterrupted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.98</td>
</tr>
<tr>
<td>2</td>
<td>9.27</td>
</tr>
<tr>
<td>3</td>
<td>18.15</td>
</tr>
<tr>
<td>4</td>
<td>38.78</td>
</tr>
</tbody>
</table>

Table 6.6b: Estimates of mean overflow and delay when priority by extension and recall is provided for buses arriving at a mean rate of 20 per hour and the signal-settings in Table 6.6a are implemented

<table>
<thead>
<tr>
<th>Stream</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>4.13</td>
<td>1.30</td>
<td>4.38</td>
<td>7.29</td>
<td>4.60</td>
<td>0.87</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>52.17</td>
<td>29.22</td>
<td>62.29</td>
<td>60.56</td>
<td>44.52</td>
<td>35.35</td>
<td>4.30</td>
</tr>
</tbody>
</table>

Table 6.6c: Estimates of the mean overflow and delay in the absence of priority when the mean bus arrival rate is 20 per hour and the signal-settings in Table 6.5a are implemented

<table>
<thead>
<tr>
<th>Stream</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>2.53</td>
<td>0.50</td>
<td>2.87</td>
<td>3.05</td>
<td>2.51</td>
<td>0.29</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>44.71</td>
<td>25.99</td>
<td>54.31</td>
<td>40.08</td>
<td>38.20</td>
<td>28.30</td>
<td>26.11</td>
</tr>
</tbody>
</table>
Suppose now that as the mean bus arrival rate varies, the signal-settings used are adjusted in accordance with the methods developed in Chapter 4. The signal-settings appropriate for a mean arrival rate of 0 buses per hour are those given in Table 6.5a; the corresponding estimates of delay are given in Table 6.8. As before, this table is appropriate for both combinations of priority methods considered in this sub-section. Figures 6.19 and 6.20 show the effects of variations in the mean bus arrival rate on the estimates of delay given in Table 6.8 when the stage durations implemented are adjusted from those given in Table 6.5a by the factors indicated by Figures 6.14 and 6.16 respectively. The minimum ratio of mean occupancy of priority to non-priority vehicles required for the implementation of responsive priority with capacity-equivalent signal-settings to give a net saving in passenger delay is shown in Figure 6.21.

In order to illustrate the degree to which differences from simple Poisson arrivals and regular departures effect the mean delay incurred by vehicles in each category, the sensitivity of the estimates given by (5.50), (5.68), (5.156) and (5.159) to changes in the values of $I_a$ and $C_s$ are examined. Inspection of (5.50) and (5.68) shows that for any given mean bus arrival rate, signal-settings and combination of priority rules, the mean delay incurred by non-priority vehicles in each stream varies linearly with $I_a$ and $C_s^2$. Thus in these circumstances, the mean delay incurred by non-priority vehicles in a stream can be expressed in the form

$$d = A + A_I I_a + A_c C_s^2$$

(6.1)

for some constant coefficients $A$, $A_I$ and $A_c$ and for arbitrary values of $I_a \geq 1$ and $C_s^2 \geq 0$. Similarly, the mean delay incurred by priority vehicles can be expressed in a similar form where $I_a$ and $C_s^2$ take the values appropriate to the stream in which priority vehicles travel. Values of the
Table 6.7a: **Signal-settings which, when implemented with priority by extension and recall with inhibition for buses arriving at a mean rate of 20 per hour, are capacity-equivalent to those given in Table 6.5a (seconds)**

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage duration (uninterrupted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.21</td>
</tr>
<tr>
<td>2</td>
<td>9.37</td>
</tr>
<tr>
<td>3</td>
<td>18.11</td>
</tr>
<tr>
<td>4</td>
<td>36.79</td>
</tr>
</tbody>
</table>

Table 6.7b: **Estimates of mean overflow and delay when priority by extension and recall with inhibition is provided for buses arriving at a mean rate of 20 per hour and the signal-settings in Table 6.7a are implemented**

<table>
<thead>
<tr>
<th>Stream</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>3.55</td>
<td>1.11</td>
<td>4.15</td>
<td>5.22</td>
<td>4.23</td>
<td>0.53</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>49.39</td>
<td>28.36</td>
<td>60.95</td>
<td>51.92</td>
<td>43.26</td>
<td>32.37</td>
<td>8.57</td>
</tr>
</tbody>
</table>
Figure 6.17:  
Priority by extension and recall  
Delay estimates for Chapel Hill junction  
with fixed signal-settings
Figure 6.18:

Priority by extension and recall with inhibition

Delay estimates for Chapel Hill junction

with fixed signal-settings
Key:
- + - Stream 1
- o - Stream 2
- * - Stream 3
- x - Stream 4
- A - Stream 5
- V - Stream 6
- - - Priority vehicles

Figure 6.19:
Priority by extension and recall
Delay estimates for Chapel Hill junction
with capacity-equivalent signal-settings
Figure 6.20:

Priority by extension and recall with inhibition
Delay estimates for Chapel Hill junction
with capacity-equivalent signal-settings
Figure 6.21: Minimum ratio of priority to non-priority vehicle occupancy to give a net reduction in passenger delay, Chapel Hill junction.
Table 6.8: Estimates of mean overflow and delay when the mean bus arrival rate tends to 0 per hour and the signal-settings in Table 6.5a are implemented.

<table>
<thead>
<tr>
<th>Stream</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean overflow</td>
<td>2.53</td>
<td>0.45</td>
<td>2.87</td>
<td>3.05</td>
<td>2.51</td>
<td>0.29</td>
<td>--</td>
</tr>
<tr>
<td>Mean delay (s)</td>
<td>44.71</td>
<td>25.69</td>
<td>54.31</td>
<td>40.08</td>
<td>38.20</td>
<td>28.30</td>
<td>3.62</td>
</tr>
</tbody>
</table>

Mean delay incurred by priority vehicles in the absence of priority: 25.80 s.
Table 6.9a: Values of the coefficients in Equation (6.1) when priority is provided by extension and recall for buses arriving at a mean rate of 20 per hour and the signal-settings given in Table 6.6a are implemented (seconds)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>40.40</td>
<td>27.46</td>
<td>45.12</td>
<td>47.83</td>
<td>36.79</td>
<td>33.23</td>
<td>4.08</td>
</tr>
<tr>
<td>A_I</td>
<td>9.71</td>
<td>1.64</td>
<td>12.51</td>
<td>8.99</td>
<td>6.69</td>
<td>1.79</td>
<td>0.20</td>
</tr>
<tr>
<td>A_C</td>
<td>10.88</td>
<td>2.36</td>
<td>14.44</td>
<td>9.72</td>
<td>7.37</td>
<td>3.29</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Table 6.9b: Values of the coefficients in Equation (6.1) when priority is provided by extension and recall with inhibition for buses arriving at a mean rate of 20 per hour and the signal-settings given in Table 6.7a are implemented (seconds)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Stream 1</th>
<th>Stream 2</th>
<th>Stream 3</th>
<th>Stream 4</th>
<th>Stream 5</th>
<th>Stream 6</th>
<th>Priority vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>36.58</td>
<td>25.86</td>
<td>41.87</td>
<td>38.53</td>
<td>34.75</td>
<td>28.75</td>
<td>8.14</td>
</tr>
<tr>
<td>A_I</td>
<td>9.61</td>
<td>1.64</td>
<td>12.49</td>
<td>8.94</td>
<td>6.77</td>
<td>1.80</td>
<td>0.22</td>
</tr>
<tr>
<td>A_C</td>
<td>10.78</td>
<td>2.38</td>
<td>14.42</td>
<td>9.67</td>
<td>7.46</td>
<td>3.29</td>
<td>0.41</td>
</tr>
</tbody>
</table>
coefficients $A$, $A_I$, and $A_C$ are given for each category of vehicles at the Chapel Hill junction in Tables 6.9. These figures are calculated for a mean bus arrival rate of 20 per hour, signal-settings given in Tables 6.6a and 6.7a and priority by extension and recall and by extension and recall with inhibition respectively.

6.4 Discussion

6.4.1 Introduction.

In this section, the results presented in the last two sections are compared and discussed. Some similarities between the two examples can be attributed to behaviour which will occur in most cases. However, many of the details of the behaviour are specific to the example junctions considered. The salient features of the examples are classified in accordance with this.

6.4.2 Capacity-equivalent signal-settings.

When priority by extension is implemented, the amount by which the duration of stage $k$ must be reduced in order to achieve capacity-equivalence is illustrated in Figure 6.2. The size of this reduction depends only on the mean bus arrival rate and the duration of an extension period. Thus the results depicted in Figure 6.2 can be applied quite generally.

The graphs in Figure 6.2 show relatively little curvature. This is because the first (linear) term in the power-series expansion in $\beta$ (4.97) for the compensated duration of stage $k$ is the largest one wherever $\beta \leq 3$. The point in Figure 6.2 with the largest value of $\beta$ is that corresponding to 100 buses per hour and an extension period of duration 30
seconds, giving $\beta t = 0.83$. Thus in the range illustrated, the first order or linear approximation $[\exp(\beta t) - (1 + \beta t)]/\beta \approx \beta t^2/2$ would yield fair estimates for the correction required.

When priority by extension and recall or extension and hurry-call is implemented, the graphs of the signal-settings required to achieve capacity-equivalence are similar in character. This is illustrated for the two example junctions by Figures 6.3-4 and 6.13-14. While the details differ in each of these cases, the behaviour is broadly similar.

In all of the cases illustrated, the correction factor for the duration of stage $m$ dominates those for the other stages at all bus arrival rates. In both of the examples considered, the capacity-equivalent stage durations are larger for any given mean bus arrival rate when there is priority by extension and hurry-call than when there is priority by extension and recall. However, these details of behaviour will not occur universally. Consider a junction for which $\lambda_m^* = \gamma_m$ and $m \geq k + 1$. According to the result of Theorem 4.38, when priority by extension and hurry-call is implemented, the duration of stage $m$ for which capacity-equivalence is achieved satisfies $\lambda_m^* = \gamma_m$ for all mean bus arrival rates. If $\lambda_m^* > \gamma_m$, then the correction factor for the duration of stage $m$ will dominate those for the others as the mean bus arrival rate increases towards the least upper bound for which capacity-equivalence can be achieved. Furthermore, the correction factor for the duration of stage $m$ when priority is implemented by extension and recall will dominate that for priority by extension and hurry-call.

In each of these four cases, the correction factor for the duration of stage $k$ is the smallest one. This will always happen since stage $k$ is the only one for which the mean duration exceeds the undisturbed one.
Now consider the effect of introducing the inhibition rule in each of these four cases. The factors by which each of the stage durations must be multiplied in order to achieve capacity-equivalence are shown in Figures 6.5-6 and 6.15-16. Again, these graphs have many features in common.

As occurred in the absence of inhibition, in each of the cases considered, the correction factor for stage $m$ dominates those for the other stages, the smallest of which is that for stage $k$. When the inhibition rule is used, the range of the correction factors at any given mean bus arrival rate is considerably reduced. The limiting values for the correction factors for stages other than $k$ can be found from the results of lemmas 4.56 and 4.57. As before, the correction factors for priority by extension and hurry-call with inhibition are larger than the corresponding ones for priority by extension and recall with inhibition. Again, this behaviour would not occur in all cases but rather is dependent upon details of the examples under consideration.

6.4.3 Estimates of delay.

The estimates of delay given in sub-sections 6.2.3 and 6.3.3 confirm that responsive priority can give rise to considerable reductions in the mean delay incurred by priority vehicles in a variety of circumstances. Since the estimates of delay incurred by other traffic are continuous functions of the mean bus arrival rate, they will be subject only to small changes when priority is implemented for buses arriving at a sufficiently low rate.

The graphs shown in Figures 6.7-8 and 6.17-18 show the effects of variations in the mean bus arrival rate on the estimates of delay when the signal-settings used remain fixed. These graphs illustrate the consequences of variations in the mean bus arrival rate from that for which the priority schemes were designed: in all four cases this rate is 20 buses per hour.
In each of these cases, the mean delay incurred by non-priority traffic travelling in the same stream as the priority vehicles is reduced by any increase in the mean arrival rate of priority vehicles. This shows that some non-priority traffic would benefit from the implementation of responsive priority if the signal-settings were not adjusted. On the other hand, in each of these circumstances, some streams experience increases in delay for any increase in the mean arrival rate of priority vehicles. This is most pronounced when priority is implemented without the inhibition rule.

When priority by extension and recall is implemented, traffic in any stream which has right of way only between the end of stage \( k+1 \) and the end of stage \( m \) will necessarily incur delays which increase without bound as the mean bus arrival rate increases towards some finite value. This is a consequence of the result of lemma 4.24 which shows that the limit of the expected duration of each stage \( i \) in the range \( k+1 \leq i \leq m \) as the mean bus arrival rate increases without bound is 0. Thus adequate capacity will be provided for streams such as those described above only in some bounded range of mean bus arrival rates. Any other stream which has right of way during the period described above and at other times may or may not receive adequate capacity in the limit as the mean bus arrival rate increases without bound: this depends upon whether or not the appropriate limiting value of the capacity of the stream exceeds the mean arrival rate of traffic in that stream.

In the case of the Derby junction, the delay incurred by vehicles travelling in stream 3 increases most rapidly with the mean bus arrival rate when either of the combinations of priority rules is used. The effect of a change in the mean bus arrival rate on the delay incurred by vehicles
in stream 4 is similar but less pronounced. Inspection of the data given in Table 6.1 shows that these two streams receive right of way only during the last stage of the cycle. The ratio of the mean arrival rate to the saturation departure rate is 0.284 for stream 3 and 0.233 for stream 4: since both streams have right of way for the same amount of time in any cycle, stream 3 is more heavily loaded than stream 4. Thus the behaviour of the mean delays incurred by vehicles in these streams is as would be expected: they both increase with the mean bus arrival rate and that for the more heavily loaded stream does so most. Similar observations apply to streams 4 and 6 of the Chapel Hill junction.

Both of the examples show that if the inhibition rule is used, then the mean delay incurred by priority vehicles increases with their mean arrival rate. This is to be expected since in these circumstances, the probability that priority by recall is inhibited when a bus arrives at the detector increases with the mean bus arrival rate. By contrast, when the inhibition rule is not used, the mean delay incurred by priority vehicles decreases with their mean arrival rate. This too is to be expected since here the probability that priority will not be required by a bus increases with the mean bus arrival rate.

The graphs shown in Figures 6.9-10 and 6.19-20 show the effects of variations in the mean bus arrival rate on the estimates of mean delay when the signal-settings used are adjusted according to the methods of Chapter 4 to take this into account. These graphs show how the effects of implementing a priority scheme vary with the mean arrival rate of the vehicles for which priority is provided.

The graphs show that in these circumstances, the delay incurred by traffic in most streams increases with any increase in the mean bus arrival rate. The single exception in these examples is stream 2 at the Derby
Table 6.10: Estimates of the mean delay incurred by priority vehicles arriving at a mean rate of 20 per hour when the signal-settings used are not compensated (seconds)

<table>
<thead>
<tr>
<th>Priority method</th>
<th>Junction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Derby</td>
</tr>
<tr>
<td>Extension and recall</td>
<td>4.24</td>
</tr>
<tr>
<td>Extension and recall with inhibition</td>
<td>9.76</td>
</tr>
</tbody>
</table>
junction. This stream has right of way during stages 3 and 1 so will have an effective green period in every cycle. In some cycles where priority is granted by recall, the duration of the effective red period for this stream will be reduced albeit at the expense of a reduction in the duration of the next effective green period. In cycles where no recall is granted, this stream will benefit from an effective green period which is extended by an amount corresponding to the correction to the duration of stage 3. There is no clear indication of whether or not this behaviour is typical of a stream in Class 5 or 6.

In all four of these cases, the mean delay incurred by priority vehicles increases with their mean arrival rate. Estimates are given in Table 6.10 for the mean delay which would be incurred by priority vehicles arriving at a mean rate of 20 per hour if the signal-settings given in Tables 6.1a and 6.5a were implemented. These estimates can be compared with those given in Tables 6.2b, 6.3b, 6.6b and 6.7b which are calculated for circumstances which differ only in that the signal-settings used have been compensated. This shows that compensating the signal-settings to achieve capacity-equivalence causes an increase in the mean delay incurred by priority vehicles.

The value of the mean delay incurred by priority vehicles is particularly sensitive to their mean arrival rate when the inhibition rule is used in conjunction with the compensated signal-settings. In both of the examples considered, when the mean arrival rate of priority vehicles is 60 per hour, the mean delay incurred by them is over 4 times that experienced at low arrival rates and is a little over half of that incurred in the absence of priority.

The results presented in Tables 6.9 quantify the sensitivity of the estimates of the mean delay incurred by vehicles in each category at the
Chapel Hill junction to changes in the variability of the arrival and departure patterns of non-priority vehicles. The values given for the coefficients of $I_a$ and $C_s^2$ show considerable variation between different streams. Thus the estimates of mean delay for some categories are rather more sensitive to the values of these parameters than are others. The values for stream 3 are as large as 25 and 30 per cent of that of the constant term so these kinds of variability can contribute significantly to the mean delay incurred by vehicles in this stream.

There is little difference between corresponding coefficients of $I_a$ and $C_s^2$ in Tables 6.9a and 6.9b. Inspection of (5.50) shows that the explicit coefficients of $I_a$ and $C_s^2$ in that expression are respectively

$$\frac{1}{2} \left( \frac{s}{s-kq} \left( \frac{r}{s-kq} \left( \frac{E(r)}{E(r)+E(g)} \right) \right) \right) \quad \text{and} \quad \frac{1}{2} \left( \frac{kq}{s-kq} \left( \frac{r}{s-kq} \left( \frac{E(r)}{E(r)+E(g)} \right) \right) \right)$$

Since the signal-settings used with each of the combinations of priority rules are devised so that the ratio $E(r)/[E(r)+E(g)]$ is similar in each case, the values of these coefficients will be similar. Thus the differences between the coefficients in Table 6.9a and those in Table 6.9b are largely attributable to differences in the sensitivity of the overflow in the two cases.

---

If the implementation of a priority scheme is to be justified solely on the grounds of a net reduction in passenger delay, then the decrease in the rate of delay for the priority vehicles multiplied by their mean occupancy should exceed the increase in that for non-priority vehicles. Thus whether or not a priority scheme could be justified in this way depends critically on the ratio of the mean occupancies of vehicles in these two categories. Figures 6.11 and 6.21 show how the minimum ratio which would enable this justification varies with the mean bus arrival rate: both combinations of priority rules are considered and in all cases the signal-settings are supposed to be adjusted according to the methods developed in Chapter 4.
The values of the occupancy ratios required are large in all of the circumstances considered: for the Derby junction it is never less than 40 and for the Chapel Hill junction it is never less than 75. None of these priority schemes is likely to be justified on these grounds alone. However, the review and discussion presented in Section 2.2 shows that there are many benefits of priority schemes besides the reduction in total delay incurred by bus passengers. These include improved regularity of journey times for bus passengers, reductions in the mean time spent by passengers waiting at bus stops and reduced costs for the bus operators. Justification for responsive priority at either of the example junctions would depend to some extent upon considerations such as these.

The graphs shown in Figures 6.11 and 6.21 for priority by extension and recall show that in both examples any increase in the mean bus arrival rate causes an increase in the occupancy ratio required for priority to give a net reduction in passenger delay. In both cases, this ratio is nearly twice as large when the mean bus arrival rate is 60 per hour as it is in the limit as the mean arrival rate tends to 0.

By contrast, when priority is provided by extension and recall with inhibition, the required occupancy ratio is fairly stable over the range of mean bus arrival rates considered. Indeed, in the case of the Derby junction, it is somewhat smaller when the mean bus arrival rate is as high as 60 per hour than it is in the limit as the mean arrival rate tends to 0. Thus despite the marked increase in the mean delay incurred by individual buses as their mean arrival rate increases, which is illustrated in Figures 6.10 and 6.20, the reduction rate of delay for buses increases with their mean arrival rate.
CHAPTER 7

SUGGESTIONS FOR FURTHER INVESTIGATION

The material presented in this thesis is restricted in several respects. By their very nature, these restrictions suggest topics for further investigation. Some of the methods developed here for the investigation of responsive priority are sufficiently versatile to be suitable for application in other circumstances. Thus these too suggest topics for further investigation. A list is given here of major topics which arise in these two ways.

a) The bus arrival pattern.

The assumption of Poissonian bus arrivals pervades this thesis. This assumption leads to considerable simplification in the analysis. The extent to which departures from this pattern affect the numerical results is certainly of interest. Headway distributions which might be considered include the normal, Erlang or gamma and possibly some with serial correlations.

b) Priority rules.

The responsive priority rules analysed here are the ones used most commonly in practice. However, many others could be treated in a similar manner. Two rules for which an analysis of delay would certainly be of interest are priority by extension and hurry-call and the active compensation rule described in section 3.5.
c) Optimisation problems.

No explicit optimisation problems which are appropriate to junctions where responsive priority is provided for buses have been considered here. As was discussed in section 1.2, once any particular evaluation criterion has been adopted, it should be optimised with respect to the signal-settings. Suppose, for example, that the minimum ratio of priority to non-priority vehicle occupancy required for responsive priority to give a net reduction in passenger delay is selected as an evaluation criterion. Then this quantity should be minimised before any comparisons are made. Thus the values of this ratio shown in figures 6.11 and 6.21 represent upper bounds for the optima.

d) Estimates of the overflow.

The estimates (5.68) and (5.27) for the mean overflow and the probability of a zero overflow are only approximate. These would certainly merit further investigation and might well be improved.

e) Estimation of mean delay.

The formula (5.50) for the mean delay incurred by traffic in a stream which experiences effective red and green periods of random duration could be applied in a variety of other circumstances. These include cases where the flow of traffic is interrupted sporadically. Any such application would involve estimation of some statistics of the duration of the effective red and green periods experienced.
REFERENCES


ALLSOP, R.E. (1972b) Estimating the traffic capacity of a signalized road junction. Transportation Research, 6(3), 245-55.


ALLSOP, R.E. (1973) Effects of errors in lost times on the delay to traffic at an isolated road junction controlled by fixed-time signals. Transportation Research, 7(2), 145-62.

ALLSOP, R.E. (1977a) Treatment of opposed turning movements in traffic signal calculations. Transportation Research, 11(6), 405-11.


HUTCHINSON, T.P. (1972) Delay at fixed time traffic signals - II: numerical comparison of some theoretical expressions. Transportation Science, 6(3), 286-305.


MURCHLAND, J.D. (1977) Webster's two-term traffic delay formula when there are several green intervals. Note from the Transport Studies Group, University College London, JDM-270 (unpublished).


OSTROWSKI, A.M. (1960) Solution of equations and systems of

OSUNA, E.E. and NEWELL, G.F. (1971) Control strategies for an
idealised public transportation system. Transportation
Science, 6(1), 52-72.

movers, followers and others: a reconsideration of headway
distributions. Traffic Engineering and Control, 21(8/9),
425-8.

Road Research, 2(2), 3-9.

RHEINBOLDT, W.C. (1980) Solution fields of nonlinear equations and
continuation methods. SIAM Journal on Numerical Analysis,
17(2), 221-37.


RICHARDSON, A.J. and OGDEN, K.W. (1979) Evaluation of active bus-
priority signals. Transportation Research Record, 718, 5-12.

RICHBELL, L.E. and Van AVERBEKE, B.A. (1972) Bus priorities at
traffic control signals. Traffic Engineering and Control,
14(2), 70-5.

ROAD RESEARCH LABORATORY (1963) A method of measuring saturation flow
at traffic signals. Road Note No. 34. London: HMSO.


traffic signals to reduce fuel consumption. Department of the
Environment and Department of Transport TRRL Report LR 934.
Crowthorne.

ROBERTSON, D.I. and VINCENT, R.A. (1975) Bus priority in a network of
fixed-time signals. Department of the Environment TRRL Report
LR 666. Crowthorne.

Greater Manchester. Traffic Engineering and Control, 15(9),
442-5.

SEWARD, S.R. and TAUBE, R.N. (1977) Methodology for evaluating bus-
actuated, signal-preemption systems. Transportation Research
Record, 630, 11-7.

SHIELDS, L.R. (1976) Bus reliability - study of a Central London bus
route. London Transport Planning Research Office, Operational


APPENDIX 1

IMPLICIT FUNCTIONS

A1.1 Introduction

In several places in this thesis, functions are defined by means of equations which are satisfied by the functions rather than by explicit formulae. Several properties which are apparent for explicit functions can be established only by analysis for these implicit ones. Results are given in this appendix which provide conditions that are strong enough to guarantee the existence, uniqueness and continuity of implicit functions and the convergence of a particular solution method to the required value.

Throughout this appendix, symbols of the form \( y \) are used to represent the ordered set \((y_1, \ldots, y_n)\) of all elements \( y_i \) which have been defined. This notation is used, where it results in no loss of clarity, for ordered sets of functions as well as for elements of \( \mathbb{R}^n \).

A1.2 Existence results

The main result of this section is the implicit function theorem which gives a sufficient condition for the existence of a unique function satisfying an equation in some neighbourhood of a point at which a solution is known to exist. The theorem also gives an explicit form for the partial derivatives of the implicit function in terms of the derivatives of the defining function: the one-dimensional version of this result is used frequently in the text. No proof of this theorem is given here as it is a standard result of functional analysis. A proof can be found in Dieudonné (1969, p 272).
Theorem A1.1 (Implicit function theorem, finite dimensional real version)

Let \( \tilde{F}_i(x,y) \) \((1 \leq i \leq m)\) be \( m \) scalar functions defined and continuously differentiable in a neighbourhood \( X \times Y \) of a point \((x^#, y^#)\) of \( \mathbb{R}^m \times \mathbb{R}^n \) such that \( \tilde{F}_i(x^#, y^#) = 0 \) \((1 \leq i \leq m)\), and the determinant of the Jacobian matrix \( \{ \partial \tilde{F}_i / \partial x_j \} \) is not 0 at \((x^#, y^#)\). Then there is an open neighbourhood \( W_0 \subset Y \) of \( y^# \) such that, for any connected open neighbourhood \( W \subset W_0 \) of \( y^# \), there is a unique system of \( m \) scalar functions \( \tilde{G}_i(y) \) \((1 \leq i \leq m)\), defined and continuous in \( W \) such that \( \tilde{G}_i(y^#) = x^#_i \) and for any \( y \in W \),

\[
\tilde{F}(\tilde{G}(y), y) = 0
\]

(A1.1)

Moreover, the functions \( \tilde{G}_i(y) \) are continuously differentiable for \( y \in W \) and the Jacobian matrix \( \{ \partial \tilde{G}_i / \partial y_j \} \) is equal to \( -B^{-1}A \), where \( A \) (resp. \( B \)) is obtained by replacing \( x_i \) by \( \tilde{G}_i(y) \) in the Jacobian matrix \( \{ \partial F_i / \partial y_j \} \) (resp. \( \{ \partial F_i / \partial x_j \} \)).

When applied to the problem of solving capacity-equivalence relations for signal settings \( \{ \lambda_j \} \) to be implemented with priority for buses arriving with a mean rate of \( \beta / c^# \), the implicit function theorem guarantees the existence of solutions for each \( \beta \) in some interval \([0, \beta_e]\). Clearly, if the hypotheses of the theorem were satisfied at \( \beta = \beta_e \), then this interval could be extended. The following result, given by Rheinboldt (1980, p 224) describes the possibilities for the behaviour of the functions \( \tilde{x}_j(\beta) \) \((1 \leq j \leq m)\) as \( \beta \) tends to the least upper bound for which solutions exist.

**Corollary A1.2**

Suppose that \( \tilde{F}_i(\lambda, \beta) \) \((1 \leq i \leq m)\) are \( m \) scalar functions defined and continuously differentiable in a neighbourhood \( X \times Y \) of the point \((\lambda^#, 0)\) of \( \mathbb{R}^m \times \mathbb{R} \) such that \( \tilde{F}(\lambda^#, 0) = 0 \) and the determinant of the Jacobian matrix
\( \{ \partial F_i / \partial \lambda_j \} \) is not 0 at \((\lambda^#, 0)\). Then there exists a unique continuously differentiable set of functions \( \tilde{\lambda}_i(\beta) \) \((1 \leq i \leq m)\) for which \( \text{det}(\{ \partial F_i / \partial \lambda_j \}) \) is not 0 which are defined on an interval \([0, \beta_c]\) where \( \beta_c \) is maximal. There are three possibilities for the behaviour near \( \beta = \beta_c \):

(a) \( \beta_c = \infty \)

(b) \( \left( \sum_{i=1}^{m} (\tilde{\lambda}_i(\beta))^2 \right)^{1/2} \rightarrow \infty \) as \( \beta \rightarrow \beta_c^- \)

(c) \( \text{det}(\{ \partial F_i / \partial \lambda_j \}) \left( \tilde{\lambda}(\beta), \beta \right) \rightarrow 0 \) as \( \beta \rightarrow \beta_c^- \)

A1.3 The Gauss-Seidel solution method

An iterative procedure is normally required to find the numerical values of an implicit function. In cases where there is only one variable to be evaluated, a variety of comparatively simple methods can be used and a suitable one is here supposed to be available. When there are several variables to be calculated simultaneously, the problem of finding the solutions is rather more complicated. An extensive treatment of both of these topics is given by Ortega and Rheinboldt (1970). The Gauss-Seidel method for solving simultaneous equations is described in this section and results are given which specify conditions which are sufficiently strong to guarantee convergence.

Suppose that \( m \) scalar equations in \( m \) variables \( \tilde{F}_i(x) = 0 \) \((1 \leq i \leq m)\) are to be solved simultaneously for \( x_i \) \((1 \leq i \leq m)\). In addition to the form of the functions \( \tilde{F}_i(x) \) \((1 \leq i \leq m)\), the Gauss-Seidel method requires an initial estimate of the solution \( x^* \). This estimate is denoted by \( x^{(0)} \).
A single iteration of the Gauss-Seidel method proceeds from the estimate \( \mathbf{x}^{(n)} \) to \( \mathbf{x}^{(n+1)} \) \((n \geq 0)\) as follows. Starting with \( i = 1 \), the single equation 
\[
\tilde{F}_i(x_1^{(n+1)}, \ldots, x_{i-1}^{(n+1)}, x_i^{(n)}, x_{i+1}^{(n)}, \ldots, x_m^{(n)}) = 0
\]
is solved for \( x_i^{(n+1)} \) and \( x_i^{(n+1)} \) is set to the value of this solution. If \( i = m \), then the iteration is complete; otherwise, the value of \( i \) is incremented and the above process is repeated.

This procedure is iterated until some termination criterion is satisfied. Two possible criteria for termination after iteration number \( n \) \((n \geq 1)\) are

\[
(a) \quad \left[ \sum_{i=1}^{m} \left( x_i^{(n)} - x_i^{(n-1)} \right)^2 \right]^{1/2} < \varepsilon_x \quad \text{for some tolerance } \varepsilon_x \quad (A1.2)
\]

\[
(b) \quad \left[ \sum_{i=1}^{m} \left( \tilde{F}_i(x^{(n)}) \right)^2 \right]^{1/2} < \varepsilon_F \quad \text{for some tolerance } \varepsilon_F \quad (A1.3)
\]

Since the Gauss-Seidel method uses the functions \( \tilde{F}_i(x) \) \((1 \leq i \leq m)\) sequentially rather than simultaneously, the order in which the functions are numbered can affect the behaviour of the method. Any material effects of re-numbering the functions will certainly be apparent in the satisfaction or otherwise of the hypotheses of the following results.

The first result of this section states a sufficient condition for the Gauss-Seidel method to converge to a solution of a set of equations. This result is given by Ortega and Rheinboldt (1970, p 326).

**Theorem A1.3**

Let \( \tilde{F}_i(x) \) \((1 \leq i \leq m)\) be \( m \) scalar functions defined and continuously differentiable in an open neighbourhood \( X \) of a point \( x^* \) for which 
\[
\tilde{F}(x^*) = 0
\]
. Consider the decomposition of the Jacobian matrix \( J = \{ \partial F_i / \partial x \} \) into diagonal, strictly lower-triangular and strictly upper-triangular parts:
\[ J = D - L - U \]

where
\[
D_{ij} = \begin{cases} 
\frac{\partial F_i}{\partial x_i} & (1 \leq i \leq m, j = i) \\
0 & \text{otherwise}
\end{cases}
\]
\[
L_{ij} = \begin{cases} 
\frac{\partial F_i}{\partial x_j} & (1 \leq i \leq m, j < i) \\
0 & \text{otherwise}
\end{cases}
\]
\[
U_{ij} = \begin{cases} 
\frac{\partial F_i}{\partial x_j} & (1 \leq i \leq m, j > i) \\
0 & \text{otherwise}
\end{cases}
\]

and let \( H = (D - L)^{-1} U \).

Denoting the maximum of the absolute values of the eigenvalues of a matrix \( M \) (the spectral radius of \( M \)) by \( \rho(M) \), suppose that

\[ \det[D(x^\#)] \neq 0 \]

(A1.6)

and \( \rho[H(x^\#)] < 1 \) (A1.7)

then for some \( \varepsilon > 0 \), if \( x^{(0)} \in X \) and \[ \sum_{i=1}^{m} (x^{(0)} - x^\#)^2_i < \varepsilon \], then there is a unique sequence \( \{ x^{(j)} \}_{j=0}^{\infty} \subset X \) with \[ \sum_{i=1}^{m} (x^{(j)} - x^\#)^2_i < \varepsilon \] \( (j=1, 2, ...) \) given by the Gauss-Seidel method. Moreover,

\[ \lim_{j \to \infty} x^{(j)}_i = x^\#_i \quad (1 \leq i \leq m) \]

(A1.8)

and \[ \sup_{x^{(0)}} \{ \lim_{j \to \infty} \sup_{j > J} \left[ \sum_{i=1}^{m} (x^{(j)} - x^\#)^2_i \right] \} = \rho[H(x^\#)] \]

(A1.9)
Note that the hypothesised continuity of the functions \( \tilde{F}_i(x) \) at \( x = x^* \) guarantees the convergence of each of the values \( \tilde{F}_i(x^{(j)}) \) to 0 (1 ≤ i ≤ m) as \( j \to \infty \) and thus that both of the convergence criteria (A1.2) and (A1.3) will be met. The next result considers the application of the Gauss-Seidel method to implicit functions.

**Corollary A1.4**

Suppose that the hypotheses of theorem A1.1 are satisfied for certain variables \( y \) and a point \( x^* \) and that

\[
\det[\tilde{D}(x^*, y^*)] \neq 0 \quad (A1.10)
\]

\[
\rho[\tilde{H}(x^*, y^*)] < 1 \quad (A1.11)
\]

where \( \tilde{D} \) and \( \tilde{H} \) are given by (A1.4) and (A1.5) respectively. Then there is an open neighbourhood \( W \) of \( x^* \) where \( x \in W \) of theorem A1.1 such that if \( x \in W \), then for some \( \varepsilon(y) > 0 \), if \( x^{(0)} \in X \) and \( \left[ \sum_{i=1}^{m} (x_i^{(0)} - \tilde{G}_i(y))^2 \right]^{1/2} < \varepsilon(y) \), where \( \tilde{G}(y) \) satisfies (A1.1), then there is a unique sequence \( \{x^{(j)} : j = 0, 1, \ldots\} \subseteq X \) with \( \left[ \sum_{i=1}^{m} (x_i^{(j)} - \tilde{G}_i(y))^2 \right]^{1/2} < \varepsilon(y) \) (j = 0, 1, ...) given by the Gauss-Seidel method.

Moreover,

\[
\lim_{j \to \infty} x_i^{(j)} = \tilde{G}_i(y) \quad (1 \leq i \leq m) \quad (A1.12)
\]

\[
\sup_{x^{(0)}} \left( \limsup_{j \to \infty} \left[ \sum_{i=1}^{m} (x_i^{(j)} - \tilde{G}_i(y))^2 \right]^{1/2} \right) = \rho[\tilde{H}(\tilde{G}(y), y)] \quad (A1.13)
\]
Proof

From the implicit function theorem (A1.1), \( \tilde{G}(y) \) is well defined for each \( y \in W_0 \). Now \( \det(M) \) is a continuous function of the elements of the matrix \( M \), so by the continuity of each of the partial derivatives of \( \tilde{F}_i(x) \) (1\( \leq i \leq m \)) and of the functions \( \tilde{G}_i(y) \) (1\( \leq i \leq m \)), there is an open neighbourhood \( W_D \) of \( y^* \) for which if \( y \in W_D \) then

\[
\det[\tilde{D}(\tilde{G}(y),y)] \neq 0
\]  

(A1.14)

Similarly, since the eigenvalues of a matrix are continuous functions of the elements of the matrix (Ostrowski, 1960, p 192) and the elements of \( H \) are continuous functions of the elements of \( J \) in some neighbourhood of \( (x^*,y^*) \), there is an open neighbourhood \( W_s \) of \( y^* \) for which if \( y \in W_s \), then

\[
\rho[\tilde{H}(\tilde{G}(y),y)] < 1
\]  

(A1.15)

Let \( W = W_0 \cap W_D \cap W_s \), then for each \( y \in W \), all of the hypotheses of theorem A1.3 are satisfied, so the theorem can be applied to establish the required result.

The last result of this appendix, which follows from the continuity of the spectral radius of the matrix \( H \) (Ortega and Rheinboldt, 1970, p 334) shows that any solution of a system of implicit functions is a suitable starting point for a slightly perturbed problem. In the particular case of solving capacity-equivalence relationships for new signal settings \( \tilde{\chi}_i(\beta) \) (1\( \leq i \leq m \)), since \( \lambda^*_i \) (1\( \leq i \leq m \)) are solutions when \( \beta = 0 \), they are suitable starting points for the Gauss-Seidel method for each \( \beta \) in some interval \([0,\beta_G)\).
Corollary A1.5

Suppose that the hypotheses of corollary A1.4 are satisfied. Then there is an open neighbourhood \( W_G \) of \( x^* \) where \( W_G \subseteq W_0 \) of theorem A1.1 such that if \( y \in W_G \), then there is a unique sequence \( \{x^{(j)}_i \mid j \geq 0, 1, \ldots \} \subseteq \mathbb{X} \) given by the Gauss-Seidel method where

\[
x^{(0)}_i = x^*_i
\]

and

\[
\lim_{{j \to \infty}} x^{(j)}_i = \tilde{G}_i(y) \quad (1 \leq i \leq m)
\]

When the result of corollary A1.5 is used in the text, \( x \) represents \( \lambda \) and \( y \) represents the scalar \( \beta \). Thus only the case where \( y \) is of order 1 is required.
APPENDIX 2

AN UNEXPECTED PROPERTY OF CAPACITY-EQUIVALENT SIGNAL-SETTINGS
WHEN PRIORITY IS PROVIDED BY RECALL

A2.1 Introduction

The result of theorem 4.53 in sub-section 4.5.5 provided formulae for the behaviour of signal-settings which, when implemented with priority by recall, would be capacity-equivalent to a given set as the mean bus arrival rate varies. Formulae were given for the total derivatives \( \frac{d\lambda_i}{d\beta} \) (1 \( \leq i \leq m \)) in terms of the solutions of the Equations (4.127). In the discussion which followed, the possibility was mentioned that \( \frac{d\lambda_i}{d\beta} \) could be strictly negative for some \( i \) (1 \( < i \leq m \)) in some circumstances. If this occurs, then in order to maintain capacity-equivalence as the mean bus arrival rate increase, the maximum duration allowed for stage \( i \) must be reduced.

In this appendix, an example is given which illustrates this possibility. A formula is given for the total derivative \( \frac{d\lambda_{k+1}}{d\beta} \) when the mean bus arrival rate is 0 and there are exactly 2 stages in the cycle during which buses do not have right of way. The data for one of the example junctions considered in Chapter 6 are then modified to provide an example for which \( \frac{d\lambda_{k+1}}{d\beta} < 0 \) when \( \beta = 0 \).

A2.2 Analysis

Theorem A2.1

Suppose that priority by recall is implemented at a junction where there are exactly 2 stages during which buses do not have right of way and that the signal-settings \( \{\tilde{\lambda}_i(\beta)\} \) are calculated to be capacity-equivalent to
some given set \( \{ \lambda_1^* \} \) when these are implemented in the absence of priority. Then the total derivative of the duration of stage \( k+1 \) with respect to the mean bus arrival rate satisfies

\[
\lim_{\beta \to 0} \frac{d\lambda_{k+1}}{db} = (\lambda_{k,0}^* + \gamma_{k+1})(\lambda_{k+1,0}^* - \lambda_{k+1})^2/2 - \\
\lambda_{k+1}^* (\lambda_{k,0}^* + \lambda_{k+1}^*) (\lambda_{k+1,0}^* + \lambda_{k+1,2,1} - \lambda_{k+1,1,1})/\lambda_0^* \tag{A2.1}
\]

**Proof**

From (4.140), given as the result of theorem 4.53, the total derivative is given by

\[
\frac{d\lambda_{k+1}}{d\beta} = \frac{\partial \lambda_{k+1}}{\partial \beta} + \frac{\partial \lambda_{k+1}}{\partial \epsilon_0} \frac{\partial \epsilon_0}{\partial \beta} \tag{A2.2}
\]

where

\[
\frac{\partial \lambda_{k+1}}{\partial \beta} = - \left[ \left( \frac{\partial \epsilon_{k+1}}{\partial \lambda_{k+1}} \right)^{-1} \frac{\partial \epsilon_{k+1}}{\partial \beta} \right] \tag{A2.3}
\]

\[
\frac{\partial \lambda_{k+1}}{\partial \epsilon_0} = \left( \frac{\partial \epsilon_{k+1}}{\partial \lambda_{k+1}} \right)^{-1} \frac{\lambda_{k+1}}{\lambda_0^*} \tag{A2.3}
\]

\[
\frac{\partial \epsilon_0}{\partial \beta} = - (\lambda_{k,0}^* + \lambda_{k+1}) \tilde{p}_0 (\lambda_{k,0}^* + \lambda_{k+1}) (\lambda_{k+1,0}^* + \lambda_{k+1,2,1} - \lambda_{k+1,1,1}) \quad \text{(since } m=k+2) \tag{A2.4}
\]

\[
\frac{\partial \epsilon_{k+1}}{\partial \lambda_{k+1}} = \tilde{p}_0 (\lambda_{k,0}^* + \lambda_{k+1}) \tag{A2.4}
\]

and \( \frac{\partial \epsilon_{k+1}}{\partial \beta} = - (\lambda_{k,0}^* + \gamma_{k+1}) \tilde{p}_0 (\lambda_{k,0}^* + \gamma_{k+1}) \left[ 1 - \tilde{p}_0 (\lambda_{k+1} - \gamma_{k+1}) \right] / \beta - \\
- \tilde{p}_0 (\lambda_{k,0}^* + \gamma_{k+1}) \left[ 1 - \tilde{p}_0 (\lambda_{k+1} - \gamma_{k+1}) \right] / \beta^2
\]

Taking the limit of (A2.4) as the mean bus arrival rate tends to 0, using the rule of Count de l'Hopitale and that \( \lim_{\beta \to 0} \lambda_1^*(\beta) = \lambda_1^* \) gives
\[
\lim_{\beta \to 0} \frac{\partial \varepsilon_0}{\partial \beta} = -(\lambda_{k,0}^{+\lambda_{k+1}}) (\lambda_{k+1,0}^{+\lambda_{k+2,1}^{+\lambda_{k+1,1}}} \\
\lim_{\beta \to 0} \frac{\partial \varepsilon_{k+1}}{\partial \lambda_{k+1}} = 1 \\
\text{and } \lim_{\beta \to 0} \frac{\partial \varepsilon_{k+1}}{\partial \beta} = -(\lambda_{k,0}^{+\gamma_{k+1}}) (\lambda_{k+1,0}^{+\gamma_{k+1}}) - (\lambda_{k+1,0}^{+\gamma_{k+1}})^2/2
\] 

(A2.5)

Substituting these limiting forms into (A2.3) and using these in (A2.2) gives (A2.1).

\[\text{A2.3 An example}\]

In order to illustrate the possibility that the derivative (A2.1) can take strictly negative values, the Derby junction example used in Section 6.2 is modified. The desired result can be achieved by increasing the minimum permissible duration for stage 2 from 7 seconds to 20 seconds. The resulting signal-settings are given in Table A2.1. Using these data in the formula (A2.1) gives

\[
\lim_{\beta \to 0} \frac{d\lambda_2}{d\beta} = -146 \text{ s}^2/\text{bus}.
\]

Thus the following result has been established.

\[\text{Corollary A2.2}\]

Under the hypotheses of theorem 4.46, there are cases in which \(\frac{d\lambda_1}{d\beta} < 0\) for some \(i\) \((k<i≤m)\).
Table A2.1: Signal-settings used for the example in Section A2.3 (seconds)

<table>
<thead>
<tr>
<th>Stage number</th>
<th>Stage durations</th>
<th>Transition period following</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>uninterrupted</td>
<td>minimum</td>
</tr>
<tr>
<td>1</td>
<td>16.0</td>
<td>7.0</td>
</tr>
<tr>
<td>2</td>
<td>24.0</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>28.0</td>
<td>7.0</td>
</tr>
</tbody>
</table>
APPENDIX 3
INDEX OF MAJOR NOTATION

A3.1 Introduction

A considerable amount of symbolic notation has been introduced in the course of this thesis. This appendix provides an index of the first usage of the major items of notation. In most cases, brief definitions are given for these items. The index is not exhaustive in the sense that some notation which is used only where it is introduced in the text is not included here. Furthermore, not all of the variations in symbols by way of subscripts, superscripts and other decorations are given here as many of these are explained within the text.

A3.2 Lists of notation

A3.2.1 Introduction.

Three separate lists of symbols are given in this section. These are one each for upper and lower case Roman letters and one for all Greek letters. Each list is ordered lexicographically.
A3.2.2  Upper case Roman letters.

\( A(t) \)  5.2.1  Number of vehicular arrivals in an interval of duration \( t \)

\( B \)  2.2  Total distance travelled by vehicles in unit time

\( C_s \)  5.2.1  Coefficient of variation of departure headways

\( C(j) \)  5.3.1  Class to which stream \( j \) belongs (1\( \leq j \leq M \))

\( D_t \)  2.2  Total rate of delay

\( D \)  5.2.1  Total delay incurred in a stream of traffic during a single pair of effective red and green periods

\( F \)  2.2  Total rate of fuel consumption

\( G_{iM} \)  4.2.3  Minimum permissible duration for stage \( i \) (1\( \leq i \leq m \))

\( H \)  2.2  Total number of stops made per unit time

\( H_s \)  5.2.2  Mean number of stops incurred by a vehicle

\( I_a \)  4.4.1  Index of dispersion of vehicular arrivals

\( I \)  5.3.3  The state 'inhibition is in effect'

\( K_j \)  4.2.1  Capacity of stream \( j \) (pou's per unit time) (1\( \leq j \leq M \))

\( L \)  4.2.3  Total duration of the transition periods in an uninterrupted cycle

\( M \)  4.2.1  Number of streams at a junction

\( \tilde{N}(g) \)  5.2.1  Number of vehicular departures from the queue in an effective green period of duration \( g \)

\( P_n \)  4.2.3  Probability that there are exactly \( n \) stages in a cycle (k\(<n\leq m\))

\( P_I \)  4.4.5  Probability that inhibition is in effect at the start of stage 1

\( P_s \)  5.2.2  Probability that a vehicle will have to stop at a set of traffic lights

\( \tilde{p}^{(n)}_{S_1S_2} \)  (N)  5.3.3  Transition probability

\( Q_n \)  5.2.1  Overflow (n\( \geq 1 \))

\( \tilde{Q}(t) \)  5.2.1  Number of vehicles in the queue at time \( t \) after the end of the last effective green period
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>S(n)</td>
<td>Time taken for n vehicles to depart from the queue</td>
</tr>
<tr>
<td>U</td>
<td>The state 'inhibition is not in effect'</td>
</tr>
<tr>
<td>V</td>
<td>Total number of priority vehicles to arrive during a single pair of effective red and green periods</td>
</tr>
<tr>
<td>W</td>
<td>Total delay incurred by priority vehicles during a single pair of effective red and green periods</td>
</tr>
</tbody>
</table>
A3.2.3 Lower case Roman letters.

\[ a_{ij} \quad 4.2.3 \quad \text{Proportion of the saturation departure rate } a_j \text{ which obtains for stream } j \text{ during stage } i \quad (1 \leq i \leq m, 1 \leq j \leq M) \]

\[ a_{0j} \quad 4.2.3 \quad \text{Effective proportion of all transition periods for which stream } j \text{ has right of way } (1 \leq j \leq M) \]

\[ a_{i0j} \quad \{ \quad 4.2.3 \quad \text{Effective proportion of specified transition periods for which stream } j \text{ has right of way } (1 \leq i \leq m, 1 \leq j \leq M) \]

\[ a_{i1j} \quad \} \quad 4.2.3 \quad \text{Duration of an uninterrupted cycle} \]

\[ c \quad 4.2.3 \quad \text{Mean delay incurred by non-priority vehicles} \]

\[ d \quad 5.2.1 \quad \text{Mean time between consecutive bus arrivals} \]

\[ e \quad 4.2.1 \quad \text{Duration of an effective green period} \]

\[ k \quad 4.2.3 \quad \text{Number of stages during which priority vehicles have right of way} \]

\[ m \quad 4.2.3 \quad \text{Number of stages in the control sequence} \]

\[ n_1(j) \quad \{ \quad 5.3.1 \quad \text{Numbers of the stages which form the first and last parts of the periods during which stream } j \text{ has right of way in uninterrupted cycles } (1 \leq j \leq M) \]

\[ n_2(j) \quad \} \]

\[ p_j \quad 4.2.2 \quad \text{Maximum acceptable degree of saturation for stream } j \quad (1 \leq j \leq M) \]

\[ p_0(\xi) \quad 4.4.1 \quad \text{Probability that no priority vehicles arrive in an interval of duration } \xi \]

\[ q_j \quad 4.2.1 \quad \text{Mean arrival rate of vehicles in stream } j \text{ (vehicles per unit time) } (1 \leq j \leq M) \]

\[ r \quad 4.2.1 \quad \text{Duration of an effective red period} \]

\[ s \quad 2.2 \quad \text{Standard deviation of the time between consecutive bus arrivals} \]

\[ s_j \quad 4.2.1 \quad \text{Saturation departure rate for stream } j \text{ (pceu's per unit time) } (1 \leq j \leq M) \]

\[ t \quad 5.2.1 \quad \text{Time since the end of the last effective green period} \]
\( w \) 2.2 Mean waiting time for bus passengers
\( w \) 5.4.2 Delay incurred by a priority vehicle
\( x_j \) 4.3.3 Degree of saturation for stream \( j \) (\( 1 \leq j \leq M \))
A3.2.4 Greek letters.

$\beta$ 4.4.1 Mean bus arrival rate (buses per uninterrupted cycle)

$\gamma_i$ 4.2.3 Minimum permissible duration for stage $i$ (proportion of $c$)  
($1 \leq i \leq m$)

$\delta_{ij}$ 4.2.3 Kronecker delta

$\delta_n$ 5.2.4 Unused capacity in service batch $n$ ($n \geq 1$)

$\epsilon_0$ 4.2.3 Mean sum of duration of transition periods in a cycle  
(proportion of $c$)

$\epsilon_i$ 4.2.3 Mean duration of stage $i$ (proportion of $c$) ($1 \leq i \leq m$)

$\eta_i$ 4.4.4 Conditional mean duration of stage $i$ given that there are  
equally $i$ stages in the cycle (proportion of $c$) ($1 \leq i \leq m$)

$\eta_{2i}$ 4.4.4 Conditional second moment of the duration of stage $i$ given  
equally $i$ stages in the cycle (units of $c^2$)  
($1 \leq i \leq m$)

$\Lambda_j$ 4.2.1 Effective proportion of all time for which stream $j$ has  
right of way with saturation departure rate $s_j$ ($1 \leq j \leq M$)

$\lambda_0$ 4.2.3 Total duration of transition periods in an uninterrupted  
cycle (proportion of $c$)

$\lambda_i$ 4.2.3 Duration of stage $i$ when uninterrupted (proportion of $c$)  
($1 \leq i \leq m$)

$\lambda_{i0}$ 4.2.3 Duration of specified transition periods (proportion of $c$)

$\lambda_{i1}$ 4.2.3 ($1 \leq i \leq m$)

$\tau$ 4.4.1 Duration of a priority extension period (proportion of $c$)

$\upsilon_k$ 4.4.2 Standard deviation of duration of stage $k$ (units of $c^2$)